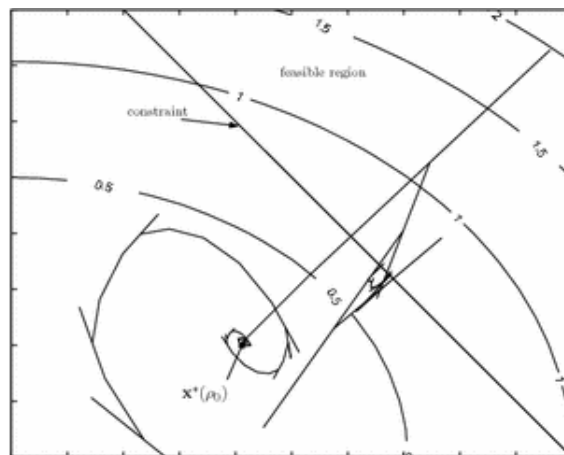


Ανάπτυξη και Θεμελίωση νέων μεθόδων Υπολογιστικών Μαθηματικών στην Υπολογιστική Νοημοσύνη

Διδακτορική Διατριβή

Σταμάτιος-Άγγελος Ν. Αλεξανδρόπουλος



Πανεπιστήμιο Πατρών
Σχολή Θετικών Επιστημών
Τμήμα Μαθηματικών
Πάτρα

Επιβλέπων: Καθηγητής Μιχαήλ Ν. Βραχάτης

(Μάρτιος 2020)

« Το έργο συγχρηματοδοτείται από την Ελλάδα και την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) μέσω του Επιχειρησιακού Προγράμματος «Ανάπτυξη Ανθρώπινου Δυναμικού, Εκπαίδευση και Διά Βίου Μάθηση», στο πλαίσιο της Πράξης «Ενίσχυση του ανθρώπινου ερευνητικού δυναμικού μέσω της υλοποίησης διδακτορικής έρευνας» (MIS-5000432), που υλοποιεί το Ίδρυμα Κρατικών Υποτροφιών (ΙΚΥ) »



Επιχειρησιακό Πρόγραμμα
Ανάπτυξη Ανθρώπινου Δυναμικού,
Εκπαίδευση και Διά Βίου Μάθηση
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Η παρούσα Διδακτορική Διατριβή του Σταματίου-Αγγέλου Ν. Αλεξανδρόπουλου με τίτλο «Ανάπτυξη και Θεμελίωση νέων μεθόδων Υπολογιστικών Μαθηματικών στην Υπολογιστική Νοημοσύνη» εξετάστηκε και εγκρίθηκε από την ακόλουθη Επιταμελή Εξεταστική Επιτροπή :

Μ.Ν. Βραχάτης, Καθηγητής
Τμήμα Μαθηματικών
Πανεπιστήμιο Πατρών

Π. Αλεβίζος, Αναπληρωτής Καθηγητής
Τμήμα Μαθηματικών
Πανεπιστήμιο Πατρών

Σ. Κωτσιαντής, Επίκουρος Καθηγητής
Τμήμα Μαθηματικών
Πανεπιστήμιο Πατρών

Θ. Γράφα, Καθηγήτρια
Τμήμα Μαθηματικών
Πανεπιστήμιο Πατρών

Κ. Παρσόπουλος, Αναπληρωτής Καθηγητής
Τμήμα Μηχανικών Η/Υ και Πληροφορικής
Πανεπιστήμιο Ιωαννίνων

Γ. Ανδρουλάκης, Αναπληρωτής Καθηγητής
Τμήμα Διοίκησης Επιχειρήσεων
Πανεπιστήμιο Πατρών

Σ. Αδάμ, Επίκουρος Καθηγητής
Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Πανεπιστήμιο Ιωαννίνων

© 2020

Πανεπιστήμιο Πατρών, Τμήμα Μαθηματικών
Σταμάτιος-Άγγελος Ν. Αλεξανδρόπουλος
Με την επιφύλαξη παντός δικαιώματος

ΑΥΤΗ Η ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ ΣΤΟΙΧΕΙΟΘΕΤΗΘΗΚΕ ΜΕ ΤΟ ΠΡΟΓΡΑΜΜΑ $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ (ΔΙΑΝΟΜΗ $\text{M}\text{I}\text{K}\text{T}_{\text{E}}\text{X}$). Η ΣΥΓΓΡΑΦΗ ΕΓΙΝΕ ΜΕ ΤΗ ΒΟΗΘΕΙΑ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ $\text{W}\text{I}\text{N}\text{E}\text{D}\text{T}$ (ΣΤΟ ΛΕΙΤΟΥΡΓΙΚΟ ΣΥΣΤΗΜΑ $\text{M}\text{I}\text{C}\text{R}\text{O}\text{S}\text{O}\text{F}\text{T}$ $\text{W}\text{I}\text{N}\text{D}\text{O}\text{W}\text{S}$ 10) ΚΑΙ ΤΟΥ $\text{O}\text{N}\text{L}\text{I}\text{N}\text{E}$ $\text{E}\text{D}\text{I}\text{T}\text{O}\text{R}$ $\text{O}\text{V}\text{E}\text{R}\text{L}\text{E}\text{A}\text{F}$ (ΣΤΗΝ ΙΣΤΟΣΕΛΙΔΑ $\text{W}\text{W}\text{W}.\text{O}\text{V}\text{E}\text{R}\text{L}\text{E}\text{A}\text{F}.\text{C}\text{O}\text{M}$). Η ΤΕΛΙΚΗ ΗΛΕΚΤΡΟΝΙΚΗ ΜΟΡΦΗ ($\text{P}\text{O}\text{R}\text{T}\text{A}\text{B}\text{L}\text{E}$ $\text{D}\text{O}\text{C}\text{U}\text{M}\text{E}\text{N}\text{T}$ $\text{F}\text{O}\text{R}\text{M}\text{A}\text{T}$ - PDF) ΔΗΜΙΟΥΡΓΗΘΗΚΕ ΜΕ ΤΟ ΠΡΟΓΡΑΜΜΑ $\text{P}\text{D}\text{F}\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ ΚΑΙ Η ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΣΧΗΜΑΤΩΝ ΕΓΙΝΕ ΜΕ ΤΑ ΠΡΟΓΡΑΜΜΑΤΑ $\text{A}\text{D}\text{O}\text{B}\text{E}$ $\text{P}\text{H}\text{O}\text{T}\text{O}\text{S}\text{H}\text{O}\text{P}$ ΚΑΙ $\text{G}\text{I}\text{M}\text{P}$.

Στο εξώφυλλο της Διδακτορικής Διατριβής ο αναγνώστης συναντά τις δυναμικές τροχιές που παράγονται από μια νέα μέθοδο που στηρίζεται στην κλίση (gradient) μιας αντικειμενικής συνάρτησης, όπως αυτή παρουσιάζεται στην εργασία [367]. Ο αναγνώστης παρατηρεί τις ισοϋψείς καμπύλες της συνάρτησης καθώς και τις διαδοχικές τροχιές που παράγονται στην πορεία τους προς το χ^2 .

Στο δάσκαλό μου Μιχάλη

Στη μνήμη των αγαπημένων μου
φίλων Παναγιώτη και Ηλία

Περίληψη

Τρία είναι τα μυστικά της επιτυχίας: δουλειά, δουλειά, δουλειά.
Και ένα τέταρτο για να μην τα ξεχνάς: δουλειά.



Δύο πολύ σημαντικά επιστημονικά πεδία, αυτά της Υπολογιστικής Νοημοσύνης και των Υπολογιστικών Μαθηματικών, ενδείκνυνται για την αποτελεσματική και αποδοτική αντιμετώπιση σύνθετων προβλημάτων του πραγματικού κόσμου. Ένα ευρύ φάσμα μεθόδων και τεχνικών, έχει αναπτυχθεί με βάση τα δύο προαναφερθέντα επιστημονικά πεδία, συγκεντρώνοντας την προσοχή της επιστημονικής κοινότητας. Αυτό που παρατηρείται συχνά είναι αλγόριθμοι και τεχνικές, που είναι προσαρμοσμένες σε ένα συγκεκριμένο πρόβλημα, δίχως να είναι σε θέση να ανταποκριθούν το ίδιο καλά σε άλλα, παρόμοια προβλήματα. Η «συνεργασία» των δύο παραπάνω επιστημονικών τομέων δύναται να παράσχει αλγορίθμους με μαθηματική θεμελίωση για την αξιόπιστη αντιμετώπιση μιας πληθώρας συγγενικών προβλημάτων.

Οι αλγόριθμοι και τα μοντέλα που συναντώνται στον τομέα της Υπολογιστικής Νοημοσύνης και έχουν τις βάσεις τους σε μεθόδους των Υπολογιστικών Μαθηματικών (Αριθμητικής Επίλυσης Συνήθων Διαφορικών Εξισώσεων), αποτελούν ένα σημαντικό μέρος αυτής της διατριβής. Η Υπολογιστική Νοημοσύνη, σαν νέος επιστημονικός κλάδος, περιγράφει υπολογιστικές μεθόδους που εμπνέονται από φυσικά και βιολογικά συστήματα και δύναται να συνδυάσει επιστημονικούς τομείς, όπως τα Υπολογιστικά Μαθηματικά και την Επιστήμη των Υπολογιστών, με αρκετές επεκτάσεις και εφαρμογές σχεδόν σε όλους τους τομείς των θετικών επιστημών και όχι μόνο.

Πολλές και διαφορετικές εφαρμογές που άπτονται σε αρκετούς τομείς της έρευνας και της τεχνολογίας εμφανίζονται ή μπορούν να αντιμετωπιστούν ως προβλήματα Βελτιστοποίησης. Αρκετές μέθοδοι της Υπολογιστικής Νοημοσύνης αναφέρονται σε μια ειδική κλάση μεθόδων βελτιστοποίησης που περιλαμβάνει τη μελέτη και τη θεμελίωση υπολογιστικών μεθόδων, οι οποίες αντιμετωπίζουν αποτελεσματικά, δύσκολα προβλήματα του φυσικού κόσμου. Οι μέθοδοι αυτοί είναι πολύ σημαντικές, καθώς προσεγγίζουν προβλήματα Βελτιστοποίησης χωρίς περιορισμούς, κάτι που συναντάται συχνότερα στα περισσότερα προβλήματα της τεχνολογίας. Ακόμα, λόγω του εύρους των εφαρμογών, και κυρίως της ανάπτυξης νέων τεχνολογιών, έχουν αναπτυχθεί προβλήματα που επιζητούν λύση και άπτονται σε διαφορετικά είδη Βελτιστοποίησης, όπως η πολυ-αντικειμενική ή Βελτιστοποίηση με περιορισμούς. Για τον χειρισμό αυτών των δύσκολων προβλημάτων, χρησιμοποιούνται μέθοδοι της Υπολογιστικής Νοημοσύνης και συγκεκριμένα, τεχνικές και αλγόριθμοι που ανήκουν στον τομέα της Μηχανικής Μάθησης και των Τεχνητών Νευρωνικών Δικτύων.

Στα πλαίσια της παραπάνω συνεργατικής αντιμετώπισης, είναι απαραίτητη η μελέτη των προαπαιτούμενων αυτών των αλγορίθμων. Έτσι, πριν την εφαρμογή αυτών των τεχνικών, είναι αναγκαία η προεργασία και προετοιμασία του διαθέσιμου συνόλου δεδομένων, προκειμένου να δημιουργηθούν αξιόπιστα μοντέλα, με καλή απόδοση και καλή ικανότητα γενίκευσης. Κατόπιν, αυτές οι μέθοδοι μπορούν να εφαρμοστούν και να αντιμετωπίσουν προβλήματα που άπτονται σε αρκετούς τομείς της επιστήμης, όπως η μηχανική, η φυσική, η βιολογία, τα μαθηματικά, η ιατρική, η επιστήμη των υπολογιστών, η βιομηχανία, η μουσική,

η οικονομία, η κρυπτογραφία κ.ά.

Στη σύγχρονη βιβλιογραφία, παρουσιάζεται ένα τεράστιο πλήθος αλγορίθμων και τεχνικών που άπτονται στις παραπάνω κατευθύνσεις. Παρόλα αυτά, η πλειοψηφία τους αφορά ένα συγκεκριμένο πλαίσιο, το οποίο είναι προσαρμοσμένο στις ανάγκες του εκάστοτε προβλήματος. Συνήθως περιορίζονται σε στατιστικά μοντέλα και σε παραμέτρους που πληρούν ένα συγκεκριμένο πρόβλημα, αποκλείοντας, έτσι, την επιτυχή εφαρμογή σε οποιοδήποτε άλλο πρόβλημα, ή ακόμα και στο ίδιο πρόβλημα με διαφορετικές ανάγκες. Μια αξιόπιστη λύση σε αυτό το μείζον ζήτημα παρέχεται μέσω της ανάπτυξης αλγορίθμων με μαθηματική θεμελίωση και μαθηματική απόδειξη. Ένας αλγόριθμος με ισχυρό μαθηματικό υπόβαθρο μπορεί να χρησιμοποιηθεί σε περισσότερες εφαρμογές με μεγαλύτερη επιτυχία.

Προς αυτή την κατεύθυνση στηρίζεται ο πυρήνας της διδακτορικής διατριβής, δηλαδή στη μαθηματική θεμελίωση αλγοριθμικών τεχνικών, οι οποίες θα απευθύνονται σε μια πλειάδα προβλημάτων, καλύπτοντας, έτσι, υπαρκτά κενά στην επιστημονική βιβλιογραφία ως προς την Υπολογιστική Νοημοσύνη και τα Υπολογιστικά Μαθηματικά. Πιστεύουμε και προσδοκούμε ότι τα αποτελέσματα της διατριβής θα συνεισφέρουν στη δημιουργία αξιόπιστων μεθόδων, οι οποίες, μεταξύ άλλων, θα δώσουν τη δυνατότητα σε επιστήμονες άλλων περιοχών ως προς την αξιόπιστη αντιμετώπιση μεγάλης κλίμακας και πολυπλοκότητας προβλημάτων, τα οποία, συχνά, εμφανίζονται σε πολλούς επιστημονικούς, οικονομικούς, βιομηχανικούς και εμπορικούς τομείς με προφανή οφέλη.

Συγκεκριμένα, η διδακτορική διατριβή ξεκινά με μια σύντομη εισαγωγή στο Κεφάλαιο 1, στην οποία ο αναγνώστης συναντά τις βασικές πτυχές της Υπολογιστικής Νοημοσύνης, των Υπολογιστικών Μαθηματικών, καθώς και σημαντικά σημεία της Βελτιστοποίησης και Μηχανικής Μάθησης. Στο Κεφάλαιο 2, παρουσιάζονται πολύ σημαντικά ζητήματα της Μαθηματικής Βελτιστοποίησης και αναλυτικότερα, αποτελέσματα που σχετίζονται με το γνωστό θεώρημα 'No free lunch theorems for optimization' αλλά και την πιθανή ύπαρξη 'Free lunches'. Στο επόμενο κεφάλαιο, ο αναγνώστης συναντά την ανάπτυξη μιας νέας οικογένειας μεθόδων Βελτιστοποίησης, όπως αυτή εμπνεύστηκε από τις δυναμικές τροχιές ανίχνευσης των Snyman-Fatti και τις μεθόδους Runge-Kutta για την αριθμητική επίλυση συνήθων διαφορικών εξισώσεων. Επίσης, ο αγνώστης μπορεί να μελετήσει το προτεινόμενο θεώρημα μέσω του οποίου τεκμηριώνεται αυτή η οικογένεια, όπως και τη σχετική απόδειξη. Το Κεφάλαιο 4 καταπιάνεται με πτυχές της Μηχανικής Μάθησης, και συγκεκριμένα με τα απαιτούμενα των αλγορίθμων (προεπεξεργασία των δεδομένων), ώστε αργότερα να μπορούν αυτοί να εφαρμοστούν αποτελεσματικά σε διάφορα προβλήματα, όπως αυτό της ελαχιστοποίησης, της ταξινόμησης κ.ά. Στο κεφάλαιο που ακολουθεί, παρουσιάζεται μια νέα, υβριδική μέθοδος για την αντιμετώπιση του προβλήματος αναγνώρισης ακραίων τιμών σε ένα σύνολο δεδομένων, όπως και οι σχετικές συγκρίσεις αυτής της μεθόδου με γνωστές και ευρέως χρησιμοποιούμενες μεθόδους αυτής της κατηγορίας. Στο Κεφάλαιο 6, αναπτύσσεται μια συνεργατική μέθοδος για την προσέγγιση του προβλήματος ταξινόμησης σε ποικίλα σύνολα δεδομένων, μαζί με τα αντίστοιχα αποτελέσματα και τις συγκρίσεις που διεξήχθησαν. Ακόμα, στο Κεφάλαιο 7, αναλύονται βασικά ζητήματα της πολυ-αντικειμενικής Βελτιστοποίησης, χαρτογραφούνται οι περιοχές εφαρμογής της, όπως και οι αλγόριθμοι Υπολογιστικής Νοημοσύνης, οι οποίοι χρησιμοποιούνται για την προσέγγιση τέτοιων προβλημάτων. Η διδακτορική διατριβή ολοκληρώνεται με το Κεφάλαιο 8, στο οποίο ο αναγνώστης συναντά μια σύνοψη, χρήσιμα συμπεράσματα, όπως και μελλοντικές ερευνητικές προσπάθειες.

Abstract

There are three secrets of success: work, work, work.
And a fourth one in order to remember the above mentioned: work.

—M. N. Vrahatis

Two very important scientific fields, Computational Intelligence and Computational Mathematics, are useful for addressing efficiently and effectively complex real-world problems. A variety of methods and techniques have been developed based on the two aforementioned scientific fields, detaching the attention of the scientific community. It is often observed algorithms and techniques, which are adapted to a particular problem without being able to respond equally well to other, similar problems. The 'collaboration' of the two abovementioned disciplines can provide mathematical algorithms in order to reliably deal with a plethora of problems.

The algorithms and the models that are included in the field of Computational Intelligence and are based on Computational Mathematics (Numerical Solution of Ordinary Differential Equations) are an important part of this thesis. Computational Intelligence as a new scientific field, describes computational methods that are inspired by nature and biological systems and can combine scientific fields, such as Computational Mathematics and Computer Science, and it has several extensions and applications in almost all science.

Many and varied applications affecting several areas of research and technology arise or can be addressed as Optimization problems. Several methods of Computational Intelligence refer to a special class of optimization methods that involves the study and the creation of computational methods that effectively deal with difficult problems of real-world. These methods are very important as they approach optimization problems without limitations, which are more common in most problems. Nevertheless, due to the amount of applications, and in particular the development of new technologies, new problems that require solution and are related to different kinds of optimization, such as multi-objective or constrained optimization, have been arisen. To handle these difficult problems, Computational Intelligence methods are used, and in particular techniques and algorithms belonging to the field of Machine Learning and Artificial Neural Networks.

In the context of the above collaborative approach, it is necessary to study the prerequisites of these algorithms. Thus, before applying these techniques, it is necessary to prepare and preprocess the available data set in order to create reliable models with good performance and good generalization ability. Then, these methods can be applied to tackle problems in several fields, such as engineering, physics, biology, mathematics, medicine, computer science, industry, music, economics, cryptography etc.

Modern literature presents a huge number of algorithms and techniques related to the above directions. However, the majority of them are related to a specific framework that is tailored to the needs of the problem. The algorithms are usually limited to statistical models and parameters that satisfy a particular problem, thus excluding the successful application to any other problem or even to the same problem with different needs. A reliable solution to this major problem is provided by the development of algori-

thms with mathematical foundation and mathematical proof. An algorithm with a strong mathematical background can be used in many applications with greater success.

Towards this direction is based our thesis, that is the mathematical foundation of algorithmic techniques, which will address as many problems as possible and thus, will fill the existing gaps in the literature concerning Computational Intelligence and Computational Mathematics. We believe and expect that the results of our thesis will contribute to the creation of reliable methods that may, among the others, enable scientists in other scientific fields to reliably deal with large-scale and complex problems that often occur in many scientific, industrial and financial fields with obvious benefits.

Ευχαριστίες

Χάριν λαβών μέμνησο και δους επιλαθού.

— (4 . .)

Συνηθίζεται σε αυτή την ενότητα να διαβάζουμε λόγια τυποποιημένα, φορμαρισμένα και καλά προβαρισμένα. Από τη μεριά μου θα ήθελα να μοιραστώ λόγια καρδιάς, αληθινά και καθόλου ξύλινα. Θα προσπαθήσω να είμαι σύντομος, αν και θα είναι δύσκολο...

Το πόνημα αυτό θα ήταν αδύνατο να περατωθεί εάν δεν είχα στηριχθεί πάνω στους ώμους ενός γίγαντα—στα μάτια μου, τουλάχιστον, πάντα ως τέτοιος φάνταζε. Τα οφέλη και οι εμπειρίες που αποκόμισα δίπλα του είναι ανεκτίμητα. Για τον λόγο αυτό δεν μπορώ να γράψω «ευχαριστώ», καθώς αισθάνομαι, θα ήταν πολύ λίγο για να εκφράσει την ευγνωμοσύνη μου. Από τα φοιτητικά μου κιόλας χρόνια, με ενέμπνευσε και μου έδωσε κίνητρο να πετύχω τους στόχους μου. Σε όλη αυτή την πορεία είχαμε μια εξαιρετική συνεργασία, που με γέμισε με ποικίλες γνώσεις και με διαμόρφωσε σημαντικά σαν άνθρωπο. Τολμώ να πω, μια συνεργασία που ξεπέρασε τα στενά όρια μαθητή-δασκάλου και τις προσδοκίες που είχα. Δάσκαλε η τιμή ήταν τεράστια και η συνεισφορά σας καθοριστική για ό,τι έχω πετύχει στην ακαδημαϊκή μου πορεία. Χωρίς τη συμβολή σας «...θα κραύγαζα για βοήθεια στα όνειρά μου».

Ο αναπληρωτής Καθηγητής του Τμήματος κος Παναγιώτης Αλεβίζος, όποτε χρειάστηκε ήταν διαθέσιμος. Η πόρτα του γραφείου του ήταν πάντα ανοιχτή και η αισιοδοξία που τον χαρακτηρίζει, με επανέφερε στο σωστό δρόμο τη στιγμή που χρειαζόταν. Κύριε Τάκη σας ευχαριστώ πολύ!

Στην πορεία μου αυτή θεωρώ πως αποκόμισα έναν καλό συνεργάτη και φίλο, τον επίκουρο Καθηγητή του Τμήματος και μέλος της Τριμελούς Συμβουλευτικής Επιτροπής μου, κο Σωτήρη Κωτσιαντή. Η βοήθειά του σε πολλούς τομείς ήταν καταλυτική. Σας ευχαριστώ πολύ για τις συμβουλές, την υποστήριξη, την ενθάρρυνση και τις ωραίες συζητήσεις που είχαμε.

Θα ήταν παράλειψη η μη αναφορά στον συνεργάτη, φίλο και σύντεκνο υποψήφιο διδάκτορα κο Χρήστο Αριδά. Δίχως την πολύτιμη στήριξη και βοήθειά του είναι πολύ πιθανόν να μην τα είχα καταφέρει. Έστεκε πλάι μου σαν καλός σύμβουλος, μοιράστηκε τους προβληματισμούς μου και χάρισε απλόχερα τις πολλές και ποικίλες γνώσεις του. Μια σχέση έντονη και ιδιαίτερα επικοινωνιακή (όπως και ο δυνατός χαρακτήρας του), με στηγμάτισε. Μου έκανε τεράστια τιμή να συνεργαστεί μαζί μου, αλλά ακόμα περισσότερο να ανοίξει την πόρτα της οικογένειάς του. Χρήστο σε ευχαριστώ πολύ για όλα!

Συνοδοιπόρος και συμπαραστάτης σε όλη αυτή την πορεία στάθηκε ο συνάδελφος και φίλος υποψήφιος διδάκτορας του Τμήματος κος Γεώργιος Τεμπονέρας. Φίλε Τέμπο, μοιραστήκαμε μαζί δυσκολίες ακαδημαϊκές αλλά και εμπειρίες ζωής. Όποτε σε χρειάστηκα ήσουν δίπλα μου και έβαλες πλάτη σαν στρατιώτης στη μάχη. Το εκτιμώ απεριόριστα και ελπίζω κάποια στιγμή να έχω την ευκαιρία να ανταποδώσω τα καλά που απλόχερα εισέπραξα κατά τη διάρκεια της συνεργασίας μας.

Στο σημείο αυτό, ευχαριστώ θερμά τον υποψήφιο διδάκτορα του Τμήματος κο Εμμανουήλ Οικονομάκη για τις χρήσιμες συμβουλές του και την προθυμία που έδειξε όποτε τον χρειάστηκα. Μάνο σε ευχαριστώ πολύ για το «ξεμπλοκάρισμα». Οι κουβέντες σου μπορεί να ήταν μετρημένες αλλά υπήρξαν τρομερά εύστοχες. Ακόμα, ευχαριστώ θερμά όλα τα υπόλοιπα μέλη του Εργαστηρίου Υπολογιστικής Νοημοσύνης (τωρινά και παρελθόντα), για όλες τις

στιγμές που περάσαμε μαζί.

Ευχαριστώ μέσα από τα βάθη της καρδιάς μου τη μητέρα μου Παναγιώτα, για τη στήριξη, τα ξενύχτια της και την άπειρη αγάπη που εισέπραξα. Ομοίως, τον πατέρα μου Νίκο και ιδιαίτερα, για την ηθική και υλική του υποστήριξη και ενίσχυση, αλλά και το παράδειγμα που μου έδωσε σε πολύ δύσκολες καταστάσεις της ζωής μας. Ακόμα, είμαι ευγνώμων στον αδερφό μου Σωτήρη, που μέσα από τη ζωή του μου δίδαξε πώς είναι ένας πραγματικός αγωνιστής. Χάριν στην οικογένειά μου, κατάφερα να φτάσω ως εδώ...

Τέλος, θα ήθελα να ευχαριστήσω θερμά την αγαπημένη μου σύζυγο Λίνα, που μοιράστηκε όλες τις ανησυχίες, τους προβληματισμούς και τις άπειρες-αριθμησίμου δε πλήθους-δυσκολίες που συνάντησα όλο αυτό το διάστημα και τις έκανε πιο ανεκτές. Σίμπα δίχως εσένα θα ήμουν ακόμα στην αρχή! Με τη δική της βοήθεια, όπως και της Βίβιαν Παναγοπούλου και του Πανοσιολογιωτάτου Αρχιμανδρίτου π. Ίωακείμ Σταματόπουλου κατάφερα-πιστεύω-να αποφύγω το σημαντικότερό μου φόβο, όπως τον διατύπωσε ο Γερμανός φιλόσοφος Fridrich Nietzsche: «

μ μ , μ
μ μ .»

Σταμάτιος-Άγγελος Ν. Αλεξανδρόπουλος
Μάρτιος, 2020.

Περιεχόμενα

Περίληψη	vii
Ευχαριστίες	xi
Κατάλογος Σχημάτων	xv
Κατάλογος Πινάκων	xvii
I Εισαγωγή	1
1 Εισαγωγή	3
1.1 Υπολογιστική Νοημοσύνη	3
1.2 Υπολογιστικά Μαθηματικά	6
1.3 Μαθηματική Βελτιστοποίηση	7
1.4 Εξελικτικοί αλγόριθμοι και Μηχανική Μάθηση	9
1.5 Συνεισφορά	11
II Μαθηματική Βελτιστοποίηση	13
2 Δεν υπάρχει δωρεάν γεύμα για τη Μαθηματική Βελτιστοποίηση	15
2.1 Πρώιμες Εξελίξεις	16
2.2 Δεν υπάρχει δωρεάν γεύμα για τη βελτιστοποίηση και την αναζήτηση	19
2.3 Πρόσφατες εργασίες του Wolpert	22
2.4 Μη δωρεάν γεύμα για βελτιστοποίηση και εξελικτικούς αλγορίθμους	23
2.4.1 Μη δωρεάν γεύματα και εξελικτικοί αλγόριθμοι	23
2.4.2 Μη δωρεάν γεύματα και Μετα-ευρετικές Τεχνικές	26
2.5 Μη δωρεάν γεύματα στην Επιτηρούμενη Μάθηση	32
2.5.1 Δεν υπάρχει δωρεάν γεύμα για την έγκαιρη διακοπή	34
2.5.2 Δεν υπάρχει δωρεάν γεύμα για διασταυρούμενη επικύρωση	34
2.5.3 Ταξινόμηση προβλημάτων Μηχανικής Μάθησης του πραγματικού κόσμου και μη δωρεάν γεύματα: μια πειραματική προσέγγιση	36
2.6 Σύνοψη και τελικές παρατηρήσεις	37
3 Μέθοδοι δυναμικών τροχιών αναζήτησης για ολική βελτιστοποίηση	39
3.1 Γνωστές μέθοδοι τροχιάς και εφαρμογές τους	41
3.1.1 Μέθοδοι δυναμικών τροχιών ανίχνευσης	41
3.1.2 Εφαρμογές μεθόδων δυναμικών τροχιών ανίχνευσης	52
3.2 Πρόσφατες μέθοδοι δυναμικών τροχιών και εφαρμογές αυτών	62
3.3 Οικογένειες μεθόδων δυναμικής τροχιάς που προέρχονται από τις μεθόδους Runge-Kutta	65
3.4 Σύνοψη και μελλοντικές ερευνητικές επιδιώξεις	78

III Μηχανική Μάθηση	79
4 Προεπεξεργασία δεδομένων στην Εξόρυξη Δεδομένων	81
4.1 Μια σύντομη ανασκόπηση	81
4.2 Εντοπισμός θορύβου και ακραίων τιμών	84
4.3 Ελλειπείς τιμές χαρακτηριστικών	92
4.4 Κανονικοποίηση και διακριτοποίηση	97
4.5 Επιλογή παραδειγμάτων	104
4.6 Επιλογή χαρακτηριστικών	111
4.7 Σύνοψη και παρατηρήσεις	117
5 Εντοπισμός και αναγνώριση ακραίων τιμών	121
5.1 Μια σύντομη επισκόπηση	121
5.2 Γνωστές τεχνικές: Μια εμπειρική κατηγοριοποίηση	123
5.2.1 Μέθοδοι που βασίζονται στον πλησιέστερο γείτονα	123
5.2.2 Πιθανοτικές μέθοδοι	123
5.2.3 Μέθοδοι που βασίζονται στον υπόχωρο	124
5.2.4 Συνδυαστικές μέθοδοι	124
5.2.5 Μέθοδοι που βασίζονται στον ταξινομητή	125
5.2.6 Μέθοδοι που βασίζονται στη συσταδοποίηση	125
5.3 Προτεινόμενη μεθοδολογία και πειραματικά αποτελέσματα	125
5.4 Συμπεράσματα - Μελλοντικές μελέτες	127
6 Συνεργατικές μέθοδοι για προβλήματα ταξινόμησης	133
6.1 Γνωστές τεχνικές σχεδιασμού συνεργατικών μεθόδων	134
6.2 Προτεινόμενο υβριδικό σχήμα ταξινόμησης	137
6.3 Συγκρίσεις και πειραματικά αποτελέσματα	138
6.4 Σύνοψη	140
IV Βελτιστοποίηση στη Μηχανική Μάθηση	145
7 Εξελικτικοί Αλγόριθμοι πολυ-αντικειμενικής βελτιστοποίησης στη Μηχανική Μάθηση	147
7.1 Βασικές έννοιες της βελτιστοποίησης πολλαπλών αντικειμένων	150
7.2 Προεπεξεργασία δεδομένων	151
7.2.1 Επιλογή χαρακτηριστικών	151
7.2.2 Επιλογή παραδειγμάτων	153
7.2.3 Ελλιπή δεδομένα	154
7.2.4 Διακριτοποίηση	154
7.2.5 Μη ισορροπημένα σύνολα δεδομένων	154
7.3 Επιτηρούμενη μάθηση	155
7.3.1 Δέντρα απόφασης	155
7.3.2 Κανόνες μάθησης	156
7.3.3 Bayesian ταξινομητές	156
7.3.4 Μηχανές διανυσματικής υποστήριξης	157
7.3.5 Τεχνητά νευρωνικά δίκτυα	157
7.3.6 Lazy Learners	158
7.3.7 Συνεργατικές μέθοδοι	158
7.4 Μη επιτηρούμενη μάθηση	159
7.4.1 Συσταδοποίηση	159
7.4.2 Κανόνες συσχέτισης	160
7.5 Πρόσφατες εφαρμογές	161

7.6 Σύνοψη	161
V Συμπεράσματα–Βιβλιογραφία–Δημοσιεύσεις	163
8 Σύνοψη	165
8.1 Σύνοψη και συμπεράσματα	165
8.2 Μελλοντικές προεκτάσεις	168
Βιβλιογραφία	171

Κατάλογος Σχημάτων

4.1	Μοντελοποίηση Προγνωστικής διαδικασίας	83
5.1	Ένα απλοποιημένο επεξηγηματικό παράδειγμα	126
6.1	Διαδικασία δημιουργίας συνεργατικής μεθόδου με βάση τη μεθοδολογία στεί- βαξης	139

Κατάλογος Πινάκων

2.1	Αριθμός αναφορών που συγκεντρώνουν οι εργασίες που σχετίζονται με τα θεωρήματα NFL και ζητήματα βελτιστοποίησης και εξελικτικών αλγορίθμων όπως παρουσιάζονται στην Ενότητα 2.4. Όπου «ΣΠΑ» δηλώνει το συνολικό πλήθος αναφορών που έλαβε ένα άρθρο, ενώ «ΑαΕ» συμβολίζει κατά μέσο όρο τις αναφορές που λαμβάνει ανά έτος κάθε εργασία.	30
3.1	Σύντομη περιγραφή, εφαρμοσιμότητα και αριθμός αναφορών σχετικά με τις μεθόδους δυναμικής τροχιάς που παρουσιάστηκαν στην Ενότητα 3.1 και Ενότητα 3.2. Όπου «ΣΠΑ» υποδηλώνει το σύνολο των αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.	66
3.2	Συγκρίσεις - Αποτελέσματα	72
3.3	Συγκρίσεις - Αποτελέσματα	73
3.4	Συγκρίσεις - Αποτελέσματα	74
3.5	Συγκρίσεις - Αποτελέσματα	75
3.6	Συγκρίσεις - Αποτελέσματα	76
3.7	Συγκρίσεις - Αποτελέσματα	77
4.1	Συνοπτική περιγραφή, δυνατότητα ανίχνευσης θορύβου και ακραίων τιμών, αριθμός των αναφορών που σχετίζονται με τα θέματα ανίχνευσης θορύβου και ακραίων τιμών που παρουσιάζονται στην Ενότητα 4.2. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.	91
4.2	Συνοπτική περιγραφή και αριθμός των αναφορών που σχετίζονται με τα θέματα ελλειπουσών τιμών χαρακτηριστικών που παρουσιάζονται στην Ενότητα 4.3. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.	96
4.3	Σύντομη περιγραφή και αριθμός των αναφορών που σχετίζονται με τα θέματα κανονικοποίησης και διακριτοποίησης που παρουσιάζονται στην Ενότητα 4.4. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.	103
4.4	Σύντομη περιγραφή και αριθμός αναφορών που σχετίζονται με τα θέματα επιλογής παραδειγμάτων, όπως παρουσιάζονται στην ενότητα 4.5. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.	110
4.5	Σύντομη περιγραφή και αριθμός αναφορών, που σχετίζονται με τα θέματα επιλογής χαρακτηριστικών όπως παρουσιάζονται στην ενότητα 4.6. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.	116
4.6	Εφαρμογή/χρησιμότητα των διαδικασιών προεπεξεργασίας σχετικά με τους γνωστούς μαθητές. Τα 'LMs' υποδηλώνουν τα γραμμικά μοντέλα, το 'DTs' υποδεικνύει τα δέντρα απόφασης, το 'ANN' υποδεικνύει τα τεχνητά νευρωνικά δίκτυα, το 'RLs' υποδηλώνει τους μαθητές κανόνων, τα 'SVM' χαρακτηρίζουν τις μηχανές διανυσματικής υποστήριξης και τα 'LLs' υποδηλώνουν τους «τεμπέληδες» μαθητές.	119

5.1	Receiver Operating Characteristics (ROC) curves	128
5.2	Precision-Recall Curve (PRC)	129
5.3	Κατάταξη των αλγορίθμων χρησιμοποιώντας το Friedman test (Using ROC) . .	130
5.4	Post-hoc Bonferroni-Dunn (Using Proposed as control method - Using ROC)	130
5.5	Κατάταξη των αλγορίθμων χρησιμοποιώντας το Friedman test (Using PRC) . .	131
5.6	Post-hoc Bonferroni-Dunn (Using Proposed as control method - Using PRC)	131
6.3	Κατάταξη των αλγορίθμων χρησιμοποιώντας το Friedman test	139
6.4	Post-hoc Holm test using Stacking as control method	140
6.1	Συλλογή 38 προβλημάτων πολλαπλών κλάσεων από το UCI Machine Learning Repository. Στον πίνακα παρουσιάζεται ο αριθμός των παραδειγμάτων, ο αριθμός των χαρακτηριστικών, καθώς και ο αριθμός των κλάσεων.	141
6.2	Ακρίβεια ταξινόμησης και τυπική απόκλιση των συγκρινόμενων μεθόδων, όπου το 'ET' δηλώνει τον αλγόριθμο ExtraTrees, το 'GB' αντιπροσωπεύει τη μέθοδο Gradient Boosting, ενώ το 'RF' δηλώνει τον αλγόριθμο Random Forest. . . .	142

Μέρος Ι

Εισαγωγή

Εισαγωγή

Τας τύχας εκ των πόνων εράν.

— (480–406 . . .)

Βασικό αντικείμενο της διατριβής είναι η ανάπτυξη και θεμελίωση νέων μεθόδων Υπολογιστικών Μαθηματικών που συσχετίζονται με την Υπολογιστική Νοημοσύνη. Οι υπολογιστικές μέθοδοι και τα μοντέλα που περιλαμβάνονται στην θεματολογία της Υπολογιστικής Νοημοσύνης και έχουν τις ρίζες τους σε μεθόδους Υπολογιστικών Μαθηματικών (Αριθμητικής Ανάλυσης, Αριθμητικής Επίλυσης Συστημάτων Μη-Γραμμικών Εξισώσεων, Αριθμητικής Επίλυσης Διαφορικών Εξισώσεων, Αριθμητικής Βελτιστοποίησης και Μη-Γραμμικού Προγραμματισμού κ.ά.), θα αποτελέσουν τη βάση αυτής της διατριβής.

1.1 Υπολογιστική Νοημοσύνη

Η « μ » (Computational Intelligence) [140] είναι ένας σχετικά νέος επιστημονικός κλάδος, ο οποίος περιγράφει τις υπολογιστικές μεθόδους που χρησιμοποιούν ιδέες εμπνευσμένες από φυσικά και βιολογικά συστήματα (ο λεγόμενος « μ » (Natural Computing) [82]). Το γεγονός ότι συνδυάζει επιστημονικούς τομείς, όπως τα Υπολογιστικά Μαθηματικά, την Επιστήμη των Υπολογιστών, την Υπολογιστική Βιολογία, κ.ά., έχοντας επεκτάσεις και εφαρμογές σχεδόν σε όλους τους τομείς των θετικών επιστημών, καταδεικνύει με emphaticό τρόπο το λόγο για τον οποίο το ενδιαφέρον της επιστημονικής κοινότητας συγκεντρώνεται γύρω από το συγκεκριμένο τομέα της έρευνας.

Ο αυστηρός ορισμός ερευνητικών πεδίων όπως αυτό της Υπολογιστικής Νοημοσύνης, είναι ένα δύσκολο έργο, καθώς θα πρέπει να «περιορίσουμε» έννοιες που από τη φύση τους είναι αφηρημένες ή έως και «αόριστες», όπως η έννοια και ιδέα της «νοημοσύνης». Για τον λόγο αυτό, οι ειδικοί συχνά αποφεύγουν να ορίσουν τέτοια επιστημονικά πεδία, και έτσι, σπάνια ο αναγνώστης συναντά κάποιο «φορμαλιστικό» ορισμό στη βιβλιογραφία. Ακόμα, στις περιπτώσεις που μπορεί να συμβεί κάτι τέτοιο, οι ορισμοί αυτοί δύνανται να μηπεδέσουν τον αναγνώστη, οπότε η λειτουργική τους αξία θα μπορούσαμε να ισχυριστούμε πως είναι «φτωχή». Ωστόσο, σε αυτή την ενότητα θα προσπαθήσουμε να δώσουμε «ανεπίσημα» κάποιους ορισμούς του επιστημονικού πεδίου της Υπολογιστικής Νοημοσύνης ή θα αναφερθούμε σε ορισμένες αυθεντίες του χώρου, που έδωσαν αντιπροσωπευτικούς, ως προς το περιεχόμενο, ορισμούς. Έτσι, η έννοια της νοημοσύνης συνδέεται σε πολλές περιπτώσεις με την ικανότητα που διαθέτει κάποιος να ανταπεξέλθει με επιτυχία σε πολύ δύσκολα προβλήματα. Όπως γίνεται εύκολα κατανοητό, ήδη υπεισέρχεται μια έννοια που χαρακτηρίζεται ως «υποκειμενική». Με άλλα λόγια, πώς αντιλαμβάνεται κάποιος ένα «δύσκολο πρόβλημα», ή ποιο πρόβλημα αποφασίζεται πως είναι «δύσκολο» και από ποιον [63]. Επιπλέον, η «νοημοσύνη» συνδέεται επίσης, με το πόσο έγκαιρα, αξιόπιστα ή/και «έξυπνα» λαμβάνει κάποιος μια απόφαση ή ανταποκρίνεται σε ένα πρόβλημα. Σε αυτό το σημείο όμως, αναφέραμε μια

ακόμα υποκειμενική έννοια, αυτή του τι θεωρεί κάποιος ως «έξυπνο». Η έννοια αυτή, συχνά συγχέεται με την ιδέα της «ευφυΐας». Κατ' επέκταση, επιζητούμε υπολογιστικά συστήματα τα οποία θα λαμβάνουν αποφάσεις έγκαιρα, αξιόπιστα, αποδοτικά και αποτελεσματικά, σύμφωνα με τα διαθέσιμα ερεθίσματα, τα οποία λαμβάνουν από το περιβάλλον τους. Συνεπώς, θα είναι σε θέση να εκτελέσουν συγκεκριμένες διεργασίες και να προσεγγίσουν προβλήματα συγκεκριμένου σκοπού.

Όπως αναφέραμε πιο πάνω, οι περισσότερες έννοιες που συναντούνται στον τομέα της Υπολογιστικής Νοημοσύνης έχουν εμπνευστεί από τη φύση. Στη φύση, οι πρωταρχικοί στόχοι των ζώντων οργανισμών είναι να επιβιώσουν, να επιλέξουν κατάλληλα ταίρια, να αναπαραχθούν, έτσι ώστε να δημιουργήσουν νέους απογόνους και συνεπώς, να εξελίξουν τον κύκλο της ζωής, όπως μαθαίνουμε από την εξελικτική διαδικασία της βιολογίας. Στην πορεία αυτή της ζωής, οι πιο ισχυροί οργανισμοί επιβιώνουν, ενώ οι πιο αδύναμοι δεν τα καταφέρνουν και τελικά εξαλείφονται. Ως αδύναμος, χαρακτηρίζεται ένας ζωντανός οργανισμός που δε διαθέτει τα κατάλληλα χαρακτηριστικά ώστε να επιβιώσει και να αναπαραχθεί, οπότε το ίδιο το περιβάλλον του τον «αποβάλλει». Η παραπάνω φιλοσοφία αναπτύσσεται και επαναλαμβάνεται εδώ και χιλιάδες χρόνια, με αποτέλεσμα να έχουμε πολλές και διαφορετικές γενιές ζώντων οργανισμών με παρόμοια χαρακτηριστικά. Σε αυτό το σημείο, δεν θα θέλαμε ο αναγνώστης να συνδυάσει την έννοια της νοημοσύνης μόνο με τους ζωντανούς οργανισμούς. Για το σκοπό αυτό, παραθέτουμε τον ακόλουθο ορισμό για τη νοημοσύνη [63]:

« μ , μ , μ .»

Χρονολογικά, η δημιουργία της Υπολογιστικής Νοημοσύνης τοποθετείται το 1994 στο μ . . . Ο βασικός σκοπός αυτού του επιστημονικού πεδίου είναι η δημιουργία υπολογιστικών συστημάτων, τα οποία θα είναι «εξοπλισμένα» με τις ικανότητες μ , μ , - , κ.ά., ώστε κατά κάποιον τρόπο να μπορούν να εφαρμοστούν για την προσέγγιση προβλημάτων του πραγματικού κόσμου, με «έξυπνο» τρόπο. Σύμφωνα με τα παραπάνω, η «Υπολογιστική Νοημοσύνη» ορίζεται ως η ικανότητα που δύναται να αποκτήσει οποιαδήποτε μηχανή, να αποκρίνεται σε ερεθίσματα που λαμβάνει από το περιβάλλον, παίρνοντας αποφάσεις με «έξυπνο» τρόπο για κάθε νέο ερέθισμα, λαμβάνοντας ωστόσο υπόψιν τις γνώσεις που έχει αποκτήσει από προηγούμενα ερεθίσματα. Στο πεδίο συναντώνται οι ακόλουθοι τομείς :

- (α) μ (Evolutionary Computation),
- (β) (Artificial Neural Networks),
- (γ) μ (Fuzzy Systems),
- (δ) ή/και οποιοσδήποτε συνδυασμός αυτών.

Συνεπώς, η Υπολογιστική Νοημοσύνη είναι υπεύθυνη για το «νου» των σύγχρονων υπολογιστικών μηχανών.

Παρά το γεγονός ότι ως επιστημονικός τομέας συγκεντρώνει το ενδιαφέρον της επιστημονικής κοινότητας εδώ και αρκετά χρόνια, και επιπλέον συναντάται στη βιβλιογραφία τεράστιο πλήθος επιστημονικών δημοσιεύσεων, ακόμα και σήμερα διατυπώνονται διαφορετικές ερμηνείες, χωρίς να υπάρχει μία που να είναι αποδεκτή. Αυτό που γενικά είναι αποδεκτό από την επιστημονική κοινότητα, είναι ότι η Υπολογιστική Νοημοσύνη αντιμετωπίζεται ως ένα «υπερσύνολο», το οποίο με την πάροδο των χρόνων συμπεριλαμβάνει όλο και περισσότερες μεθόδους, τεχνικές, υβριδικά σχήματα και συνδυασμούς που δύνανται να χαρακτηριστούν

ως «ευφείς». Επιπροσθέτως, εξαιτίας της ενασχόλησης όλο και περισσότερων επιστημόνων με τον κλάδο, από ποικίλα ερευνητικά πεδία, η προσθαφαίρεση μεθόδων σε αυτό το υπερ-σύνολο είναι διαρκής. Αυτό, μπορεί να δικαιολογήσει σε ένα βαθμό και το λόγο για τον οποίο δεν έχει επικρατήσει ένας ενιαίος και αποδεκτός ορισμός για την Υπολογιστική Νοημοσύνη.

Ακολουθώντας, θεωρούμε χρήσιμο για τον αναγνώστη να παραθέσουμε τους επικρατέστερους ορισμούς που συναντώνται στη βιβλιογραφία [35, 67, 111, 324]. Συγκεκριμένα, ο Bezdek αναφέρει [35]:

«... ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα. Ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα. Ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα.»

Οι Eberhart et al. αναφέρουν [111]:

«... ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα. Ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα. Ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα.»

Παραθέτουμε δύο ακόμα αντιπροσωπευτικούς ορισμούς, όπως ορίστηκαν [67] και [324]:

«... ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα.»

και

«... ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα. Ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα.»

Τέλος ο Fogel στην εργασία [136] ανέφερε το εξής:

«... ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα. Ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα. Ο υπολογιστής, με τη βοήθεια της τεχνητής νοημοσύνης, μπορεί να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος αντιμετωπίζει τα προβλήματα.»

Μπορεί οι ορισμοί που έχουν σημειωθεί όλα αυτά τα χρόνια να είναι πολλοί και ίσως θα ήταν ελαφρώς αποπροσανατολιστικό να σταθούμε περαιτέρω σε αυτό το θέμα. Αυτό που έχει μεγάλη σημασία είναι να συμφωνήσουμε στα χαρακτηριστικά και τις ιδιότητες που πρέπει να διέπουν τις μεθόδους που εντάσσονται στο πλαίσιο της Υπολογιστικής Νοημοσύνης. Έτσι,

επιζητούμε οι μέθοδοι και οι τεχνικές που ανήκουν σε αυτό το πεδίο, να έχουν ισχυρές δυνατότητες προσέγγισης εργασιών της πραγματικής ζωής, την ιδιότητα να αποκτούν γνώση από το περιβάλλον στο οποίο ανήκουν, να μπορούν να αυτο-οργανωθούν σε περιβάλλοντα με διαφορετικές συνθήκες και να ανασύρουν τη γνώση που έχουν αποκτήσει για άγνωστα δεδομένα.

Μια μέθοδος με αυτά τα χαρακτηριστικά δύναται να αντιμετωπίσει απαιτητικά προβλήματα που συναντώνται στην πραγματική ζωή, τα οποία δε μπορούν να προσεγγιστούν με αναλυτικές μεθόδους ή άλλες τεχνικές. Οι εφαρμογές στις οποίες συναντάμε τέτοιου είδους τεχνικές είναι πολλές και άπτονται της Μηχανικής, της Φυσικής, της Βιολογίας, της Βιομηχανίας, των Μαθηματικών, της Ιατρικής και πολλών άλλων επιστημονικών τομέων. Ακόμα, αξίζει να σημειώσουμε ότι συχνά παρατηρούνται μέθοδοι στις τομές των παραπάνω επιστημονικών πεδίων, οι οποίες, όταν αξιοποιούν γνώση από διαφορετικά πεδία, συνηθίζεται να χαρακτηρίζονται ως «υβριδικές».

Οι κεντρικοί άξονες και οι τεχνικές που παρουσιάζονται στην παρούσα Διδακτορική Διατριβή άπτονται της θεματολογίας της Υπολογιστικής Νοημοσύνης και όχι μόνο. Συγκεκριμένα, αναπτύσσονται νέες, καινοτόμες μεθοδολογίες ολικής Βελτιστοποίησης, που ανήκουν κυρίως στην κατηγορία της κλασικής Βελτιστοποίησης, και προκύπτουν μετά από το συνδυασμό τους με μεθόδους των Υπολογιστικών Μαθηματικών και συγκεκριμένα μεθόδων αριθμητικής επίλυσης συνήθων διαφορικών εξισώσεων. Αυτή η νέα οικογένεια μεθόδων που προτείνεται, δύναται να εφαρμοστεί σε δύσκολα προβλήματα που, μεταξύ άλλων, περιλαμβάνουν την αποτελεσματική εκπαίδευση τεχνητών νευρωνικών δικτύων ή/και να συνδυαστούν με μεθόδους που ανήκουν στην οικογένεια του Εξελικτικού Υπολογισμού.

Προκειμένου να δοθεί μία ολοκληρωμένη περιγραφή αυτών στις παρακάτω ενότητες, παρουσιάζονται συνοπτικά οι κύριες συνιστώσες της Υπολογιστικής Νοημοσύνης, των Υπολογιστικών Μαθηματικών όπως και της μαθηματικής Βελτιστοποίησης. Ιδιαίτερη έμφαση δίνεται στην παρουσίαση των δυνατοτήτων των Υπολογιστικών Μαθηματικών, των αλγορίθμων Εξελικτικού Υπολογισμού, αλλά και στο πως μέσα από τη συνεργεία αυτών των κλάδων, μπορούν να δημιουργηθούν τομές για την ανάπτυξη και θεμελίωση νέων μεθόδων που θα εφαρμόζονται σε διαφορετικούς τομείς της Υπολογιστικής Νοημοσύνης.

1.2 Υπολογιστικά Μαθηματικά

Τα Υπολογιστικά Μαθηματικά και οι αριθμητικές μέθοδοι κατά μια έννοια καταπιάνονται με τους αριθμούς και ευθύνονται σημαντικά για την προσφορά προσεγγίσεων όταν η αναλυτική λύση ενός προβλήματος είναι αδύνατο να βρεθεί. Οι αριθμητικοί υπολογισμοί συναντώνται από τα αρχαία κιόλας χρόνια σε τομείς της ανάλυσης, της αστρονομίας και των κατασκευών. Κατά το 1845, ο τομέας εδραιώνεται και κατά τα τέλη του 19ου αιώνα ισχυροποιείται με τις πρώτες υπολογιστικές μηχανές.

Η ανάπτυξη των Ηλεκτρονικών Υπολογιστών (H/Y) κατά το τέλος του δεύτερου παγκοσμίου πολέμου, συνέβαλε σημαντικά στη ραγδαία εξέλιξη των Υπολογιστικών μεθόδων και την ανάπτυξη των πρώτων αλγορίθμων. Έτσι, όπως γίνεται εύκολα αντιληπτό ο τομέας των Υπολογιστικών Μαθηματικών είναι άμεσα συνιφασμένος με τους H/Y. Η σημαντικότητά τους έγκειται σε πολλούς παράγοντες, τους οποίους θα προσπαθήσουμε ενδεικτικά να παρουσιάσουμε στη συνέχεια αυτού του κεφαλαίου, πριν προχωρήσουμε στο κυρίως σώμα της διατριβής.

Πολλά και δύσκολα προβλήματα του πραγματικού κόσμου μπορούν να μετασχηματιστούν κατά τέτοιο τρόπο ώστε να δύνανται να επεξεργαστούν με τη βοήθεια ενός H/Y. Τα προβλήματα αυτά μπορούν να προσεγγιστούν αριθμητικά και άπτονται πολλών γνωστών πεδίων, όπως η επίλυση συστημάτων μη γραμμικών αλγεβρικών ή/και υπερβατικών εξισώσεων, η αριθμητική παραγωγή και ολοκλήρωση, η επίλυση προβλημάτων αρχικών τιμών, η βελτιστοποίηση μιας αντικειμενικής συνάρτησης κ.ά. Συμπερασματικά, πρόκειται για έναν επι-

στημονικό κλάδο που είναι χρήσιμος, ή και κατά περιπτώσεις απαραίτητος σε κλάδους των εφαρμοσμένων μαθηματικών, της πληροφορικής, της επιχειρησιακής έρευνας, της εξόρυξης δεδομένων, της κρυπτογραφίας κ.ά.

Για την προσέγγιση των προβλημάτων επιζητούνται οι κατάλληλες αριθμητικές μέθοδοι. Με τον όρο «κατάλληλη» αναφερόμαστε σε μια μέθοδο που μπορεί να μας εξασφαλίσει με σιγουριά αλλά και με δεδομένη ακρίβεια το επιθυμητό αποτέλεσμα, συναρτήσει δύο πολύ σημαντικών παραγόντων: (α) το μικρότερο δυνατό υπολογιστικό κόστος και (β) το μικρότερο δυνατό χώρο αποθήκευσης (μνήμη). Σύμφωνα με τα παραπάνω, η αναφορά σε αριθμητικές μεθόδους είναι ταυτόσημη με την ύπαρξη αλγορίθμων, δηλαδή μια πεπερασμένη ακολουθία βημάτων που για συγκεκριμένη είσοδο, μετά από συγκεκριμένο αριθμό επαναλήψεων δύναται να μας επιστρέψει ένα ικανοποιητικό αποτέλεσμα. Οι μέθοδοι αυτοί τεκμηριώνονται θεωρητικά και βρίσκουν πρακτική εφαρμογή μέσω της υλοποίησής τους από έναν Η/Υ.

Το επιθυμητό σενάριο θα ήταν μία μόνο μέθοδος να μπορούσε να ανταποκριθεί αξιόπιστα και αποτελεσματικά για το σύνολο των προβλημάτων που θα μπορούσαμε να συναντήσουμε. Κάτι τέτοιο όμως δε μπορεί να συμβεί. Συγκεκριμένα, κάθε μέθοδος διακρίνεται για μια σειρά πλεονεκτημάτων και μειονεκτημάτων. Έτσι, αφού μελετήσουμε καλά τους πόρους και τα ζητούμενα του εκάστοτε προβλήματος είμαστε σε θέση να επιλέξουμε την κατάλληλη μέθοδο για την προσέγγιση ενός συγκεκριμένου προβλήματος. Συμπερασματικά, με τη βοήθεια των Υπολογιστικών Μαθηματικών έχουμε τη δυνατότητα μέσω αλγορίθμων να προσεγγίσουμε μεγάλο πλήθος προβλημάτων που δεν επιδέχονται αναλυτική λύση (δεν υπάρχουν κλειστοί μαθηματικοί τύποι για την επίλυση του προβλήματος) και βρίσκουν εφαρμογές στην επιστήμη και την τεχνολογία. Ένα επιπλέον, σημαντικό πλεονέκτημα, είναι η ανάπτυξη των υπολογιστικών εργαλείων ή/και γλωσσών προγραμματισμού ή/και υπολογιστικών λογισμικών, όπως η γλώσσα προγραμματισμού Python και Fortran, αλλά και τα προγραμματιστικά περιβάλλοντα MATLAB, Octave, Mathematica κ.ά, τα οποία εύκολα και γρήγορα δίνουν τη δυνατότητα – ακόμα και σε μη έμπειρους χρήστες – να προσεγγίσουν δύσκολα προβλήματα, με πλήθος βιβλιοθηκών. Ακόμα, ένα τεράστιο μέρος της κοινότητας, συμβάλλει σε forums και γνωστές πλατφορμες όπως το GitHub, GitLab, StackOverflow κ.ά.

1.3 Μαθηματική Βελτιστοποίηση

Σε αρκετούς τομείς της επιστήμης και της τεχνολογίας, εμφανίζονται πολλές σημαντικές εφαρμογές [228,397] που μπορούν να αντιμετωπιστούν ως προβλήματα Βελτιστοποίησης [92]. Οι μέθοδοι Εξελικτικού Υπολογισμού, αποτελούν μια ειδική κλάση μεθόδων βελτιστοποίησης που περιλαμβάνει τη μελέτη και τη θεμελίωση υπολογιστικών μεθόδων, οι οποίες αντιμετωπίζουν αποτελεσματικά δύσκολα προβλήματα που άπτονται σε διαφορετικούς τομείς [110]. Η σπουδαιότητά τους έγκειται στο γεγονός ότι αντιμετωπίζουν αποτελεσματικά προβλήματα βελτιστοποίησης χωρίς περιορισμούς, κάτι που απαιτούν τα περισσότερα προβλήματα του πραγματικού κόσμου [211].

Το πρόβλημα που αντιμετωπίζεται δυσκολότερα είναι αυτό της ολικής βελτιστοποίησης, καθώς δεν υπάρχει μια γενική «συνταγή» την οποία μπορούμε να ακολουθήσουμε για να εντοπίσουμε ένα ολικό βέλτιστο της υπό εξέταση συνάρτησης. Κάθε αλγόριθμος που εφαρμόζουμε σε ένα πρόβλημα βελτιστοποίησης μπορεί να συναντήσει πολλά τοπικά βέλτιστα στην προσπάθειά του να «κινηθεί» προς το ολικό βέλτιστο. Αυτό επιβαρύνει τον αλγόριθμο με μια τρομερά δύσκολη απόφαση, αυτή της απόφασης για το αν έχει καταλήξει σε ένα τοπικό ή ολικό βέλτιστο, προκειμένου να συνεχίσει να εκτελείται ή αντιστοίχως προκειμένου να τερματιστεί.

Οι αλγόριθμοι ολικής βελτιστοποίησης χρειάζεται να διαιρεθούν σε δύο μη-επικαλυπτόμενες ομάδες σε σχέση με την ακρίβειά τους [384]. Συγκεκριμένα, υπάρχουν οικογένειες μεθόδων που έχουν ακρίβεια και οικογένειες που δεν έχουν εγγυημένη ακρίβεια. Οι μέθοδοι της δεύτερης κατηγορίας, αναμένεται να έχουν εκτελέσει μια εξαντλητική ανα-

ζήτηση στο χώρο των λύσεων. Κάτι τέτοιο όμως είναι δύσκολο να συμβεί, καθώς απαιτούνται *a priori* γνώσεις και πληροφορίες για το πρόβλημα, κάτι που δε συμβαίνει σε αρκετές των περιπτώσεων. Επιπλέον, ένα ακόμα μειονέκτημα αυτής της κατηγορίας είναι το μεγάλο υπολογιστικό κόστος, καθώς στην περίπτωση εύρεσης του ολικού βέλτιστου απαιτούνται εξαντλητικές επαναλήψεις του επαναληπτικού σχήματος. Εξαιτίας αυτής της διαφοράς, δεν είναι «πρέπον» να συγκρίνονται οι αποδόσεις αλγορίθμων που ανήκουν σε διαφορετικές κατηγορίες. Έτσι, οι αλγόριθμοι ολικής βελτιστοποίησης μπορούν να χωριστούν στις παρακάτω κατηγορίες:

- (α) μ (Methods with guaranteed accuracy). Οι πιο απλές μέθοδοι αυτής της κατηγορίας, οι μ (covering methods) στηρίζονται στον εντοπισμό των υπο-διαστημάτων, στα οποία δεν περιέχεται κάποιο ολικό ελάχιστο. Ακολουθώντας, αποκλείουν αυτές τις περιοχές και ο αλγόριθμος δε χρειάζεται να τις εξετάσει. Αυτό μπορεί να επιτευχθεί μέσω μίας τοπικής διαδικασίας, ή μίας διαδικασίας υποδιαίρεσης ή διχοτόμησης η οποία καταλήγει στην εύρεση μικρότερων υπο-περιοχών (με συγκεκριμένη ακρίβεια) γύρω από την περιοχή του ολικού βελτιστοποιητή.
- (β) μ (Direct Methods). Στη συγκεκριμένη οικογένεια μεθόδων συναντάμε τρεις υποκατηγορίες:
1. Στην πρώτη υποκατηγορία συναντώνται οι μ μ μ (generalized descent methods), που αξιοποιούν τα πλεονεκτήματα των μεθόδων τοπικής αναζήτησης. Οι πιο γνωστές μέθοδοι είναι οι μ (trajectory methods) (αναλύονται διεξοδικά στο Κεφάλαιο 3, με την ανάπτυξη μιας νέας οικογένειας μεθόδων Βελτιστοποίησης, όπως εμπνεύστηκε από τις δυναμικές τροχιές ανίχνευσης των Snyman-Fatti και τις μεθόδους Runge-Kutta για την αριθμητική επίλυση συνήθων διαφορικών εξισώσεων), που χρησιμοποιούν τροχιές δυναμικών συστημάτων για να επισκεφτούν όλα τα τοπικά ακρότατα της αντικειμενικής συνάρτησης.
 2. Η δεύτερη υποκατηγορία περιλαμβάνει τις μεθόδους (clustering), οι οποίες εκτελούν αρχικά μία δειγματοληψία του χώρου αναζήτησης και στη συνέχεια, συγκεντρώνουν τα δείγματα σε συστάδες. Ακολουθώντας, σε κάθε συστάδα επιλέγεται ένα αντιπροσωπευτικό δείγμα και εφαρμόζεται μια μέθοδος τοπικής αναζήτησης.
 3. Στην τρίτη υποκατηγορία ανήκουν οι μ (random search methods), οι οποίες με τη σειρά τους διακρίνονται σε δύο υποκατηγορίες. Η πρώτη περιέχει μεθόδους που δεν χρησιμοποιούν πληροφορίες από προηγούμενα δείγματα. Γνωστές μέθοδοι αυτών αποτελούν οι μ Monte Carlo και οι αλγόριθμοι (multi-start methods). Στην άλλη κατηγορία ανήκουν μέθοδοι που χρησιμοποιούν πληροφορίες από προηγούμενα δείγματα. Γνωστές μέθοδοι αυτής της υποκατηγορίας αποτελούν ο αλγόριθμος Simulated Annealing, όπως και οι αλγόριθμοι μ (Evolutionary Computing).
- (γ) $\mu\mu$ (Indirect Methods). Οι μέθοδοι αυτής της κατηγορίας αξιοποιούν τοπικές πληροφορίες, όπως τιμές της αντικειμενικής συνάρτησης, για τη δημιουργία ενός μοντέλου προσέγγισης της συνάρτησης. Βασικός στόχος αυτών των μοντέλων αποτελεί η καθοδήγηση της διαδικασίας αναζήτησης, ελαττώνοντας τον αριθμό των πραγματικών υπολογισμών της αντικειμενικής συνάρτησης.

Στο σημείο αυτό, δεν θα δώσουμε περισσότερες λεπτομέρειες, καθώς ο αναγνώστης θα βρει μια λεπτομερή ανάλυση για το πρόβλημα της Βελτιστοποίησης στο Κεφάλαιο 2. Στη συνέχεια

της Εισαγωγής, θα εξετάσουμε το πεδίο των Εξελικτικών Αλγορίθμων και της Μηχανικής Μάθησης.

1.4 Εξελικτικοί αλγόριθμοι και Μηχανική Μάθηση

Οι μ (Evolutionary algorithms) είναι στοχαστικοί αλγόριθμοι αναζήτησης, εμπνευσμένοι από τη διαδικασία της φυσικής εξέλιξης. Το κίνητρο για την εφαρμογή των Εξελικτικών Αλγορίθμων είναι σημαντικό, εφόσον αποτελούν μια αποδοτική και αποτελεσματική λύση για τη διαχείριση του προβλήματος της Εξόρυξης Δεδομένων. Αυτό συμβαίνει, λόγω του ότι για πολλά προβλήματα του πραγματικού κόσμου δε μπορούν να υπάρξουν αναλυτικές λύσεις, δηλαδή λύσεις που παρέχονται μέσω κλειστών μαθηματικών σχέσεων. Για αυτό τον λόγο, η χρήση των Υπολογιστικών Μαθηματικών και του Εξελικτικού Υπολογισμού ενδείκνυται και συμβάλλει σημαντικά στο πρόβλημα αυτό, ως μια αξιόπιστη εναλλακτική τακτική. Οι Εξελικτικοί Αλγόριθμοι θεωρούνται πολύ ισχυρές, προσαρμοστικές τεχνικές αναζήτησης μέσω των οποίων μπορεί να διεξαχθεί μια καθολική αναζήτηση στο χώρο των λύσεων ενός δεδομένου προβλήματος.

Ο μ , τα (Artificial Neural Networks), τα μ κ.ά. αποτελούν τα βασικότερα παραδείγματα, μεταξύ άλλων τεχνικών, που ανήκουν στις προαναφερθείσες κατηγορίες και χρησιμοποιούνται σε ένα ευρύ φάσμα εφαρμογών. Οι αλγόριθμοι Εξελικτικών Υπολογισμών, όπως οι μ (Genetic Algorithms) ή οι (Evolutionary Strategies), οι αλγόριθμοι μ (Particle Swarm Optimization) και οι μ (Differential Evolution Algorithms) είναι στοχαστικές μέθοδοι ολικής βελτιστοποίησης, που βασίζονται στους πληθυσμούς και εφαρμόζονται για την επίλυση σημαντικών προβλημάτων σε διάφορους επιστημονικούς κλάδους, όπως η ιατρική [276], η οικονομία [409], η μηχανική [239], η ρομποτική [81] καθώς και η μηχανική μάθηση και εξόρυξη δεδομένων [105] κ.ά.

Η κάθε κατηγορία των Εξελικτικών Αλγορίθμων έχει τα δικά της ξεχωριστά γνωρίσματα, αλλά η αυστηρή διαφοροποίησή τους δεν είναι εφικτή, καθώς οι σύγχρονοι Εξελικτικοί Αλγόριθμοι «αλληλοδανείζονται» στοιχεία, όπως τον τρόπο κωδικοποίησης ή τους τελεστές εξέλιξης, ανάλογα με την καταλληλότητά τους στην κάθε εφαρμογή [443]. Οι Εξελικτικοί Αλγόριθμοι χρησιμοποιούν έναν αλγόριθμο επιλογής προκειμένου να διαλέξουν τα άτομα υψηλότερης ποιότητας. Η πιθανότητα επιλογής ενός συγκεκριμένου ατόμου για γονέα, εξαρτάται προφανώς από την ποιότητά του. Η τελική φάση των Εξελικτικών Αλγορίθμων είναι η (reproduction) με την (crossover) των γονιδίων τους. Στο χρωμόσωμα που προκύπτει, υπάρχει και η πιθανότητα της μ (mutation) με αποτέλεσμα το νέο χρωμόσωμα να αποκτά και κάποια δικά του ιδιαίτερα χαρακτηριστικά, εκτός αυτών που κληρονόμησε από τους γονείς του. Οι παραπάνω τελεστές είναι ανάμεσα στους βασικότερους [260] και συναντώνται σε όλο το φάσμα των εξελικτικών διαδικασιών. Από τη στιγμή που υψηλής ποιότητας άτομα είναι πιθανότερο να γίνουν γονείς, οι μετέπειτα γενιές είναι πιθανό να αποτελούνται από ορισμένα (τουλάχιστον) άτομα, που όχι μόνο συνδυάζουν τα καλά χαρακτηριστικά των γονιών τους, αλλά είναι και υψηλότερης ποιότητας από αυτούς. Το σημαντικότερο πλεονέκτημα των Γενετικών Αλγορίθμων και γενικότερα Εξελικτικών αλγορίθμων έναντι των κλασικών μεθόδων βελτιστοποίησης, είναι ότι δεν απαιτούν καμία προϋπόθεση (π.χ. κυρτότητα, παραγωγισιμότητα) για την εύρεση του ολικού ακροτάτου. Η ικανότητά τους αυτή οφείλεται στη μορφή της μεθόδου, η οποία χρησιμοποιεί την αρχή της «τυχαίας αναζήτησης λύσης» (random search). Η αναζήτηση της λύσης έχει κατά μία έννοια τυχαίο χαρακτήρα, ο οποίος όμως συνδυάζεται με την ύπαρξη μιας συγκεκριμένης κατεύθυνσης. Οι Γενετικοί Αλγόριθμοι αξιοποιούν τη γνώση από τις προηγούμενες γενιές χρωμοσωμάτων για να κατασκευάσουν μία νέα γενιά που θα πλησιάζει τη βέλτιστη λύση.

Στην μ (single-objective) [87] ο χώρος των λύσεων είναι

(συνήθως) εύκολα προσδιορίσιμος και η βέλτιστη λύση ορίζεται μονότιμα. Η εισαγωγή επιπλέον αντικειμενικών συναρτήσεων σε ένα πρόβλημα βελτιστοποίησης - μ (multi-objective optimization) [73] (αναλύονται στο Κεφάλαιο 7) - και η απαίτηση της ταυτόχρονης βελτιστοποίησης έκαστης, έχει σαν αποτέλεσμα τόσο την αύξηση του πλήθους των ενδεχόμενων λύσεων, (η βέλτιστη λύση δεν είναι μία αλλά πολλές), όσο και τη δυσχέρεια του ακριβούς προσδιορισμού του χώρου των λύσεων. Οι λύσεις αυτές είναι βέλτιστες, υπό την έννοια ότι δεν υπάρχουν άλλες λύσεις στο χώρο των λύσεων που να είναι καλύτερες από αυτές· Δηλαδή δεν μπορεί να υπάρξει ένα καθολικό κριτήριο που να ορίσει μία από αυτές ως καλύτερη. Σκοπός, λοιπόν, της βελτιστοποίησης πολλαπλών στόχων, είναι η εύρεση του συνόλου των βέλτιστων λύσεων, γνωστού και ως Pareto βέλτιστο σύνολο.

Το γεγονός ότι ολοένα και περισσότεροι ερευνητές καταπνάνονται με προβλήματα που ανήκουν στο χώρο της μ , όπως η μ (feature selection), η διαχείριση μ - μ μ (imbalanced dataset) ή δεδομένων με μ μ (missing data), το πρόβλημα μ μ (classification) κ.ά., έχει ως αποτέλεσμα την παραγωγή μιας πληθώρας μεθόδων Υπολογιστικής Νοημοσύνης. Αυτό βέβαια δε σημαίνει πως περιορίζεται ο χώρος για την παραγωγή νέων, καινοτόμων και ευφυών μεθόδων και τεχνικών Μηχανικής Μάθησης (όπως αναλύεται στο Κεφάλαιο 2 της διατριβής), που θα είναι βέλτιστες, πιο ακριβείς και πιο αποτελεσματικές από τις ήδη υπάρχουσες. Με άλλα λόγια, μέσα από τη συγκεκριμένη διδακτορική διατριβή, φιλοδοξούμε να παράγουμε νέα Υπολογιστικά Συστήματα και υβριδικούς αλγορίθμους Μηχανικής Μάθησης και Εξόρυξης Δεδομένων με τη βοήθεια Εξελικτικών Υπολογισμών, για την καλύτερη αντιμετώπιση γνωστών προβλημάτων Υπολογιστικής Νοημοσύνης (όπως αναφέρθηκαν παραπάνω).

Η μ (Machine Learning) [271] είναι ένας τομέας της επιστήμης που επεξεργάζεται δεδομένα και προσπαθεί να δημιουργήσει ευφυή Υπολογιστικά Συστήματα. Το πώς ορίζεται αυτή η «ευφυία», όπως αναλύσαμε παραπάνω, είναι πολλές φορές υποκειμενικό. Ωστόσο μπορούμε απλά να το αποδοσουμε ως «σκέψη» ή «απόκριση» σε διαφόρων ειδών προβλήματα, όπως η ταξινόμηση, η διακριτοποίηση, η παλινδρόμηση, η ομάδοποίηση κ.ά. Με άλλα λόγια, προσπαθούμε να κάνουμε έναν υπολογιστή να έχει δική του «σκέψη». Τα σύνολα δεδομένων που καλούμαστε να επεξεργαστούμε σήμερα γίνονται ολοένα και μεγαλύτερα (big data). Έτσι, μπορεί να γίνει εύκολα αντιληπτό γιατί ο τομέας της Μηχανικής Μάθησης συγκεντρώνει τόσο μεγάλο ενδιαφέρον. Κάτι τέτοιο, σε συνδυασμό με την ραγδαία ανάπτυξη των Υπολογιστικών Μηχανών, μας δίνει τη δυνατότητα να εξαγάγουμε χρήσιμες πληροφορίες και να διακρίνουμε μια πλειάδα από προσεγγιστικές διαδικασίες όπως τα μ - μ (decision trees), τα μ (artificial neural networks), οι μ μ (support vector machines), οι μ μ (Bayesian classifiers), τα μ (deep neural networks) κ.ά.

Η μ μ (Data mining) [172] συνδέεται άμεσα με τα μοτίβα τα οποία «κρύβονται» στα σύνολα δεδομένων τα οποία πρέπει να επεξεργαστούμε για να πάρουμε χρήσιμες πληροφορίες. Για το σκοπό αυτό, χρησιμοποιούνται πολλές τεχνικές από το πεδίο της Μηχανικής Μάθησης, της Στατιστικής, της Υπολογιστικής Νοημοσύνης αλλά και άλλων επιστημονικών πεδίων. Κύριος στόχος, αφότου γίνουν οι απαραίτητες ενέργειες, όπως για παράδειγμα η μ μ (data preprocessing) (ο αναγνώστης θα βρει λεπτομέρειες στο Κεφάλαιο 4) ή η προετοιμασία του δείγματος (data preparation), είναι τα παραγόμενα μοντέλα να έχουν την ικανότητα αυτοματοποίησης της γνώση που έλαβαν, ώστε να μπορούν να μ (generalization ability) και να δώσουν απαντήσεις για διαφορετικά, άγνωστα σύνολα δεδομένων.

Φυσικά, κάτι τέτοιο είναι πολλές φορές δύσκολο να πραγματοποιηθεί, καθώς τα σύνολα δεδομένων που λαμβάνουμε, είτε λόγω του όγκου τους είτε λόγω της ποιότητάς τους, δε μπορούν να χρησιμοποιηθούν άμεσα. Τα προβλήματα τα οποία μπορεί να συναντήσουμε κατά τη διάρκεια της παραπάνω προσπάθειας είναι πολλά και ποικίλα. Ενδεικτικά, σημειώνουμε τα εξής: επιλογή κατάλληλων χαρακτηριστικών (feature selection), επιλογή παραδειγμάτων

(instance selection), ελλιπή σύνολα δεδομένων (missing values), μη-ισοροπημένα δεδομένα (imbalance datasets) κ.ά. Τα παραπάνω προβλήματα μπορούν να «καμφθούν» αποτελεσματικά με τη βοήθεια Εξελικτικών Τεχνικών και μεθόδων Υπολογιστικής Νοημοσύνης. Για παράδειγμα, έχουν αναπτυχθεί τεχνικές Εξελικτικού Υπολογισμού [174] και συγκεκριμένα, αλγοριθμικά σχήματα που πραγματοποιούν βελτιστοποίηση με χρήση σμηνών σωματιδίων, επιλέγοντας έτσι, κατάλληλα χαρακτηριστικά του αρχικού συνόλου δεδομένων, ανταποκρινόμενα στο πρόβλημα της ταξινόμησης. Ακόμα, δύσκολα διαχρονικά προβλήματα, όπως αυτό της επιλογής των (clusters), στα οποία διακρίνεται ένα σύνολο δεδομένων, προσεγγίστηκε [408] αποτελεσματικά και αποδοτικά με τη χρήση ενός εξελικτικού αλγορίθμου πολυ-αντικειμενικής βελτιστοποίησης.

Τέλος, η ιδανική κατάσταση, όπως αναφέραμε παραπάνω, είναι το αναπτυσσόμενο μοντέλο να είναι σε θέση να γενικεύσει πάνω σε άγνωστα δεδομένα και μάλιστα με την ίδια ακρίβεια. Σε εφαρμογές του πραγματικού κόσμου όμως, τα μοντέλα/αλγόριθμοι μάθησης έρχονται αντιμέτωπα με μη-ισοροποιημένα σύνολα δεδομένων, τα οποία μπορούν να οδηγήσουν το μοντέλο προς μια κλάση, και συνεπώς σε λάθος ταξινόμηση. Αυτή η «προκατάληψη» είναι απόρροια μιας υπο-εκπροσωπούμενης κλάσης δεδομένων κατά την εκπαίδευση του αλγορίθμου, γεγονός που καθιστά την εξισορρόπηση του συνόλου δεδομένων απαραίτητη. Τα παραπάνω παραδείγματα, είναι μερικά μόνο αποτελέσματα στα οποία έχει καταλήξει η επιστημονική κοινότητα και καταδεικνύουν τόσο τα προβλήματα που υπάρχουν στους συγκεκριμένους επιστημονικούς τομείς, όσο και τις λύσεις που έχουν δοθεί.

1.5 Συνεισφορά

Κατά συνέπεια, προσδοκούμε ότι τα αποτελέσματα της διατριβής θα συνεισφέρουν στη δημιουργία αξιόπιστων μεθόδων, οι οποίες, μεταξύ των άλλων, ενδεχομένως να δώσουν τη δυνατότητα σε επιστήμονες άλλων περιοχών για την αξιόπιστη αντιμετώπιση μεγάλης κλίμακας και πολυπλοκότητας προβλημάτων, τα οποία συχνά εμφανίζονται σε πολλούς επιστημονικούς, οικονομικούς, βιομηχανικούς και εμπορικούς τομείς με προφανή οφέλη. Το συμπέρασμα που απορρέει από τα παραπάνω είναι ότι, αν και τα τελευταία χρόνια υπάρχουν μέθοδοι βελτιστοποίησης που έχουν μελετηθεί ευρέως και έχουν κερδίσει την εμπιστοσύνη της επιστημονικής κοινότητας, ωστόσο υπάρχουν πάντα περιθώρια ανάπτυξης νέων μεθόδων, οι οποίες βασίζονται σε καινοτόμες ιδέες και μπορούν να αντιμετωπίσουν αποδοτικότερα κλάσεις δύσκολων προβλημάτων του πραγματικού κόσμου.

Οι παρακάτω δημοσιεύσεις αφορούν εργασίες, οι οποίες δημοσιεύτηκαν στα πλαίσια εκπόνησης της διδακτορικής διατριβής:

(α) Εργασίες σε έγκριτα διεθνή επιστημονικά περιοδικά με σύστημα κριτών:

(J1) Alexandropoulos S.-A.N., Pardalos P.M., Vrahatis M.N., Dynamic search trajectory methods for global optimization, *Annals of Mathematics and Artificial Intelligence*, 88(1-3), pp.3-37, 2020.

(J2) Alexandropoulos S.-A.N., Kotsiantis S.B., Vrahatis M.N., Data preprocessing in predictive data mining, *Knowledge Engineering Review*, 34, Article No. e1, pp.1-33, 2019.

(β) Εργασίες σε πρακτικά συνεδρίων με σύστημα κριτών:

(C1) Alexandropoulos S.-A.N., Kotsiantis S.B., Piperigou V.E., Vrahatis M.N., A new ensemble method for outlier identification, *Proceedings of the IEEE Tenth International Conference on Cloud Computing, Data Science Engineering (Confluence 2020)*, January 29-31, 2020, Noida, Uttar Pradesh, India, pp.786-791, IEEE 2020.

(γ) Εργασίες σε κεφάλαια βιβλίων και συλλογικούς τόμους με κριτές:

(V1) Alexandropoulos S.-A.N., Aridas C.K., Kotsiantis S.B., Vrahatis M.N., A deep dense neural network for bankruptcy prediction, *Communications in Computer and Information Science (CCIS)*, vol.1000, pp.435-444, Springer Nature Switzerland AG, Switzerland, 2019.

(V2) Alexandropoulos S.-A.N., Aridas C.K., Kotsiantis S.B., Vrahatis M.N., Stacking strong ensembles of classifiers, *IFIP Advances in Information and Communication Technology*, vol.559, pp.545-556, Springer Nature Switzerland AG, Switzerland, 2019.

(δ) Εργασίες σε κεφάλαια βιβλίων μετά από πρόσκληση και κρίση:

(Ch1) Alexandropoulos S.-A.N., Aridas C.K., Kotsiantis S.B., Vrahatis M.N., Multi-objective evolutionary optimization algorithms for machine learning: A recent survey, *Approximation and Optimization*, I.C. Demetriou and P.M. Pardalos (eds.), Chapter 4, pp.35-55, Springer Optimization and Its Applications 145, Springer Nature Switzerland AG 2019 [ISBN: 978-3-030-12766-4, ISBN: 978-3-030-12767-1 (eBook)].

Μέρος II

Μαθηματική Βελτιστοποίηση

Δεν υπάρχει δωρεάν γεύμα για τη Μαθηματική Βελτιστοποίηση

Labor omnia vincit.

— (70-19 . .)

Tο θεώρημα «μη δωρεάν γεύματος» (No Free Lunch (NFL) theorem) δηλώνει ότι, κατά μέσο όρο για όλα τα προβλήματα βελτιστοποίησης, χωρίς επαναδειγματοληψία, όλοι οι αλγόριθμοι βελτιστοποίησης αποδίδουν εξίσου καλά. Η βελτιστοποίηση, η αναζήτηση και η επιτηρούμενη μάθηση είναι οι τομείς που έχουν ωφεληθεί περισσότερο από αυτό το σημαντικό θεωρητικό αποτέλεσμα. Η διατύπωση του αρχικού NFL θεώρηματος, πολύ σύντομα, οδήγησε σε μια σειρά ερευνητικών εργασιών, που ακολούθως οδήγησαν σε μια σειρά θεωρημάτων που ορίζουν ένα ολόκληρο ερευνητικό πεδίο με σημαντικά αποτελέσματα σε άλλους επιστημονικούς τομείς, όπου η επιτυχής διερεύνηση ενός χώρου αναζήτησης είναι απαραίτητη και αρκετά δύσκολη εργασία. Στόχος του παρόντος κεφαλαίου είναι να παρουσιάσει τις κύριες ερευνητικές προσπάθειες που συνέβαλαν σε αυτόν τον ερευνητικό τομέα, να αποκαλύψει τα βασικά ζητήματα και να χαρτογραφήσει εκείνα τα σημεία που είναι χρήσιμα στην κατανόηση των υποθέσεων, των περιορισμών ή ακόμα και της ανικανότητας εφαρμογής των θεωρημάτων που σχετίζονται με «δωρεάν γεύματα» (Free Lunches).

Τα προβλήματα βελτιστοποίησης που συναντώνται σε διάφορους τομείς της επιστήμης, της πληροφορικής και της μηχανικής εξαρτώνται από τον αριθμό των παραμέτρων, το μέγεθος του χώρου λύσεων, και κυρίως από την αντικειμενική συνάρτηση, της οποίας ο ορισμός είναι κρίσιμος, καθώς καθορίζει σε μεγάλο βαθμό το επίπεδο δυσκολίας του προβλήματος. Ως εκ τούτου, ο ορισμός και η επίλυση ενός προβλήματος βελτιστοποίησης είναι μερικές φορές ένα εξαιρετικά δύσκολο και απαιτητικό έργο. Ερευνητές από διάφορους τομείς έχουν εμπλακεί με την επίλυση προβλημάτων βελτιστοποίησης, είτε επειδή αποτελούν μέρος της κύριας έρευνάς τους, είτε επειδή το ζήτημα που αντιμετωπίζουν μπορεί να αντιμετωπιστεί ως ένα τέτοιο πρόβλημα. Οι ερευνητικές προσπάθειες σε αυτό το θέμα επέτρεψαν τη δημιουργία πολλών μεθόδων και τεχνικών, βασισμένων σε συμπαγείς μαθηματικές έννοιες, των οποίων η εφαρμογή είχε σημαντικά θετικά αποτελέσματα.

Ωστόσο, εν αντιθέσει με κάθε αντίθετο ισχυρισμό, καμία από αυτές τις μεθόδους δεν έχει αποδειχθεί επιτυχής σε όλους τους τύπους των προβλημάτων που εφαρμόστηκαν. Αυτό το επιχείρημα ήταν ο στόχος της σημαντικής θεωρητικής εργασίας που πραγματοποίησε ο David Wolpert και οδήγησε στο γνωστό θεώρημα No Free Lunch (NFL). Εν συντομία, το θεώρημα NFL δηλώνει ότι:

« μ μ , μ , . »

«... μ μ μ μ μ - μ - μ μ μ μ μ a priori μ μ μ.»

Ο Wolpert βασίζει τη θεωρητική του εργασία σε παλαιότερες εξελίξεις που αναπτύχθηκαν στην εργασία του «Σχετικά με τη σύνδεση μεταξύ δοκιμής εντός δείγματος και σφάλματος γενίκευσης» [418]. Σε αυτό το έργο, το σφάλμα γενίκευσης λαμβάνεται ως το σφάλμα εκτός του συνόλου εκπαίδευσης (Off-Training Set (OTS)) και η ερώτηση που τίθεται αφορά το συσχετισμό του με το σφάλμα που παράγεται χρησιμοποιώντας τις δοκιμές εντός του δείγματος. Επιπλέον, ο Wolpert ασχολείται με το πώς «... μ

μ μ », καθώς οποιαδήποτε θεωρία γενίκευσης είναι άσχετη όσον αφορά την εφαρμογή της στα πραγματικά προβλήματα, αν δεν αντιμετωπίσει το προηγούμενο πρόβλημα. Ορισμένα - αλλά όχι όλα - από τα σημαντικά ζητήματα που εγείρονται σε αυτή την εργασία είναι τα ακόλουθα:

- (α) « μ μ » Με άλλα λόγια, δεδομένης της επίδοσης ενός αλγορίθμου μάθησης στο σύνολο δεδομένων εκπαίδευσης, είναι δυνατόν να ληφθούν πληροφορίες σχετικά με την ικανότητά του να παρέχει μια ακριβή αναπαράσταση της συνάρτησης στόχου, για παραδείγματα εκτός του συνόλου δεδομένων;
- (β) Αν δεν μπορέσουμε να απαντήσουμε στην προηγούμενη ερώτηση, ποιες είναι οι υποθέσεις για την κατανομή πραγματικών δεδομένων (η συνάρτηση στόχου) που μπορούν να βοηθήσουν στη γενίκευση αλγορίθμων εκπαίδευσης, όπως ο αλγόριθμος οπίσθιας διάδοσης σφάλματος (back-propagation algorithm), που στοχεύουν στην ελαχιστοποίηση του σφάλματος σχετικά με τα δεδομένα εκπαίδευσης;
- (γ) Υπάρχει μαθηματική αρχή για την εκτίμηση του πότε λαμβάνει χώρα το φαινόμενο της υπερ-εκπαίδευσης, ώστε να προχωρήσουμε στην τροποποίηση του αλγορίθμου μάθησης, προκειμένου να περιοριστούν οι επιπτώσεις μιας τέτοιας υπερ-εκπαίδευσης;
- (δ) Είναι δυνατόν να εκφραστούν από μαθηματικούς όρους η ικανότητα ενός συνόλου εκπαίδευσης να αντιπροσωπεύει πιστά την κατανομή σε ολόκληρο το χώρο των δεδομένων;
- (ε) Ποιες είναι οι υποθέσεις βάσει των οποίων οι μη παραμετρικές στατιστικές τεχνικές, όπως της διασταυρωμένης επικύρωσης (cross-validation), (που έχουν σχεδιαστεί για να επιλέξουν μεταξύ αλγορίθμων μάθησης), επιτυγχάνουν να μειώσουν το σφάλμα γενίκευσης;

Κατά την επίλυση των παραπάνω ζητημάτων, ο προτεινόμενος φορμαλισμός φαίνεται να επεκτείνει τον κλασικό Bayesian φορμαλισμό χρησιμοποιώντας τη συνάρτηση της υπόθεσης (hypothesis function), δηλαδή την κατανομή του συνόλου δεδομένων, όπως αυτή μαθαίνεται από το «γενικευτή». Ο μαθηματικός φορμαλισμός που υιοθετείται προτείνει έναν τρόπο ταιριάσματος του βαθμού, στον οποίο η κατανομή που προκύπτει από τον αλγόριθμο μάθησης ταιριάζει με την κατανομή των δεδομένων εκπαίδευσης και μπορεί να χρησιμοποιηθεί για την αντιμετώπιση διαφόρων ζητημάτων γενίκευσης, όπως η υπερ-εκπαίδευση και ο ελάχιστος αριθμός παραμέτρων για το μοντέλο που χρησιμοποιείται. Από άλλη άποψη, ο φορμαλισμός αυτός προτείνεται με σκοπό να εκφράσει με μαθηματικούς όρους τις παραδοχές ενός γενικευτή, ώστε το χρησιμοποιούμενο μοντέλο να ταιριάζει καλύτερα στο σύνολο εκπαίδευσης που αντιπροσωπεύει τον πραγματικό κόσμο. Ως εκ τούτου, η επεξεργασία σημαντικών θεωρητικών αποδείξεων προτείνει μια σταθερή βάση για την αντιμετώπιση διαφόρων θεμάτων στη Μηχανική Μάθηση και οδηγεί στην ανάπτυξη εννοιών όπως τα θεωρήματα NFL.

το μέσο όρο των «γενικευτών» παρά των στόχων. Αυτό σημαίνει ότι αντί να χαρακτηρίζουμε δύο αλγορίθμους λαμβάνοντας το μέσο όρο αναφορικά με όλους τους διαθέσιμους στόχους, έστω f ; $P(f)$ ή $P(\cdot)$, κρατώντας την υπόθεση $P(h/d)$ σταθερή, είναι δύσκολο να εξεταστούν εναλλακτικά αποτελέσματα όταν κάποιος κατέχει μία από τις οντότητες σχετικά με τους στόχους, καθορίζει και υπολογίζει κατά μέσο όρο τις οντότητες υπόθεσης. Για αυτή την περίπτωση, ο Wolpert διατυπώνει κάποια επιπλέον θεωρήματα και τελικά εξετάζει την περίπτωση όπου η συνάρτηση απώλειας (loss function) $L(j)$ είναι μη ομογενής (non-homogeneous) και έτσι, τα θεωρήματα NFL δεν εφαρμόζονται, καθώς κάποιος μπορεί να κάνει a priori διακρίσεις μεταξύ αλγορίθμων.

Ως συμπέρασμα αναφέρεται στο άρθρο [420], ότι αυτές οι δύο εργασίες διερευνούν κάποια συμπεριφορά του σφάλματος OTS. Συγκεκριμένα, επισημοποιούν και διερευνούν την έννοια ότι

« μ , μ ».

2.2 Δεν υπάρχει δωρεάν γεύμα για τη βελτιστοποίηση και την αναζήτηση

Μια άλλη κατεύθυνση της έρευνας για την εφαρμογή των ιδεών που σχετίζονται με τα θεωρήματα «μη ύπαρξης δωρεάν γευμάτων», που παρουσιάστηκαν παραπάνω, αφορά τον τομέα της βελτιστοποίησης. Η εργασία « μ » [424] που δημοσιεύτηκε από τους Wolpert και McReady ασχολείται με το θέμα αυτό, με βάση δύο τεχνικές εκθέσεις που εκπονήθηκαν από τους συγγραφείς στο Ινστιτούτο του Santa Fe. Η πρώτη τεχνική αναφορά που δημοσιεύθηκε [417], με τον τίτλο « μ ;» θέτει την ερώτηση:

« μ , μ »

Η δεύτερη τεχνική αναφορά [416], με τίτλο «Δεν υπάρχουν δωρεάν γεύματα για αναζήτηση» επικεντρώνεται στην απόδειξη ότι όλοι οι αλγόριθμοι που ψάχνουν για ένα βέλτιστο σε κάποιο πρόβλημα βελτιστοποίησης, δηλαδή ένα ακραίο σημείο μιας αντικειμενικής συνάρτησης, αποδίδουν ακριβώς το ίδιο, ανεξάρτητα από το μέτρο απόδοσης που χρησιμοποιείται, όταν λαμβάνεται ο μέσος όρος για όλες τις δυνατές αντικειμενικές συναρτήσεις.

Το έργο των Wolpert και McReady «Δεν υπάρχουν δωρεάν γεύματα στη βελτιστοποίηση» [424], θέτει ένα φορμαλισμό για τη διερεύνηση της σχέσης της αποτελεσματικότητας των αλγορίθμων βελτιστοποίησης και των προβλημάτων που επιλύουν. Τα θεωρήματα «μη ύπαρξης δωρεάν γευμάτων» που αναπτύχθηκαν στην εργασία τους, καταδεικνύουν ότι η επιτυχής απόδοση οποιουδήποτε αλγορίθμου βελτιστοποίησης σε μια τάξη προβλημάτων αντιστοιχίζεται από την «υποβαθμισμένη» επίδοση σε μια άλλη κατηγορία προβλημάτων. Μια γεωμετρική ερμηνεία παρέχεται σχετικά με τη σημασία της καταλληλότητας ενός αλγορίθμου για την αντιμετώπιση ορισμένων προβλημάτων βελτιστοποίησης. Επιπλέον, όπως αναφέρθηκε στις προηγούμενες τεχνικές αναφορές, οι συγγραφείς τους εξετάζουν εφαρμογές των θεωρημάτων NFL σε θεωρητικές πτυχές της βελτιστοποίησης, καθώς και στον ορισμό μέτρων απόδοσης για πλαίσια που άπτονται της βελτιστοποίησης.

Δεδομένης της πληθώρας διαθέσιμων τεχνικών βελτιστοποίησης «μαύρου-κουτιού» (black-box optimization techniques), οι Wolpert και McReady προσπαθούν να παράσχουν τον φορ-

μαλισμό για την αντιμετώπιση του ακόλουθου προβλήματος:

« μ ; μ μ - »

Αυτό το πρόβλημα μπορεί να προβληθεί σε πολλά άλλα όπως:

- Ποια είναι τα μαθηματικά συστατικά της θεωρίας βελτιστοποίησης που κάποιος πρέπει να γνωρίζει πριν αποφασίσει σχετικά με τις απαραίτητες κατανομές πιθανοτήτων που πρέπει να εφαρμοστούν;
- Είναι η θεωρία των πληροφοριών και η Bayesian ανάλυση κατάλληλη για την κατανόηση των προηγούμενων θεμάτων;
- Δεδομένων των αποτελεσμάτων απόδοσης ενός συγκεκριμένου αλγορίθμου σε μια συγκεκριμένη κατηγορία προβλημάτων, μπορεί κανείς να παράσχει a priori γενίκευση αυτών των αποτελεσμάτων σε άλλες κατηγορίες προβλημάτων;
- Υπάρχει κατάλληλο μέτρο μιας τέτοιας γενίκευσης; Μπορεί κάποιος να αξιολογήσει την απόδοση των αλγορίθμων σε προβλήματα, ώστε να είναι σε θέση να συγκρίνει αυτούς τους αλγορίθμους;

Ο φορμαλισμός που αναπτύχθηκε από τους συγγραφείς είναι διαρθρωμένος γύρω από τις ακόλουθες έννοιες:

- Ένα δείγμα μεγέθους m είναι ένα σύνολο m διαφορετικών σημείων που επισκέπτεται ο αλγόριθμος και υποδηλώνεται από το ακόλουθο σύνολο

$$d_m = f(d_m^x(1); d_m^y(1)); (d_m^x(2); d_m^y(2)); \dots; (d_m^x(m); d_m^y(m))g$$

όπου το $d_m^x(i)$ δηλώνει την τιμή X του στοιχείου i του δείγματος και $d_m^y(m)$ είναι το σχετικό κόστος, δηλαδή η τιμή Y .

- Ένας αλγόριθμος βελτιστοποίησης είναι μια χαρτογράφηση από τα σύνολα σημείων που επισκεφθήκε προηγουμένως σε ένα καινούριο σημείο του συνόλου X , δηλαδή

$$: d \in D ! f_{X|X} \geq d^x g;$$

όπου Δ δηλώνει το χώρο όλων των (m μεγέθους) δειγμάτων και το είναι ένας ντετερμινιστικός αλγόριθμος, με την έννοια ότι κάθε δείγμα αντιστοιχεί σε ένα μοναδικό νέο σημείο.

- Η απόδοση ενός αλγορίθμου μετά από m επαναλήψεις είναι μια συνάρτηση (d_m^y) του δείγματος.
- Δοθέντος ολόκληρου του χώρου των δυνατών συναρτήσεων κόστους, για παράδειγμα τα προβλήματα βελτιστοποίησης F , η κατανομή

$$P(f) = P(f(x_1); f(x_2); \dots; f(x_{|X|}))$$

που ορίζεται αναφορικά με τα προβλήματα F παρέχει την πιθανότητα κάθε $f \in F$ να είναι το πραγματικό πρόβλημα βελτιστοποίησης που διαχειριζόμαστε.

5. Η απόδοση ενός αλγορίθμου βελτιστοποίησης αναφορικά με μια συνάρτηση κόστους f μετά από m επαναλήψεις μετράται ως $P(d_m^y f; m; \cdot)$.

Ας εξετάσουμε σε αυτό το σημείο το ακόλουθο πρόβλημα :

$$F_1 \quad F \quad \mu \quad 2 \quad F_2 \quad F \quad \mu \quad 1 \quad 2 \quad -$$

Η απάντηση που έδωσαν οι συγγραφείς βασίζεται στο άθροισμα των $P(d_m^y f; m; \cdot_1)$ και το άθροισμα των $P(d_m^y f; m; \cdot_2)$ αναφορικά με όλες τις διαθέσιμες συναρτήσεις f , δηλαδή σε όλα τα προβλήματα βελτιστοποίησης. Το ακόλουθο θεώρημα, όπως διατυπώνεται στην εργασία τους [424], διευθετεί το προηγούμενο πρόβλημα.

Θεώρημα 1

$$\times \quad \mu \quad 1 \quad 2, \\ P(d_m^y f; m; \cdot_1) = \times \quad P(d_m^y f; m; \cdot_2):$$

Στο παραπάνω θεώρημα το πρόβλημα θεωρείται σταθερό με την πάροδο του χρόνου. Αν η συνάρτηση κόστους εξαρτάται από το χρόνο, με την έννοια ότι, ενώ το πρόβλημα αρχικά εκφράζεται με κάποια συνάρτηση κόστους f_1 που υπάρχει όταν λαμβάνεται η πρώτη τιμή στο X , τότε αυτή η συνάρτηση μετασχηματίζεται πριν από οποιαδήποτε επακόλουθη επανάληψη του αλγορίθμου βελτιστοποίησης. Εάν ο μετασχηματισμός αντιπροσωπεύεται με την απεικόνιση $T : F \rightarrow F$, και $T = T_i$, τότε $f_{i+1} = T_i(f)$ και το ακόλουθο θεώρημα μπορεί να διατυπωθεί:

Θεώρημα 2

$$d_m^y; D_m^y; m > 1, \quad \mu \quad 1 \quad 2, \\ \times \quad P(d_m^y f_1; T; m; \cdot_1) = \times \quad P(d_m^y f_1; T; m; \cdot_2); \\ \times \quad P(D_m^y f_1; T; m; \cdot_1) = \times \quad P(D_m^y f_1; T; m; \cdot_2):$$

Μια από τις συνέπειες των θεωρημάτων NFL που συζητήθηκαν από τους συγγραφείς, ασχολείται με τη γεωμετρική προοπτική των θεωρημάτων NFL. Από την άποψη αυτή, θεωρούμε το χώρο F όλων των συναρτήσεων κόστους και την πιθανότητα απόκτησης ενός συγκεκριμένου d_m^y που ορίζεται από τη σχέση:

$$P(d_m^y; m; \cdot) = \times \quad P(d_m^y; m; \cdot; f) P(f);$$

με το $P(f)$ να είναι η εκ των προτέρων πιθανότητα το πρόβλημα βελτιστοποίησης που διαχειριζόμαστε να έχει ως συνάρτηση κόστους την f . Όπως σημειώνεται από τους συγγραφείς, το προηγούμενο άθροισμα μπορεί να θεωρηθεί ως εσωτερικό γινόμενο στο χώρο F . Επομένως, αν ορίσουμε τα διανύσματα $v_{d_m^y; m}$ και p , με τις f συνιστώσες τους αντίστοιχα, τότε:

$$v_{d_m^y; m}(f) = P(d_m^y; m; \cdot; f); \quad p = P(f);$$

τότε, ισχύει ότι:

$$P(d_m^y; m; \cdot) = v_{d_m^y; m} \cdot p;$$

Οι συγγραφείς σημειώνουν ότι αυτή η εξίσωση παρέχει μια γεωμετρική ερμηνεία της διαδικασίας βελτιστοποίησης. Επομένως, το d_m^y αντιπροσωπεύει το επιθυμητό δείγμα και το m λαμβάνεται ως μέτρο της υπολογιστικής προσπάθειας που απαιτείται για τον αλγόριθμο. Επιπλέον, εάν ένα διάνυσμα ρ αντιπροσωπεύει το προηγούμενο που περιλαμβάνει όλες τις γνώσεις σχετικά με τις συναρτήσεις κόστους, τότε η τελευταία εξίσωση διατυπώνει με μαθηματικούς όρους ότι:

«Η απόδοση ενός αλγορίθμου καθορίζεται από το μέγεθος της προβολής του στο ρ ή με άλλα λόγια από το πόσο ευθυγραμμισμένο είναι το διάνυσμα ρ με το διάνυσμα του προβλήματος ρ .»

Σε σχέση με τη γεωμετρική θεώρηση που προκύπτει, το θεώρημα NFL έχει ως αποτέλεσμα ότι το $P(d_m^y|f; m; \rho)$ είναι ανεξάρτητο του αλγορίθμου ρ , κάτι που σημαίνει ότι για κάθε συγκεκριμένο d_m^y και m , όλοι οι αλγόριθμοι έχουν την ίδια προβολή πάνω στην ομοιόμορφη $P(f)$ που αντιπροσωπεύεται από το διαγώνιο διάνυσμα $\mathbf{1}$.

Επιπλέον, οι συγγραφείς Wolpert και McReady διερευνούν τη σχέση των παραπάνω αποτελεσμάτων με τις θεωρητικές πτυχές της βελτιστοποίησης και παρέχουν μέτρα απόδοσης για την αξιολόγηση της αποτελεσματικότητας ενός συγκεκριμένου αλγορίθμου βελτιστοποίησης. Τέλος, συζητούνται ελάχιστες διαφορές μεταξύ αλγορίθμων αναζήτησης και παρέχονται ορισμένα μέτρα επιδόσεων για αλγόριθμους αναζήτησης.

2.3 Πρόσφατες εργασίες του Wolpert

Το άρθρο των Körpen, Wolpert και McReady «Παρατηρήσεις αναφορικά με το πρόσφατο άρθρο σχετικά με τα θεωρήματα μη ύπαρξης δωρεάν γεύματος» [220] είναι μια σύντομη εργασία που επανεξετάζει μια προηγούμενη εργασία του Körpen [219] με τον τίτλο «Ορισμένες τεχνικές παρατηρήσεις σχετικά με την απόδειξη του θεωρήματος μη ύπαρξης δωρεάν γεύματος». Σε αυτή την επιστολή, οι συγγραφείς, ακολουθώντας τις υποδείξεις δημοσιεύθηκαν στο άρθρο [219], παρέχουν μια σύντομη απόδειξη των θεωρημάτων NFL, ενώ διορθώνουν ένα λάθος ισχυρισμό που έγινε στην εργασία [219], σχετικά με τον κυκλικό ισχυρισμό της αρχικής απόδειξης των θεωρημάτων NFL που δημοσιεύθηκαν στα άρθρα [416, 424].

Ακολουθώντας, ας δώσουμε κάποιες λεπτομέρειες για αυτό το θεώρημα, όπως παρουσιάζεται στα άρθρα [416, 424]: η απόδειξή του είναι σημαντική για πολλά άρθρα σχετικά με τα θεωρήματα NFL. Αρχικά, θεωρούμε δύο πεπερασμένα σύνολα X και Y μαζί με το σύνολο όλων των συναρτήσεων κόστους $f: X \rightarrow Y$. Επιπλέον, για θετικό ακέραιο m , τέτοιο ώστε $m < |X|$, έστω $d_m = f(d_m^x(i); d_m^y(i) = f(d_m^x(i)))g$, δηλαδή τα σημεία του δείγματος που επισκέπτεται ο αλγόριθμος σε m βήματα, με $i = 1; 2; \dots; m$ $d_m^x(i) \in X$ και για κάθε $i; j$ ισχύει ότι $d_m^x(i) \neq d_m^x(j)$. Ας υποθέσουμε ότι ρ είναι ο αλγόριθμος αναζήτησης που μας ενδιαφέρει, ο οποίος είναι ένας ντετερμινιστικός «τυφλός» αλγόριθμος (blind algorithm), που αποδίδει σε κάθε πιθανό d_m ένα στοιχείο του X το οποίο δεν είναι ήδη στο d_m . Αυτό σημαίνει ότι,

$$d_{m+1}^x(m+1) = [d_m] \neq f d_m^x g;$$

Ας θεωρήσουμε ότι η ποσότητα $Y(f; a; m)$ υποδηλώνει την ακολουθία των m τιμών του Y που παράγεται από τον αλγόριθμο ρ για μια συνάρτηση f μετά τα m διαδοχικά βήματα εκτέλεσης και $\delta(\cdot; \cdot)$ είναι η συνάρτηση δέλτα του Kronecker, η οποία επιστρέφει 1 αν τα ορίσματά της είναι αυτόσημα και 0 διαφορετικά. Συνεπώς, ισχύει το ακόλουθο λήμμα:

Λήμμα 1

μ

d_m^y

$$\prod_{f \in F} (d_m^Y; Y(f; m;)) = \prod_{j \in J} j^{X_j} m;$$

$c() \mu \mu \mu \mu$
 $k \in \mathbb{R} \mu \mu \mu \mu$

Θεώρημα 3

$k \in \mathbb{R} \mu c(), \mu b, \mu$
 $\prod_{f \in F} (k; c(Y(f; m;))) = \prod_{f \in F} (k; c(Y(f; m; b)))$

Εκτός από την εξέταση της απόδειξης αυτού του θεωρήματος, σε αυτή την επιστολή, οι συγγραφείς είχαν την ευκαιρία να «υπερασπιστούν» τα θεωρήματα NFL ενάντια στην άποψη που αποκαλούν ως μηδενιστική περί μη ύπαρξης αλγορίθμων καθολικής εφαρμογής. Συμπερασματικά, τα θεωρήματα NFL θα πρέπει να θεωρούνται ως ένα ερευνητικό πεδίο και όχι, ως ένα απλό, βολικό ή ενοχλητικό αποτέλεσμα. Ως εκ τούτου, προτείνουν να επικρατήσει μια πιο «ανοιχτόμυαλη» άποψη για να διερευνηθούν τα όρια των θεωρημάτων NFL καθώς και τα δυνατά ζητήματα που προκύπτουν από την εφαρμογή τους σε διάφορους τομείς.

Αξίζει να αναφερθεί, σε αυτό το σημείο, μια σχετικά πιο πρόσφατη εργασία του Wolpert [423] με τίτλο: « $\mu \mu \mu \mu \mu \mu$ »; Στην παρούσα ερευνητική εργασία ο συγγραφέας της επανεξετάζει τις βασικές ιδέες της εργασίας του σχετικά με τα NFL, όσον αφορά τους αλγορίθμους αναζήτησης. Ο Wolpert επιμένει να αναλύει το ζήτημα ότι, ενώ τα θεωρήματα NFL έχουν ισχυρές συνέπειες, κάθε φορά που υιοθετείται ομοιόμορφη κατανομή για τις συναρτήσεις κόστους αναφορικά με τα προβλήματα βελτιστοποίησης, αυτό δεν προορίζεται να υποστηρίξει τη χρήση μιας τέτοιας κατανομής όταν κάποιος πρέπει να λύσει ένα πρόβλημα βελτιστοποίησης. Προσπαθώντας να διευκρινίσει τι σημαίνει πραγματικά το θεώρημα NFL προκειμένου να βελτιώσει τους αλγορίθμους αναζήτησης, ο Wolpert αναλύει ένα είδος « $\mu \mu \mu \mu \mu \mu$.» Ως αποτέλεσμα της ανάλυσης αυτής της σχέσης, υπάρχουν θεωρήματα «μη ύπαρξης δωρεάν γεύματος» τόσο για τη διαδικασία αναζήτησης όσο και για την επιτηρούμενη διαδικασία και, έτσι, υπάρχουν διάφοροι τρόποι επαναχρησιμοποίησης των τεχνικών που αναπτύχθηκαν αρχικά στην επιτηρούμενη μάθηση για την καθοδήγηση της αναζήτησης. Έτσι, παρουσιάζονται στην εργασία του ορισμένα πειράματα που επιβεβαιώνουν την αποτελεσματικότητα των αλγορίθμων αναζήτησης, οι οποίοι βασίζονται σε αυτές τις έννοιες.

2.4 Μη δωρεάν γεύμα για βελτιστοποίηση και εξελικτικούς αλγορίθμους

Σε αυτή την ενότητα θα παρουσιάσουμε τους συσχετισμούς των θεωρημάτων NFL με τη βελτιστοποίηση και μια συχνά χρησιμοποιούμενη και αποτελεσματική οικογένεια μεθόδων, αυτή που άπτεται του εξελικτικού υπολογισμού και συγκεκριμένα τους εξελικτικούς αλγορίθμους. Ακόμα, αναλύονται μέθοδοι ευρετικών και μετα-ευρετικών τεχνικών, όπως και η δυνατότητα «ύπαρξης δωρεάν γευμάτων».

2.4.1 Μη δωρεάν γεύματα και εξελικτικοί αλγόριθμοι

Τα θεωρήματα «μη ύπαρξης δωρεάν γεύματος» έχουν προσελκύσει το ενδιαφέρον της επιστημονικής κοινότητας και διατηρούν αυτό το ενδιαφέρον αμείωτο. Από την άλλη πλευρά,

αποτελέσαν περιστασιακά το μήλον της έριδος μεταξύ ορισμένων ερευνητών. Τέτοιες συγκρούμενες θέσεις παρατίθενται στο δοκίμιο του Perakh [301]. Σε αυτό το σημείο, αξίζει να σταθούμε σε κάποιες από τις επιφυλάξεις που έχουν διατυπωθεί διαχρονικά. Συγκεκριμένα, ας σημειώσουμε τη θέση του Orr [280] σχετικά με τα θεωρήματα NFL, που παρουσιάστηκαν στο περιθώριο της δημοσίευσης του βιβλίου του William Dembski [91]. Ο Orr δήλωσε ότι:

... μ «μ μ » μ -
 , μ μ μ μ μ ...

Ο Orr υπογράμμισε κάποιες πολύ χρήσιμες παρατηρήσεις σχετικά με τα θεωρήματα NFL σε σχέση με τη Δαρβινική θεωρία, και αυτή ήταν η ουσία της διαφοράς μεταξύ του Orr και του Dembski. Πιο συγκεκριμένα, ο Orr ισχυρίζεται ότι η εξέλιξη, σύμφωνα με τη θεωρία του Δαρβίνου, δε μπορεί να θεωρηθεί ως διαδικασία αναζήτησης, και επομένως, αντίθετα με τον Dembski, δεν μπορεί κανείς να ισχυριστεί ότι ο Δαρβινισμός αποτελεί έναν αλγόριθμο αναζήτησης. Είναι προφανές ότι τα θεωρήματα NFL δεν αποκλείουν την εξελικτική διαδικασία που ορίζεται σύμφωνα με τη δαρβινική θεωρία. Ως εκ τούτου, οι εξελικτικοί αλγόριθμοι μπορούν να χρησιμοποιηθούν κατάλληλα για αναζήτηση και είναι ικανοί να ξεπεράσουν έναν αλγόριθμο τυχαίας αναζήτησης (random search algorithm).

Στην εργασία [301] ο Perakh έδωσε μια δημοφιλή ερμηνεία των θεωρημάτων NFL. Η ερμηνεία αυτή παρουσιάζεται στη συνέχεια, μαζί με ορισμένα χρήσιμα σχόλια και παρατηρήσεις που δόθηκαν από τον συγγραφέα. Ας υποθέσουμε ότι οι A και B είναι δύο αλγόριθμοι αναζήτησης, διερευνώντας τον ίδιο χώρο αναζήτησης. Οι αλγόριθμοι διερευνούν το χώρο αναζήτησης που εφαρμόζονται σε μεταβαλλόμενα σημεία, επιλέγοντας είτε τυχαία σημεία είτε ακολουθώντας μια συγκεκριμένη ακολουθία. Κάθε αλγόριθμος εκτελεί έναν ορισμένο αριθμό κινήσεων. Σε κάθε σημείο που επισκέπτεται ο αλγόριθμος υπολογίζει την αξία της συνάρτησης καταλληλότητας (fitness function) και έτσι, μετά από, π.χ. k βήματα, ο αλγόριθμος παρέχει k μετρήσεις, οι οποίες συνιστούν αυτό που ονομάζεται μ (sample).

Στην ουσία, αυτό το δείγμα δεν είναι παρά ένας πίνακας στον οποίο καταγράφονται οι τιμές της συνάρτησης καταλληλότητας για κάθε σημείο αναζήτησης. Ωστόσο, προκύπτει ένα σημαντικό ερώτημα:

« μ μ ; » μ , μ

Προφανώς, τα δείγματα που υπολογίζονται από δύο αυθαίρετα επιλεγμένους αλγορίθμους δεν αναμένεται να είναι τα ίδια. Αυτό το επιχείρημα μπορεί εύκολα να κατανοηθεί αν το εξετάσουμε από άποψη πιθανοτήτων.

Συγκεκριμένα, ο Perakh σημειώνει ότι:

μ , μ A , μ B , μ (), μ k , μ

Ωστόσο, το πρώτο θεώρημα NFL δηλώνει ότι εάν τα αποτελέσματα αναζήτησης των δύο αλγορίθμων δεν συγκριθούν για ένα συγκεκριμένο χώρο καταλληλότητας αλλά υπολογισθούν κατά μέσο όρο για όλους τους δυνατούς χώρους αναζήτησης, τότε οι παραπάνω πιθανότητες απόκτησης του ίδιου δείγματος είναι ίσες για οποιοδήποτε ζεύγος αλγορίθμων.

Αξίζει να υπογραμμιστεί ότι τα θεωρήματα NFL ισχύουν ανεξάρτητα από το πόσες φορές χρησιμοποιούνται οι αλγόριθμοι για να ολοκληρωθεί η αναζήτηση στον υποκείμενο χώρο του προβλήματος ή το ποιες τιμές της συνάρτησης καταλληλότητας επιστρέφονται από τα διαφορετικά σημεία αναζήτησης. Ένα άλλο σημείο που αξίζει να προσέξουμε, είναι, ότι τα θεωρήματα «μη ύπαρξης δωρεάν γεύματος» δεν παρέχουν κάποιον ισχυρισμό για τη σχετική απόδοση των αλγορίθμων, όπως ορίζεται στην εργασία [301], για ένα συγκεκριμένο χώρο αναζήτησης. Ως αποτέλεσμα, από άποψη απόδοσης, οποιοσδήποτε αλγόριθμος θα μπορούσε να είναι πολύ καλύτερος από οποιονδήποτε άλλο «ανταγωνιστή» του.

Παρά το γεγονός ότι τα θεωρήματα NFL ισχύουν για εξελικτικούς αλγόριθμους, στο άρθρο [422] υποστηρίζεται ότι αυτό δεν μπορεί να ισχύει για την περίπτωση συν-εξελικτικών αλγορίθμων και έτσι, είναι δυνατή η «ύπαρξη δωρεάν γευμάτων». Το πλαίσιο NFL για την περίπτωση συν-εξελικτικών αλγορίθμων, όπως περιγράφεται στο άρθρο [380], δίνεται στη συνέχεια.

Η δήλωση σχετικά με τη μ , που αναφέρεται στις προηγούμενες παραγράφους και τις αναφορές που δίνονται σε αυτή την ενότητα, δεν έχει νόημα χωρίς τον ορισμό του τρόπου με τον οποίο μετράται αυτή η απόδοση. Με άλλα λόγια, ένα πολύ σημαντικό ζήτημα είναι να καθορίσουμε τις μετρικές που πρέπει να χρησιμοποιήσουμε για να μετρήσουμε και να συγκρίνουμε αποτελεσματικά τις επιδόσεις των αλγορίθμων.

Επιπλέον, ενδέχεται να προκύψουν άλλα σημαντικά ερωτήματα, όπως:

- (α) μ - μ NFL;
- (β) μ - μ « μ »;

Σύμφωνα με τη βιβλιογραφία [425], αυτά τα ερωτήματα είναι δύσκολο να απαντηθούν και εξακολουθούν να παραμένουν ανοιχτά προβλήματα.

Ορισμένες πρόσφατες ερευνητικές προσπάθειες σχετικά με τα θεωρήματα NFL και τη βελτιστοποίηση «μαύρου κουτιού» (black-box optimization), έχουν δείξει ότι υπάρχουν συν-εξελικτικά προβλήματα για τα οποία ισχύει η «μη ύπαρξη δωρεάν γευμάτων», ενώ τα «δωρεάν γεύματα» είναι παρόντα στο πλαίσιο άλλων συν-εξελικτικών προβλημάτων. Πιο συγκεκριμένα, στην εργασία [380], οι Service και Tauritz παρουσιάζουν ένα πλαίσιο NFL για κατηγορίες συν-εξελικτικών αλγορίθμων. Αυτό που είναι σημαντικό σε αυτή την εργασία, είναι η ταξινόμηση των συν-εξελικτικών αλγορίθμων με βάση τις λύσεις που αναζητούν. Στο πλαίσιο του συν-εξελικτικού αλγορίθμου που ορίστηκε στην παραπάνω εργασία, ο τύπος των λύσεων ή οι αντίστοιχες λύσεις (individuals), οι οποίες θεωρούνται ως αποτελεσματικές λύσεις στο πρόβλημα, εξαρτώνται αποκλειστικά από τον τύπο του προβλήματος για τον οποίο ο συν-εξελικτικός αλγόριθμος είναι σχεδιασμένος. Αξίζει να σημειωθεί ότι οι διαφορετικές έννοιες λύσεων σχετίζονται με την περίπτωση συνεργατικής συν-εξέλιξης, την περίπτωση ισορροπίας Nash, την περίπτωση *maxmin* κ.λπ.

Οι συγγραφείς ορίζουν τη λεγόμενη μ μ (weak preference relation) που είναι ένας σχετικά απλός τρόπος μέτρησης της απόδοσης των συν-εξελικτικών αλγορίθμων και ως εκ τούτου, αποτελεί μια δυνατή μετρική. Αυτή η μετρική απόδοσης είναι διαφορετική από αυτή που ορίστηκε αρχικά από τους Wolpert και Macready στο έργο τους « μ - μ » (Co-evolutionary Free Lunches) [425].

Το πλαίσιο που αναπτύχθηκε από τους Service και Tauritz μπορεί να θεωρηθεί ως ένας συνδυασμός εννοιών και ορισμών που προέρχονται από δύο θεωρητικά πλαίσια. Το πρώτο πλαίσιο ασχολείται με τα αρχικά θεωρήματα NFL [424] για τους αλγορίθμους αναζήτησης και το άλλο για τους συν-εξελικτικούς αλγορίθμους [128]. Αυτή η σύντηξη γίνεται σε σχέση με τη συνέπεια και των δύο πλαισίων. Επιπλέον, στην εργασία [380] οι συγγραφείς έδειξαν ότι στην συν-εξέλιξη μπορούν να υπάρξουν δωρεάν γεύματα. Κατά συνέπεια, το σημαντικό ερώτημα που παραμένει να απαντηθεί είναι:

« - μ « μ »»

και χρειάζονται περαιτέρω μελέτες για να εξερευνήσουν πρόσθετες κατηγορίες συν-εξέλιξης.

2.4.2 Μη δωρεάν γεύματα και Μετα-ευρετικές Τεχνικές

Είναι γνωστό ότι οι μέθοδοι βελτιστοποίησης με χρήση σμήνους σωματιδίων [296] έχουν συμβάλει σημαντικά στο πεδίο της μαθηματικής βελτιστοποίησης. Αυτές οι μέθοδοι βασίζονται σε σμήνη και αποτελούνται από έναν αριθμό ατόμων που «καθοδηγούν» τη διαδικασία βελτιστοποίησης μέσω της συλλογικής συμπεριφοράς τους, προκειμένου να επιτύχουν μια βέλτιστη λύση. Ένα μεγάλο πλεονέκτημα αυτών των μεθόδων είναι ότι, υπό κατάλληλες συνθήκες και προϋποθέσεις, είναι σε θέση να αποφύγουν τα τοπικά ελάχιστα και να εξασφαλίσουν τη σύγκλιση του αλγορίθμου με κάποια ολική βέλτιστη λύση. Ωστόσο, η ανάλυση της σύγκλισης των αλγορίθμων βελτιστοποίησης που χρησιμοποιούν σμήνη σωματιδίων παραμένει ένας ενεργός ερευνητικός τομέας. Οι πιο σημαντικές μέθοδοι που έχουν αναπτυχθεί και χρησιμοποιούνται ευρύτατα, περιλαμβάνουν, μεταξύ άλλων, τους παρακάτω αλγορίθμους:

μ μ (Particle Swarm Optimization (PSO)) [110],
 μ μ (Ant Colony Optimization (ACO)) [100], μ
 μ (Firefly Algorithm) [429] μ μ (Artificial
 Bee Colony (ABC) algorithm) [208], μ (Bat Algorithm (BA)) [430],
 μ (Cuckoo Search (CS)) [432].

Αυτές οι μέθοδοι, που ονομάζονται επίσης, μετα-ευρετικές τεχνικές (meta-heuristic techniques), περιλαμβάνουν τη διαδικασία (exploration) και μ (exploitation), δύο συγκεκριμένες διεργασίες αναζήτησης, οι οποίες, υπό κατάλληλες συνθήκες, «ελέγχουν» το σμήνος προκειμένου να αποφευχθούν τα τοπικά ελάχιστα της συνάρτησης καταλληλότητας. Οι εφαρμογές των παραπάνω μεθόδων, που βασίζονται σε σμήνη, είναι πολλές και ανήκουν σε διαφορετικούς επιστημονικούς τομείς. Ο αναγνώστης μπορεί να αντλήσει περισσότερες λεπτομέρειες σχετικά με αυτές από την εργασία [135], ιδίως όσον αφορά τις εφαρμογές που άπτονται στον τομέα μηχανικής αλλά και τη βιομηχανία. Τα τελευταία χρόνια, η εφαρμογή των μετα-ευρετικών τεχνικών αυξάνεται συνεχώς και μάλιστα, έχει εισέλθει σε πολλούς διαφορετικούς τομείς, όπως αυτόν της τέχνης [5, 331]. Πιο συγκεκριμένα, έχουν εφαρμοσθεί μετα-ευρετικές τεχνικές στις διεργασίες της μ (crowd simulation), των μ (human swarming), της μ (swarmic art) κ.ά.

Οι μετα-ευρετικές τεχνικές ή μετα-μοντέλα (meta-models), όπως αυτά που προτείνονται στα άρθρα [292-294, 296], χρησιμοποιούνται σε πολλές περιπτώσεις εξελικτικών υπολογιστικών τεχνικών [114, 115, 226], προκειμένου να δημιουργηθούν γρηγορότεροι αλγόριθμοι βελτιστοποίησης. Ειδικά, στις περιπτώσεις όπου τα σύνολα δεδομένων είναι ελλιπή, ή μη ισορροπημένα, ή η αντικειμενική συνάρτηση είναι υπολογιστικά δαπανηρή, οι μετα-ευριστικές διαδικασίες παρέχουν μια εναλλακτική, αποδοτική και αποτελεσματική λύση στο πρόβλημα της βελτιστοποίησης. Συγκεκριμένα, αυτές οι τεχνικές είναι υψηλού επιπέδου ευρετικές διαδικασίες που στοχεύουν στην επιλογή ή τη δημιουργία μετα-ευρετικών μοντέλων αναζήτησης για την αποτελεσματικότερη επίλυση προβλημάτων βελτιστοποίησης. Όπως σημειώνεται στην εργασία [25], υπό ορισμένες ήπιες συνθήκες που σχετίζονται με τις αντικειμενικές συναρτήσεις, οι «αντιπροσωπευτικοί» αλγόριθμοι (surrogate algorithms) επιτυγχάνουν ολική σύγκλιση [24]. Επιπλέον, αυτά τα μετα-μοντέλα δεν «μαστίζονται» από τα «βλαβερά» χαρακτηριστικά των κλασικών μεθόδων βελτιστοποίησης, όπως ο υπολογισμός των παραγώγων. Κατά συνέπεια, τα μετα-μοντέλα ξεπερνούν τις κλασικές μεθόδους, επιτρέποντάς τους να εφαρμοστούν αποδοτικά και αποτελεσματικά σε διάφορες εφαρμογές, όπως [392].

Μια σημαντική συζήτηση σχετικά με τα θεωρήματα NFL και τα μετα-ευρετικά μοντέλα εισάγεται από τον Yang [431]. Σε αυτό το άρθρο ο συγγραφέας σημειώνει ότι τα θεωρήματα

NFL ασχολούνται με τη μέση απόδοση των αλγορίθμων βελτιστοποίησης σε όλα τα υπάρχοντα προβλήματα. Παρ' όλα αυτά, σε πολλά πραγματικά προβλήματα αυτό δεν ισχύει, καθώς οι θεωρητικές απαιτήσεις είναι αυστηρές και δεν μπορούν να εφαρμοστούν στην πράξη. Κατά συνέπεια, η κατάσταση αυτή έχει ως αποτέλεσμα την ύπαρξη «δωρεάν γευμάτων» (free lunches) και αυτό που πρέπει να καθοριστεί είναι η απόδοση συγκεκριμένων αλγορίθμων σε συγκεκριμένες κατηγορίες προβλημάτων. Ως εκ τούτου, σε τέτοιες περιπτώσεις μπορεί να υπάρχουν αλγόριθμοι που είναι σημαντικά καλύτεροι από άλλους για μια συγκεκριμένη κατηγορία προβλημάτων. Αυτό το φαινόμενο, δηλαδή η μη εγκυρότητα των θεωρημάτων NFL, συναντάται συχνά όταν εφαρμόζονται μετα-ευρετικές προσεγγίσεις, καθώς τα πρωταρχικά θεωρήματα NFL αφορούν αλγορίθμους που αναζητούν μεμονωμένες λύσεις. Από την άλλη, οι μετα-ευρετικές προσεγγίσεις που βασίζονται σε πληθυσμούς λύσεων, διερευνούν ταυτόχρονα διαφορετικά μέρη του χώρου αναζήτησης και, υπό αυτήν την έννοια, θεωρούμε ότι ασχολούνται με σύνολα λύσεων. Ως παράδειγμα μπορούμε να εξετάσουμε τις περιπτώσεις γενετικών αλγορίθμων ή του αλγορίθμου PSO. Μια παρόμοια κατάσταση συναντάται και στην βελτιστοποίηση πολλαπλών αντικειμένων (multi-objective optimization), όπου ορισμένοι αλγόριθμοι βρέθηκαν να ξεπερνούν άλλους σε συγκεκριμένα προβλήματα, δημιουργώντας έτσι free lunches [74].

Τα θεωρητικά αποτελέσματα των θεωρημάτων NFL, ενώ είναι πολύ σημαντικά για τη μαθηματική βελτιστοποίηση με σημαντικές θεωρητικές επιπτώσεις, εντούτοις, υποκινούν μια σειρά ερωτήσεων που σχετίζονται με πρακτικές εφαρμογές όπως:

« μ μ μ μ μ NFL;»

Ο Yang [431] παρέχει μια απάντηση σε αυτή την ερώτηση και ταξινομεί τους προγραμματιστές των αλγορίθμων βελτιστοποίησης σε τρεις ομάδες:

- (α) Ένα μεγάλο μέρος των ερευνητών πιστεύει ότι οι προϋποθέσεις που θέτουν τα θεωρήματα NFL δεν μπορούν να εφαρμοστούν στην πράξη και ως εκ τούτου, δεν τις αποδέχονται.
- (β) Οι ερευνητές της δεύτερης κατηγορίας αποδέχονται την εγκυρότητα των θεωρημάτων NFL, αλλά πιστεύουν ότι για συγκεκριμένους τύπους προβλημάτων υπάρχουν βέλτιστοι αλγόριθμοι. Επομένως, εστιάζουν στην εύρεση τέτοιων αλγορίθμων για συγκεκριμένες κατηγορίες προβλημάτων.
- (γ) Η τελευταία ομάδα ισχυρίζεται ότι τα θεωρήματα NFL δεν ισχύουν για συνεχή προβλήματα ή για προβλήματα που ανήκουν στην κλάση NP-hard. Επομένως, εστιάζουν στην εύρεση προβλημάτων για τα οποία δεν ισχύουν τα θεωρήματα NFL και ως εκ τούτου, στον ορισμό «δωρεάν γευμάτων» (free lunches).

Οι επιφυλάξεις και οι διαμάχες που προκαλούνται από τα θεωρήματα NFL οδήγησαν ένα μεγάλο μέρος της επιστημονικής κοινότητας να επανεξετάσει αυτά τα θεωρήματα και να τα επαναδιατυπώσει σε διάφορες ισοδύναμες μορφές. Οι μελέτες που προέκυψαν από αυτή την τάση, οδήγησαν στη δημιουργία πολλών πλαισίων για αλγορίθμους αναζήτησης «μαύρου κουτιού», όπως το πλαίσιο που προτάθηκε από τους Schumacher et al. [343]. Αυτοί οι συγγραφείς μελέτησαν το πρόβλημα και κατέληξαν στο συμπέρασμα ότι τα αποτελέσματα των θεωρημάτων «μη ύπαρξης δωρεάν γεύματος», όπως αρχικά διατυπώθηκαν από τον Wolpert ισχύουν όχι μόνο για το σύνολο όλων των συναρτήσεων, αλλά και για ακόμα μικρότερα σύνολα. Επομένως, τα αποτελέσματα των NFL είναι ανεξάρτητα από το εάν το σύνολο συναρτήσεων είναι συμπιεσμένο ή όχι. Τέλος, οι συγγραφείς του άρθρου καταλήγουν στο συμπέρασμα ότι

τα αποτελέσματα των θεωρημάτων NFL διατηρούνται καλύτερα στην περίπτωση μιας κλειστής μετάθεσης μίας και μόνο συνάρτησης (permutation closure of a single function).

Η πληθώρα των επιστημονικών τομέων όπου εφαρμόστηκαν τα θεωρήματα NFL έκανε όλο και περισσότερους ερευνητές να μελετήσουν και να εφαρμόσουν αυτά τα θεωρήματα που οδήγησαν στην πρόταση διαφόρων επεκτάσεων των θεωρημάτων NFL. Αξίζει να σημειωθεί ότι οι Auger και Teytaud [25] πρότειναν επεκτάσεις των θεωρημάτων NFL που σχετίζονται με άπειρους χώρους, τόσο μετρήσιμους όσο και μη αριθμησίμους. Επιπλέον, μελέτησαν το σχεδιασμό βέλτιστων ευρετικών μοντέλων βελτιστοποίησης. Σύμφωνα με το πρωτότυπο έργο των Wolpert και Macready [424], τα θεωρήματα NFL για βελτιστοποίηση αφορούν πεπερασμένους χώρους αναζήτησης. Έτσι, προκειμένου να επεκταθούν τα θεωρήματα σε άπειρους χώρους αναζήτησης, εισάγονται στοχαστικοί όροι και διαδικασίες. Οι συγγραφείς του άρθρου κατέδειξαν ότι στην περίπτωση άπειρων χώρων αριθμησίμου πλήθους, η φυσική επέκταση των θεωρημάτων NFL δεν ισχύει. Επιπλέον, για την απόδειξή τους, ορίστηκαν μερικές κατανομές των συναρτήσεων καταλληλότητας, οι οποίες οδηγούν σε ίση απόδοση για όλες τις τεχνικές ευρετικής αναζήτησης.

Η παραπάνω απόδειξη οδήγησε σε θεωρήματα «δωρεάν γευμάτων» (Free Lunch) με βάση μια τυχαία συνάρτηση προσαρμογής (καταλληλότητας) και περιλαμβάνει τυχαίους χώρους αναζήτησης. Μια πρόσθετη συμβολή που έγινε στο άρθρο [25] ασχολείται με το σχεδιασμό βέλτιστων αλγορίθμων για τυχαίες συναρτήσεις καταλληλότητας σχετικά με ένα πλαίσιο βελτιστοποίησης «μαύρου κουτιού». Συγκεκριμένα, οι συγγραφείς της εργασίας παρουσίασαν έναν βέλτιστο αλγόριθμο βασισμένο στην «αρχή αποσύνθεσης του Bellman» [34], για έναν ορισμένο αριθμό επαναλήψεων του αλγορίθμου και μια δεδομένη κατανομή της καταλληλότητας. Επιπλέον, για τη διαδικασία σχεδιασμού και τα πειράματα που έγιναν, χρησιμοποιήθηκε ο αλγόριθμος μ Monte-Carlo (Monte-Carlo planning) [217] και ο αλγόριθμος μ (upper confidence tree) [382]. Ακολουθώντας αυτά τα ερευνητικά αποτελέσματα, μπορεί ευλόγως να τεθεί το ερώτημα:

« μ Auger Teytaud μ μ ; »

Στη συνέχεια, για να παρουσιάσουμε μερικά από τα αποτελέσματα των Auger και Teytaud [25], πιο τυπικά, υπενθυμίζουμε την απαραίτητη σημείωση που υιοθετήθηκε στο άρθρο [25]. Έτσι, ως θεωρήσουμε το χώρο αναζήτησης X και Y το πεδίο τιμών για μια δεδομένη αντικειμενική συνάρτηση f . Για κάθε ακέραιο $m \geq 1; 2; \dots; j \in X; j \in Y$ ως θεωρήσουμε ότι $(x_1; x_2; \dots; x_m)$ να είναι ένα διάνυσμα που παίρνουμε μετά από τις πρώτες m επαναλήψεις ενός αλγορίθμου αναζήτησης και $(f(x_1); f(x_2); \dots; f(x_m))$ είναι το διάνυσμα των αντίστοιχων αντικειμενικών τιμών. Η απόδοση ενός αλγορίθμου μ μετά από m επαναλήψεις, δίνεται με τη μέτρηση του διανύσματος των τιμών κόστους που σημειώνονται με $Y(f; m; \mu) = hf(x_1); f(x_2); \dots; f(x_m) i$.

Χρησιμοποιώντας τον προηγούμενο συμβολισμό, τα θεωρήματα NFL υποδηλώνουν τα ακόλουθα αποτελέσματα για πεπερασμένο σύνολο X , Y το αντίστοιχο πεδίο τιμών, δύο αλγορίθμους αναζήτησης μ και b , οποιουδήποτε αριθμού επαναλήψεων m και, τέλος, οποιαδήποτε αντικειμενική συνάρτηση f και ρ κάθε τυχαία κατανομή ομοιόμορφα κατανεμημένη (μεταξύ όλων των κατανομών) πάνω στο σύνολο X και $f \sim \rho$ η σύνθεση των αντίστοιχων συναρτήσεων: τα τυχαία διανύσματα:

$$Y(f \sim \rho; m; \mu) = hf \sim \rho(x_1); f \sim \rho(x_2); \dots; f \sim \rho(x_m) i;$$

και

$$Y(f \sim \rho; m; b) = hf \sim \rho(x_1); f \sim \rho(x_2); \dots; f \sim \rho(x_m) i;$$

ακολουθούν την ίδια κατανομή.

Επιπλέον, αν X είναι ένας άπειρος χώρος αριθμησίμου πλήθους και χωρίς απώλεια της γενικότητας, έστω $X = \mathbb{N}$. Εάν κάποιος είναι σε θέση να παράσχει μια μη τετριμμένη μετρήσιμη αντικειμενική συνάρτηση f , τότε ισχύει η ακόλουθη πρόταση:

Πρόταση 1 $\mu \in NFL(\mathbb{N}; p; f)$ «μ μ»
 $f(i) = (1 - p)^{i+1}$; $\forall i \in \mathbb{N}$:
 $NFL(\mathbb{N}; p; f) = NFL(\mathbb{N}; p; f)$.

Μεταξύ των διαφορετικών θεωρητικών αποτελεσμάτων, μπορεί κανείς να σταθεί στο ακόλουθο θεώρημα «μ» (Continuous Free Lunch), το οποίο θεωρείται ως το κυριότερο αποτέλεσμα των Auger και Teytaud [25].

Θεώρημα 4 (μ), μ $f \in R^{[0;1]}$, μ $GNFL([0;1]; f)$

Στο παραπάνω θεώρημα συμβολίζεται με (X) ένα συνεχές σύνολο και χωρίς απώλεια της γενικότητας θεωρούμε ότι $(X) = [0;1]$ και $U = \mathbb{R}$. Επιπλέον, ο συμβολισμός $GNFL$ χρησιμοποιείται για ένα ασθενέστερο θεώρημα NFL, το οποίο δεν περιορίζει τη συνάρτηση καταλληλότητας στη σύνθετη μορφή $f \cdot p$.

Σημείωση 1 $f \in \mu$, $m \in \mathbb{N}$ (μ) $j \in X$ μ $Y(f; m; b)$

Στον Πίνακα 2.1, παρέχουμε πληροφορίες για τον αριθμό των αναφορών¹ που έλαβαν οι σημαντικότερες εργασίες που αφορούν το πεδίο των θεωρημάτων NFL και τους εξελικτικούς αλγορίθμους για την βελτιστοποίηση. Αυτή η παρουσίαση των αναφορών μπορεί να θεωρηθεί ως πρόσθετη πληροφορία για τη σημασία και τη συμβολή αυτών των άρθρων στον τομέα των θεωρημάτων NFL.

Ο σχεδιασμός ενός αλγορίθμου βελτιστοποίησης που θα είναι πιο αποτελεσματικός από άλλες μεθόδους βελτιστοποίησης είναι μια πολύ δύσκολη διαδικασία και απαιτεί μια σειρά από προϋποθέσεις. Ο «πολυ-διάστατος σχεδιασμός βελτιστοποίησης» (Multidisciplinary Design Optimization) είναι ένα πρόβλημα που βασίζεται στην καλύτερη επιλογή αρχιτεκτονικής. Σε ένα τέτοιο πλαίσιο, μπορεί εύκολα να γίνει κατανοητό ότι η απόκτηση του πιο αποτελεσματικού σχήματος απαιτεί δοκιμές και μπορεί να οδηγήσει σε σφάλματα. Ωστόσο, η διαδικασία δοκιμής και σφάλματος δεν είναι κατάλληλη, καθώς είναι μια δαπανηρά υπολογιστική διαδικασία. Στο άρθρο [396] των Vanaret et al. προτείνεται μια γενική διαδικασία σχεδιασμού που αποφεύγει το παραπάνω πρόβλημα, καθώς και την εγγενή πολυπλοκότητα που υπάρχει σε τέτοιες εφαρμογές.

Στις περισσότερες περιπτώσεις βελτιστοποίησης πολυ-διάστατου σχεδιασμού, η ύπαρξη αποτελεσματικών αλγορίθμων βελτιστοποίησης περιορίζεται σοβαρά από την πολυπλοκότητα της αντικειμενικής συνάρτησης, που οφείλεται στο γεγονός ότι για την εργασία σχεδιασμού χρησιμοποιούνται διαφορετικές αρχιτεκτονικές. Επομένως, είναι πρωταρχικής σημασίας να διατεθεί μια μεθοδολογία που μπορεί να εφαρμοστεί σε όλες τις περιπτώσεις σχεδιασμού και να μετριάσει αυτό το μειονέκτημα. Στο άρθρο [396] αυτό επιτυγχάνεται μέσω μιας

¹πηγή: google scholar, σύμφωνα με μετρήσεις που έγιναν 1/2019

Άρθρο	ΣΠΑ	ΑαΕ
Schumacher et al. (2001) [343]	190	11.18
Droste et al. (2002) [103]	150	9.38
Perakh (2003) [301]	6	0.40
Griffiths & Orponen (2005) [165]	9	0.69
Tauritz (2008) [380]	4	0.40
Poli & Graff (2009) [309]	35	3.18
Auger & Teytand (2010) [25]	70	8.75
Yang (2012) [431]	34	5.67

Πίνακας 2.1: Αριθμός αναφορών που συγκεντρώνουν οι εργασίες που σχετίζονται με τα θεωρήματα NFL και ζητήματα βελτιστοποίησης και εξελικτικών αλγορίθμων όπως παρουσιάζονται στην Ενότητα 2.4. Όπου «ΣΠΑ» δηλώνει το συνολικό πλήθος αναφορών που έλαβε ένα άρθρο, ενώ «ΑαΕ» συμβολίζει κατά μέσο όρο τις αναφορές που λαμβάνει ανά έτος κάθε εργασία.

συνάρτησης αντικατάστασης, που μπορεί να υπολογιστεί πολύ πιο εύκολα από την αρχική. Οι συγγραφείς του άρθρου προτείνουν ένα κλιμακωτό μοντέλο αντικατάστασης, μέσω του οποίου οι αρχιτεκτονικές μπορούν να αξιολογηθούν ευκολότερα και έτσι, η επιλογή γίνεται πιο κατάλληλη. Μέσω των πειραματικών τους αποτελεσμάτων, είναι σαφές ότι η απόδοση ενός μοντέλου αρχιτεκτονικής εξαρτάται σημαντικά από τη διάσταση του αρχικού προβλήματος. Επομένως, όπως δηλώνεται από τα θεωρήματα NFL, δεν υπάρχει αρχιτεκτονική που να είναι σημαντικά πιο αποτελεσματική από όλες τις άλλες, όταν αντιμετωπίζει προβλήματα της ίδιας διάστασης. Οι συγγραφείς για τον σκοπό αυτό, υιοθετούν τα μοντέλα αρχιτεκτονικής «εφικτής πολυ-διάστασης» (Multidisciplinary Feasible) και «εφικτής ατομικής διάστασης» (Individual Disciplinary Feasible), ως πιο αντιπροσωπευτικά μεταξύ διαφορετικών μοντέλων αρχιτεκτονικής για προβλήματα πολυδιάστατης βελτιστοποίησης σχεδιασμού. Παρ' όλα αυτά, στο μέλλον συστήνεται η διευρημένη μελέτη περισσότερων μοντέλων αρχιτεκτονικής.

Στην εργασία των Kimbrough et al. [215] μελετήθηκαν αρκετές περιπτώσεις βελτιστοποίησης με περιορισμούς, χρησιμοποιώντας αλγορίθμους βελτιστοποίησης που βασίζονται στους πληθυσμούς. Συγκεκριμένα, στην έρευνά τους χρησιμοποίησαν γενετικούς αλγορίθμους σχετικά με δύο πληθυσμούς, αυτούς των εφικτών λύσεων και εκείνους που δεν ήταν εφικτοί. Θεωρητικά, σε ένα απλό, τυπικό σχήμα γενετικής εξέλιξης, τα άτομα που αξιολογούνται ως εφικτές λύσεις θα είναι τα μόνα που θα συμμετάσχουν στην εξέλιξη του πληθυσμού και έτσι, στην τελική διαμόρφωση της λύσης. Ωστόσο, αυτή η θεωρητική διάταξη δεν ισχύει στην εργασία [215], καθώς στην προκειμένη περίπτωση, οι συγγραφείς χρησιμοποιούν εφικτές λύσεις για τη βελτίωση των αξιών της αντικειμενικής συνάρτησης, ενώ μη-εφικτές λύσεις χρησιμοποιούνται για τη διόρθωση των ποινών που προκαλούνται από τους περιορισμούς του προβλήματος.

Προκειμένου να εξασφαλιστεί η ομαλότητα της διαδικασίας βελτιστοποίησης, δηλαδή η εξέλιξη των πληθυσμών, οι συγγραφείς καθόρισαν μια μετρική απόστασης μεταξύ των δύο πληθυσμών, τόσο μεταξύ των ατόμων, όσο και των κεντροειδών (centroids) των πληθυσμών. Μια σημαντική λεπτομέρεια που πρέπει να υπογραμμιστεί, είναι ότι, τα κεντροειδή των δύο πληθυσμών πλησιάζουν το ένα με το άλλο κατά τη διάρκεια της εξέλιξης.

Εκ πρώτης όψεως, μπορεί να φαίνεται παράξενο να διατηρηθεί ένας ολόκληρος πληθυσμός «αδύνατων» λύσεων. Ωστόσο, μια πιο προσεκτική ματιά αποκαλύπτει τη χρησιμότητα αυτής της θέσης, δεδομένου ότι αυτός ο πληθυσμός είναι ελεύθερος να μετακινηθεί σε περιοχές του χώρου αναζήτησης, όπου οι εφικτές λύσεις δεν μπορούν συνεπώς να εξερευνήσουν τις περιορισμένες περιοχές αναζήτησης. Οι συγγραφείς του άρθρου μελέτησαν συγκεκριμένα προβλήματα και χώρους αναζήτησης και έδειξαν ότι τα συμπεράσματα των θεωρημάτων NFL

σχετικά με την ισοδυναμία των αλγορίθμων αναζήτησης «μαύρου κουτιού» δεν ισχύουν. Επιπλέον, έδειξαν ότι τα θεωρήματα NFL δεν ισχύουν για προβλήματα με περιορισμούς και συγκεκριμένα, σε πολλά πρακτικά προβλήματα, όπου οι περιορισμοί είναι σταθεροί και προκαθορισμένοι.

Το εξελικτικό υπολογιστικό σχήμα που υιοθετήθηκε από τους συγγραφείς στην εργασία [215], είναι ένας κομψός μηχανισμός που δείχνει ότι υπάρχουν προβλήματα βελτιστοποίησης με περιορισμούς για τα οποία τα αποτελέσματα των θεωρημάτων NFL δεν ισχύουν. Ο αναγνώστης που ενδιαφέρεται να πληροφορηθεί περαιτέρω για το συγκεκριμένο ζήτημα παραπέμπεται στο άρθρο [215].

Ως συμπλήρωμα των παραπάνω ερευνητικών εργασιών, μαζί με τα θεωρητικά ορίσματα και τα συμπεράσματα των συγκεκριμένων προβλημάτων που αναφέρθηκαν παραπάνω, οι Droste et al. [103] παρέχουν κάποιες ρεαλιστικές παρατηρήσεις με βάση την υπολογιστική πολυπλοκότητα αλγορίθμων ευρετικής βελτιστοποίησης. Οι συγγραφείς του άρθρου υποστηρίζουν ότι τα θεωρήματα NFL δεν είναι δυνατά στην περίπτωση της ευρετικής βελτιστοποίησης. Εντούτοις, ένα θεώρημα «σχεδόν μη δωρεάν γεύματος» ((Almost) No Free Lunch ((A)NFL)) δείχνει ότι για κάθε συνάρτηση που μπορεί να βελτιστοποιηθεί αποτελεσματικά από μια ευρετική αναζήτηση, πολλές άλλες σχετικές συναρτήσεις, όπου το ίδιο ευρετικό μοντέλο είναι κακό, μπορούν να κατασκευαστούν. Συνεπώς, οι μέθοδοι ευρετικής αναζήτησης χρησιμοποιούν ορισμένες, εκ των προτέρων γνωστές πληροφορίες, ένα είδος «ιδεών», για το πώς να αναζητούν καλές λύσεις, και έτσι, μπορούν να είναι επιτυχείς μόνο για συναρτήσεις που δίνουν την κατάλληλη «βοήθεια».

Θεώρημα 5

$$\begin{matrix}
 \mu & S & \mu & \mu & & f \\
 \mu & f & 2 & f0;1g^n & & f0;1;2;:::;N & 1g. \\
 & N^{2^2=3} & 1 & & & g, & g: f0;1g! \\
 f0;1;2;:::;Ng & \mu & \mu & f & & 2^{n=3} & S \\
 \mu & & & g\mu & 2^{n=3} & \mu & \mu & \mu \\
 & & & & & & & 2^{n=3}.
 \end{matrix}$$

Αυτό το θεώρημα υποδηλώνει ότι οι ευρετικές μέθοδοι δεν μπορούν να επιτύχουν σε όλα τα υπάρχοντα προβλήματα. Αυτό οφείλεται στο γεγονός ότι η αποτελεσματικότητα αυτών των τεχνικών βασίζεται σε μεγάλο βαθμό σε «καλές μαντεπιές». Εάν αυτή η μαντεπιά είναι καλή, αυτές οι μέθοδοι μπορεί να είναι πολύ αποτελεσματικές. Εάν όχι, ο χρόνος αναζήτησης μπορεί να φτάσει σε εκθετικά επίπεδα και αυτό αποτελεί σοβαρό μειονέκτημα αυτής της οικογένειας μεθόδων.

Στο άρθρο [165], οι Griffiths και Orponen μελέτησαν στρατηγικές βελτιστοποίησης για ένα δεδομένο πεπερασμένο σύνολο συναρτήσεων. Συγκεκριμένα, διερεύνησαν τις συνθήκες που πρέπει να πληρούνται για τις συναρτήσεις που εξετάζονται, ώστε να έχουν τις ίδιες επιδόσεις για ομοιόμορφη κατανομή συναρτήσεων. Το αποτέλεσμα αυτής της έρευνας σχετίζεται με ορισμένες μη τετριμμένες Boolean συναρτήσεις και οριοθετημένους αλγορίθμους αναζήτησης. Ένα σημαντικό συμπέρασμα αυτής της έρευνας είναι ότι, η σχέση των θεωρημάτων NFL και των κλειστών, υπό συνθήκες, μεταθέσεων (closed under permutation conditions), δεν ισχύει πάντα. Αυτό συμβαίνει όταν εξετάζουμε συναρτήσεις που χρησιμοποιούνται για τη μεγιστοποίηση της απόδοσης των αναζητήσεων περιορισμένου μήκους.

Κλείνοντας αυτή την ενότητα, αξίζει να αναφερθεί η συμβολή των Poli και Graff [309] σχετικά με τα θεωρήματα NFL και τις υπερ-ευρετικές τεχνικές (hyper-heuristic techniques). Τα συμπεράσματά τους υποστηρίζουν περαιτέρω τα προηγούμενα αναφερόμενα άρθρα, όσον αφορά τη μη ισχύ των θεωρημάτων NFL και την ύπαρξη «δωρεάν γευμάτων». Τα θεωρήματα NFL εγγυώνται ότι αυτό το φαινόμενο συμβαίνει σε υπερ-ευρετικές τεχνικές και υψηλού επιπέδου υπερ-ευρετικές τεχνικές, αν όλα τα προβλήματα ενδιαφέροντος είναι κλειστά υπό μεταθέσεις (closed under permutation). Για πολλές πραγματικές εφαρμογές, τα αντίστοιχα προβλήματα βελτιστοποίησης δεν πληρούν αυτή την προϋπόθεση και έτσι, σε αυτές τις

Αυτός είναι ο μέσος όρος των φορών σε όλα τα $X \subseteq X$ που τα h και f διαφέρουν σε σχέση με την κατανομή δειγματοληψίας (x) που παρήγαγε τα δεδομένα εκπαίδευσης.

Στη συνέχεια, διατυπώνονται δύο σημαντικά θεωρήματα. Αυτά τα θεωρήματα είναι γνωστά ως « μ μ μ μ μ μ μ μ » (No Free Lunch theorems for supervised learning).

Θεώρημα 6

$$E(Cjd) = \int_{h,f} Er(h; f; d) P(hjd) P(fjd)$$

Οι ακόλουθες εμνηνείες δίνονται από τον Wolpert σε αυτό το θεωρήμα :

- (α) Μια απάντηση στο πόσο καλά ανταποκρίνεται ένας αλγόριθμος μάθησης σε κάποιο πρόβλημα καθορίζεται από το κατά πόσο «ευθυγραμμίζεται» ο αλγόριθμος $P(hjd)$ με το μεταγενέστερο $P(fjd)$.
- (β) Δεν μπορεί κανείς να αποδείξει τίποτα σχετικά με το πόσο καλά γενικεύεται ένας συγκεκριμένος αλγόριθμος μάθησης, καθώς γενικά δεν είναι σε θέση να αποδείξει ότι το $P(hjd)$ ευθυγραμμίζεται με το $P(fjd)$, εκτός και αν το $P(fjd)$ έχει μια συγκεκριμένη μορφή.

Η αδυναμία να αποδείξουμε ότι η $P(fjd)$ έχει μια συγκεκριμένη μορφή, επισημαίνεται από το ακόλουθο θεώρημα.

Θεώρημα 7

$$E_1(Cjf; m) = E_2(Cjf; m) = 0, \quad E_1(Cjf; d) = E_2(Cjf; d) = 0, \quad E_1(Cjm) = E_2(Cjm) = 0, \quad E_1(Cjd) = E_2(Cjd) = 0$$

Σημείωση 2

$$E(Cjd); E(Cjm); E(Cjf; d) = E(Cjf; m)$$

Τα παραδείγματα που δίδονται από τον Wolpert, αφορούν τη διασταυρούμενη επικύρωση και τη Bayesian συμπερασματολογία. Επιπλέον, παρουσιάζονται μερικές παραλλαγές του Θεωρήματος 7 και αναλύονται οι διαισθητικές ιδέες του. Αυτές οι ιδέες οδήγησαν στις ακόλουθες σημαντικές ερευνητικές προσπάθειες σχετικά με δύο κρίσιμα ζητήματα της επιτηρούμενης μάθησης, δηλαδή την και την .

2.5.1 Δεν υπάρχει δωρεάν γεύμα για την έγκαιρη διακοπή

Οι επαναληπτικές μέθοδοι, όπως αυτή της ϵ (gradient descent), εκπαιδεύουν ένα μαθητή (learner) ανανεώνοντας τις ελεύθερες παραμέτρους του μοντέλου, προκειμένου να ταιριάζει καλύτερα στα δεδομένα εκπαίδευσης και να βελτιώνει την απόδοση του εκπαιδευόμενου (learner) σε δεδομένα εκτός του συνόλου εκπαίδευσης. Μέχρις ενός σημείου αυτή είναι μια επιτυχημένη διαδικασία, αλλά πέρα από αυτό το σημείο η περαιτέρω εκπαίδευση οδηγεί σε υπερβολική προσαρμογή στα δεδομένα εκπαίδευσης (το φαινόμενο του λεγόμενου υπερ-ταιριάσματος (over-fitting)), ενώ παράλληλα αποτυγχάνουν να διαχειριστούν δεδομένα εκτός του δείγματος, αυξάνοντας έτσι το σφάλμα γενίκευσης του εκπαιδευόμενου μοντέλου. Οι τεχνικές κανονικοποίησης, συμπεριλαμβανομένης της (early stopping), χρησιμοποιούνται για να αποφευχθεί το υπερ-ταίριασμα.

Η «έγκαιρη διακοπή» παρέχει κανόνες σχετικά με τον τρόπο διεξαγωγής της εκπαίδευσης και το πότε πρέπει να σταματήσουν οι επαναλήψεις, προκειμένου να αποφευχθεί η υπερπροσαρμογή. Στη Μηχανική Μάθηση, η έγκαιρη διακοπή έχει χρησιμοποιηθεί σε πολλές περιπτώσεις και έχει υποστηριχθεί με διάφορα μαθηματικά εργαλεία. Μια πολύ γνωστή και ευρέως χρησιμοποιούμενη τεχνική, είναι να κατευθύνουμε την επικύρωση (validation) μιας διαδικασίας εκπαίδευσης μέσω έγκαιρης διακοπής, παρακολουθώντας την αύξηση του σφάλματος γενίκευσης στα δεδομένα επικύρωσης (validation data).

Στο άρθρο των Cataltepe et al. [58], οι συγγραφείς επιδιώκουν να φέρουν την ιδέα της «μη ύπαρξης δωρεάν γευμάτων» στο πλαίσιο της έγκαιρης διακοπής. Η μέθοδος επιλογής ενός μοντέλου που χρησιμοποιεί την προσέγγιση έγκαιρης διακοπής βασίζεται σε μια ομοιόμορφη επιλογή ενός μοντέλου μεταξύ εκείνων που δίνουν το ίδιο σφάλμα εκπαίδευσης. Αυτή η προσέγγιση ισχυρίζεται ότι είναι παρόμοια με τον «αλγόριθμο Gibbs». Η ομοιόμορφη πιθανότητα επιλογής γύρω από το ελάχιστο σφάλμα εκπαίδευσης είναι ισοδύναμη με τις ισοτροπικές κατανομές των Amari et al. [14], ενώ διαφέρει από αυτή την εργασία, καθώς δε λαμβάνει πολύ μεγάλο αριθμό εκπαιδευτικών παραδειγμάτων. Εκτός από τα γενικά γραμμικά μοντέλα του άρθρου [58], τεκμαίρεται ότι η πιθανότητα επιλογής μοντέλων είναι συμμετρική γύρω από το ελάχιστο σφάλμα εκπαίδευσης.

Αυτή η υπόθεση συμμετρίας είναι μια ασθενέστερη απαίτηση από την ομοιομορφία. Οι συγγραφείς του άρθρου αναλύουν την έγκαιρη διακοπή για κάποιο ελάχιστο σφάλμα εκπαίδευσης. Εάν το σύνολο εκπαίδευσης συνθέτει όλες τις πληροφορίες που έχει κάποιος σχετικά με το στόχο, τότε θα πρέπει να ελαχιστοποιηθεί το σφάλμα της εκπαίδευσης όσο το δυνατόν περισσότερο για να επιτευχθεί χαμηλότερο σφάλμα γενίκευσης. Επιπλέον, οι συγγραφείς καταδεικνύουν ότι, όταν υπάρχουν διαθέσιμες πρόσθετες πληροφορίες, η έγκαιρη διακοπή μπορεί να βοηθήσει συνολικά τη διαδικασία.

2.5.2 Δεν υπάρχει δωρεάν γεύμα για διασταυρούμενη επικύρωση

Στη Μηχανική Μάθηση και γενικά στη θεωρία της στατιστικής μάθησης, η «διασταυρούμενη επικύρωση» (cross-validation) είναι μια μέθοδος αξιολόγησης ενός μοντέλου που χρησιμοποιείται όταν μια προγνωστική διαδικασία μοντελοποίησης ή ένας οποιοσδήποτε μαθητής ή εκπαιδευόμενος καλείται να κάνει νέες προβλέψεις για δεδομένα που δεν έχει ήδη «δει». Αυτά τα δεδομένα αποτελούν το σύνολο επικύρωσης του μοντέλου. Επομένως, αντί να χρησιμοποιείται μαθηματική ανάλυση, η διασταυρούμενη επικύρωση είναι μια γενικά εφαρμόσιμη μέθοδος που χρησιμοποιείται για την εκτίμηση της απόδοσης ενός μοντέλου. Οι συγκεκριμένες μέθοδοι διασταυρούμενης επικύρωσης μπορεί να είναι είτε «εξαντλητικές» (όπως οι τεχνικές leave-p-out και leave-one-out), είτε «μη εξαντλητικές» (όπως οι μέθοδοι k fold και hold out) ή επαναλαμβανόμενες τυχαίες υπο-δειγματοληψίες (random sub-sampling) και όλες είναι σε θέση να δώσουν σημαντικά αποτελέσματα υπό την προϋπόθεση ότι το σύνολο εκπαίδευσης και το σύνολο επικύρωσης προέρχονται από τον ίδιο πληθυσμό, δηλαδή από την ίδια κατανομή.

Η διασταυρούμενη επικύρωση είναι μια στατιστική τεχνική που αποτελεί μια αντικειμενική προσέγγιση για τη σύγκριση διαφόρων διαδικασιών μάθησης, καθώς δεν βασίζεται σε ποσοστά σφάλματος εντός του δείγματος (in-sample error rates). Έτσι, πιστεύαμε για πολύ καιρό, ότι μπορεί να είναι επιτυχής, παρά τις προηγούμενες γνώσεις που υπάρχουν σχετικά με το πρόβλημα. Οι Zhu και Rohwer [445] παρέχουν ένα αριθμητικό αντιπαράδειγμα, το οποίο, παρά το γεγονός ότι είναι τεχνητό, παρέχει μια ελάχιστη απόδειξη του ότι η διασταυρούμενη επικύρωση ως διαδικασία δεν είναι μια «καθολικά ευεργετική μέθοδος». Το πρόβλημα συνίσταται στην επιλογή του κατάλληλου αμερόληπτου εκτιμητή (unbiased estimator) της προσδοκίας μιας Gaussian κατανομής μεταξύ δύο εκτιμητών, και συγκεκριμένα: (α) μια μ (unbiased) και (β) μια τρομερά προκατειλημμένη (highly biased). Οι συγγραφείς του άρθρου εφαρμόζουν την τεχνική leave-one-out και επιχειρούν να δείξουν ότι η μέθοδος αυτή είναι αναποτελεσματική ακόμη και σε μικρά προβλήματα. Ως εκ τούτου, η διασταυρούμενη επικύρωση δεν μπορεί να αψηφήσει το θεωρητικό αποτέλεσμα του θεωρήματος NFL, δηλαδή « μ μ μ μ ».

Επιπλέον, οι συγγραφείς πραγματοποιούν περαιτέρω υπολογιστικά πειράματα και δίνουν μια λεπτομερή ανάλυση, με σκοπό να δώσουν απαντήσεις σε οποιαδήποτε κριτική κατά του κυρίου ζητήματος που αντιμετωπίζει το άρθρο τους, το οποίο είναι « μ μ , μ , μ , μ , bootstrap , μ μ » .» Ως εκ τούτου, εάν χρησιμοποιείται κάποια προηγούμενη γνώση για έναν αλγόριθμο, τότε αυτό θα πρέπει να γνωστοποιείται σε κάθε ενδιαφερόμενο χρήστη, ώστε να μπορεί να αποφασίσει αν θα το χρησιμοποιήσει ή όχι.

Ο Goutte δημοσίευσε μια πιο εξειδικευμένη προσέγγιση για το θέμα αυτό στο άρθρο του [161] με τίτλο «Μια σημείωση για τα δωρεάν γεύματα και την διασταυρούμενη επικύρωση.» Σε αυτή την εργασία, η προαναφερθείσα προσέγγιση των Zhu και Rohwer για τη διασταυρούμενη επικύρωση και το NFL θεώρημα επανεξετάζεται με περαιτέρω επεξεργασία του αριθμητικού παραδείγματος. Ο συγγραφέας, επίσης, εφαρμόζει τις τεχνικές leave-one-out και m fold διασταυρούμενης επικύρωσης στο αριθμητικό αποτέλεσμα που χρησιμοποιούν οι Zhu και Rohwer και εκτελεί μια πιο λεπτομερή μαθηματική περιγραφή. Η ανάλυση των επιτευχθέντων αποτελεσμάτων υποστηρίζει το επιχείρημα ότι «δεν υπάρχει δωρεάν γεύμα για την διασταυρούμενη επικύρωση» και παρότι η μέθοδος δεν είναι η καλύτερη προσέγγιση που μπορεί να εφαρμοστεί για την αξιολόγηση της απόδοσης των εκπαιδευομένων, ωστόσο είναι ικανή να δώσει πολύ καλά αποτελέσματα σε πολλές πρακτικές καταστάσεις.

Εκτός από την παραπάνω ερευνητική εργασία, οι Rivals και Personnaz [327] καταπιάστηκαν με τη δουλειά του Goutte και συγκεκριμένα, με τη χρήση της πιθανοτικής ανάλυσης, εφαρμόζοντας τη leave-one-out διασταυρούμενη επικύρωση ως μέτρο ποιότητας ενός μοντέλου. Τα αποτελέσματα της τεχνικής leave-one-out δείχνουν ότι τα συμπεράσματα του Goutte είναι αισιόδοξα, καθώς αντιμετωπίζουν ένα τριμμένο πρόβλημα για το οποίο οποιαδήποτε «λογική» μέθοδος δεν είναι επιρρεπής σε μια λανθασμένη επιλογή. Επιπροσθέτως, πραγματοποιείται σύγκριση μεταξύ της διασταυρούμενης επικύρωσης με χρήση της leave-one-out τεχνικής και των στατιστικών δοκιμών για την επιλογή γραμμικών μοντέλων. Τα πειραματικά αποτελέσματα που προκύπτουν από ένα συγκεκριμένο επεξηγηματικό παράδειγμα δείχνουν ότι, για τους γραμμικούς εκτιμητές με μεγάλο αριθμό δειγμάτων, η διασταυρούμενη επικύρωση με χρήση της τεχνικής leave-one-out δεν λειτουργεί ικανοποιητικά σε σύγκριση με τις στατιστικές δοκιμές. Αυτό οδηγεί στο συμπέρασμα ότι είναι απίθανο η μέθοδος αυτή να είναι σε θέση να αποδίδει καλά στην περίπτωση των μη γραμμικών εκτιμητών, όπως τα τεχνητά νευρωνικά δίκτυα. Ως εκ τούτου, τονίζεται ένα σημαντικό αποτέλεσμα, δηλαδή ότι οι στατιστικές δοκιμές θα πρέπει να προτιμώνται σε σχέση με την διασταυρούμενη επικύρωση leave-one-out « $(\mu \mu \mu \mu) \mu$ μ .»

2.5.3 Ταξινόμηση προβλημάτων Μηχανικής Μάθησης του πραγματικού κόσμου και μη δωρεάν γεύματα: μια πειραματική προσέγγιση

Η πλειοψηφία της έρευνας σχετικά με τα θεωρήματα «μη ύπαρξης δωρεάν γεύματος» και την επιτηρούμενη μάθηση φαίνεται να είναι περισσότερο ή λιγότερο θεωρητική. Σε αντίθεση με την εργασία στην οποία αναφερθήκαμε προηγουμένως, οι Gómez και Rojas [158] δημοσίευσαν τα αποτελέσματα πολλών υπολογιστικών πειραμάτων που άπτονται της Μηχανικής Μάθησης, με σκοπό να βοηθήσουν στην κατανόηση του αντίκτυπου των θεωρημάτων NFL σε πραγματικά προβλήματα. Ταυτόχρονα, οι συγγραφείς του άρθρου προσπάθησαν να δώσουν επαρκή υπολογιστικά πειραματικά στοιχεία για την εγκυρότητα του θεωρήματος NFL.

Το σύνολο των αλγορίθμων Μηχανικής Μάθησης που χρησιμοποιούνται στα προαναφερθέντα πειράματα περιλαμβάνει:

- (α) Κατηγοριοποιητές της οικογένειας Naive Bayes,
- (β) Δέντρα απόφασης C4.5,
- (γ) Τεχνητά Νευρωνικά δίκτυα,
- (δ) Ταξινομητές της οικογένειας των k πλησιέστερων γειτόνων,
- (ε) Τυχαία δάση C4.5,
- (στ) AdaBoost.M1,
- (ζ) Συνδυαστικές μέθοδοι ταξινόμητων.

Η απόδοση αυτών των προσεγγίσεων εξετάστηκε με βάση τη μέση ακρίβεια σε έξι σύνολα δεδομένων που λήφθηκαν από τη βάση δεδομένων UCI. Συγκεκριμένα, τα σύνολα δεδομένων που χρησιμοποιήθηκαν στην παραπάνω έρευνα είναι: Audiology, Column, Breast cancer, Multiple features (Fourier), German credit και Nursery. Σε μεγάλο βαθμό, τα αποτελέσματα που λήφθηκαν είναι σύμφωνα με προηγούμενες έρευνες. Από την άλλη πλευρά, σύμφωνα με το θεώρημα NFL, οι αλγόριθμοι που συμμετέχουν στον πείραμα, πρέπει να εκθέτουν τον ίδιο βαθμό ακρίβειας. Ωστόσο, αυτό ισχύει όταν υπάρχει διαθέσιμος αρκετά μεγάλος αριθμός συνόλων δεδομένων. Οι συγγραφείς του άρθρου υπογραμμίζουν ότι ορισμένες κοινές υποθέσεις αφορούν τα σύνολα δεδομένων. Αυτές οι κοινές υποθέσεις αφορούν το ξυράφι του Occam και την ανεξάρτητα ταυτόσημη κατανομή των δειγμάτων, καθώς και, κυρίως τις εξαρτώμενες από τα δεδομένα, ιδιότητες που βρίσκονται στα σύνολα δεδομένων, δηλαδή το ντετερμινισμό και την αρχή του Pareto. Με βάση αυτές τις τελευταίες ιδιότητες, εξηγούν τις ιδιαιτερότητες των συνόλων δεδομένων και τα αποτελέσματα που αφορούν την ακρίβεια. Στη συνέχεια, είναι σαφές ότι όλοι οι αλγόριθμοι δεν αποδίδουν εξίσου καλά σε όλα τα προβλήματα.

Εκτός από τα παραπάνω, οι συγγραφείς του άρθρου εκτελούν μια σειρά από πειράματα χρησιμοποιώντας μ μ καθώς και μ . Τα αποτελέσματα που παρατηρήθηκαν δείχνουν ότι οι μηχανές διανυσματικής υποστήριξης ξεπερνούν σε επιδόσεις τους άλλους αλγορίθμους μάθησης, ενώ η απόδοση της βαθιάς μάθησης σε αυτά τα μικρά και σχετικά απλά προβλήματα είναι «απογοητευτική». Στην πραγματικότητα, ενώ αυτές οι αρχιτεκτονικές έχουν σχεδιαστεί για να χειρίζονται σύνθετα σύνολα δεδομένων που έχουν εγγενή πολλαπλά επίπεδα, φαίνεται ότι είναι ανίκανα να αντιμετωπίσουν απλούστερα σύνολα δεδομένων με πιθανότητα «φτωχή» άντληση δεδομένων. Αυτό δείχνει ότι το θεώρημα NFL εφαρμόζεται ακόμη και στην περίπτωση της βαθιάς μάθησης που, επίσης, υπόκειται σε περιορισμούς όπως άλλοι αλγόριθμοι Μηχανικής Μάθησης.

Συμπερασματικά, οι συγγραφείς δηλώνουν ότι:

Μέθοδοι δυναμικών τροχιών αναζήτησης για ολική βελτιστοποίηση

Φτασμένες οι προλήψεις σε μια καθαρότητα μαθηματική,
μας οδηγούν στη βαθύτερη γνώση του κόσμου.

— (1911–1996)

Στο συγκεκριμένο κεφάλαιο, ο αναγνώστης συναντά μια λεπτομερή επισκόπηση των μεθόδων δυναμικής τροχιάς αναζήτησης για ολική βελτιστοποίηση. Επιπλέον, παρουσιάζουμε μια οικογένεια μεθόδων δυναμικής τροχιάς ανίχνευσης οι οποίες δημιουργούνται χρησιμοποιώντας Αριθμητικές μεθόδους Επίλυσης Συνήθων Διαφορικών Εξισώσεων. Περαιτέρω, παρέχουμε μια στρατηγική για την ανάπτυξη μεθόδων ολικής σύγκλισης, η οποία είναι εφαρμόσιμη στην προτεινόμενη οικογένεια μεθόδων και αποδεικνύουμε το αντίστοιχο θεώρημα. Τέλος, δίνονται τα θεωρητικά αποτελέσματα για την επίτευξη μη-μονοτονικών συγκλινόντων μεθόδων που αξιοποιούν τις πληροφορίες σχετικά με τις πιο πρόσφατες τιμές της αντικειμενικής συναρτήσεως.

Το πρόβλημα της μαθηματικής βελτιστοποίησης είναι ένα από τα πιο γνωστά και ευρέως μελετημένα προβλήματα, όχι μόνο από μαθηματικούς αλλά από πολλούς άλλους επιστήμονες σε όλο τον κόσμο. Αυτό συμβαίνει επειδή οι εφαρμογές που απαιτούν τη λύση ενός προβλήματος βελτιστοποίησης είναι πολλές και ποικίλες. Επιπλέον, πολλά σημαντικά προβλήματα βελτιστοποίησης παραμένουν ανοικτά [287]. Ενδεικτικά αναφέρουμε μερικά επιστημονικά πεδία στα οποία συναντώνται συνεχώς τέτοια προβλήματα βελτιστοποίησης, όπως η βιομηχανία, η ρομποτική, τα μαθηματικά, η φυσική, η χημική μηχανική, η οικονομία, η μηχανική μάθηση, η υπολογιστική βιολογία, η γεωλογία και η κλασική μηχανική, μεταξύ άλλων. Αν και η επιστημονική κοινότητα έχει μελετήσει το πρόβλημα της βελτιστοποίησης των αντικειμενικών συναρτήσεων για πολλά χρόνια, το πρόβλημα παραμένει αρκετά ελκυστικό [108, 261, 369].

Σε αυτό το σημείο ας μας επιτρέψει ο αναγνώστης να συζητήσουμε εν συντομία μερικές βασικές έννοιες σχετικά με το πρόβλημα της ολικής βελτιστοποίησης [189]. Το πρόβλημα της ολικής βελτιστοποίησης αφορά τον υπολογισμό του ολικού ελαχίστου ή μεγίστου μιας αντικειμενικής συνάρτησης, η οποία ορίζεται σε ένα δεδομένο σύνολο (πεδίο ορισμού). Συνήθως το πρόβλημα ολικής βελτιστοποίησης αναφέρεται ως πρόβλημα ολικής ελαχιστοποίησης. Έτσι, υποθέτουμε ότι $F : D \rightarrow \mathbb{R} \cup \{\infty\}$ είναι μια μη γραμμική και μη κυρτή, συνεχής συνάρτηση και όπου D είναι ένα συμπαγές σύνολο και όχι κατ' ανάγκη κυρτό. Υποθέτουμε ότι το F^* είναι το ολικό ελάχιστο της συνάρτησης F και $X^* = \{x \in D \mid F(x) = F^*\}$ είναι το σύνολο όλων των ολικών ελαχιστοποιητών της F στο D . Έτσι, το πρόβλημα ολικής ελαχιστοποίησης δίνεται ως $\min_{x \in D} F(x)$, δηλαδή αναζητούμε έναν ολικό ελαχιστοποιητή $x^* \in X^*$. Γενικά, είναι γνωστό ότι το πρόβλημα της ολικής βελτιστοποίησης είναι ένα δύσκολο πρόβλημα, το οποίο μπορεί να κληθεί να αντιμετωπίσει κανείς. Συγκεκριμένα, είναι γνωστό ότι μια ντετερμινιστική μέθοδος για την διαχείριση του προβλήματος βελτιστοποίησης

καθώς και οι εφαρμογές τους, ποικίλουν μεταξύ διαφορετικών επιστημονικών πεδίων.

Σε αυτό το κεφάλαιο, στοχεύουμε να εκτελέσουμε μια λεπτομερή ανασκόπηση των μεθόδων βελτιστοποίησης που χρησιμοποιούν δυναμικές τροχίες αναζήτησης. Η περιγραφή ασχολείται τόσο με τις προηγούμενες όσο και με τις πρόσφατες προσεγγίσεις. Επιπλέον, παρουσιάζουμε μια νέα οικογένεια μεθόδων δυναμικής τροχιάς αναζήτησης που δημιουργούνται χρησιμοποιώντας αριθμητικές μεθόδους για την επίλυση συνήθων διαφορικών εξισώσεων. Επιπλέον, δίνουμε μια στρατηγική για την ανάπτυξη μεθόδων ολικής σύγκλισης, οι οποίες εφαρμόζονται στην προτεινόμενη οικογένεια μεθόδων, καθώς και μια απόδειξη με το αντίστοιχο θεώρημα. Τέλος, δίνουμε τα θεωρητικά αποτελέσματα για την απόκτηση «μη μονότονων» συγκλινόντων μεθόδων που αξιοποιούν τις συσσωρευμένες πληροφορίες σε σχέση με τις πιο πρόσφατες τιμές της αντικειμενικής συνάρτησης.

Ο αναγνώστης στο υπόλοιπο του κεφαλαίου θα συναντήσει αναλυτικότερα τα εξής: Στην επόμενη ενότητα παρουσιάζουμε μια εκτεταμένη επισκόπηση από γνωστές και ευρέως χρησιμοποιούμενες μεθόδους δυναμικών τροχιών που έχουν μελετηθεί στη βιβλιογραφία, όπως η γνωστή μέθοδος δυναμικής αναζήτησης των Snyman-Fatti [359]. Στην Ενότητα 3.3, παρέχουμε τις πιο πρόσφατες μεθόδους δυναμικών τροχιών αναζήτησης και τις εφαρμογές τους, εστιάζοντας στις σύγχρονες μεθόδους που αναπτύχθηκαν τα τελευταία δέκα χρόνια. Επιπλέον, στην ίδια ενότητα, παρέχουμε έναν πίνακα με τις εργασίες που έλαβαν τις περισσότερες αναφορές, προκειμένου να ενημερώσουμε τον αναγνώστη σχετικά με την εφαρμογή και το ενδιαφέρον που συγκεντρώθηκε σχετικά με τις μεθόδους δυναμικών τροχιών. Στην Ενότητα 3.4, παρουσιάζουμε μια νέα οικογένεια μεθόδων δυναμικών τροχιών για ολική βελτιστοποίηση με βάση την αριθμητική επίλυση συνήθων διαφορικών εξισώσεων. Επιπλέον, παρουσιάζουμε μια στρατηγική για την ανάπτυξη μεθόδων ολικής σύγκλισης, καθώς και το αντίστοιχο θεώρημα με την απόδειξή του. Επιπλέον, δίνουμε θεωρητικά αποτελέσματα για την απόκτηση «μη μονότονων» συγκλινόντων μεθόδων. Τέλος, το κεφάλαιο ολοκληρώνεται με μια σύνοψη και μια σύντομη συζήτηση για μελλοντικές ερευνητικές εργασίες.

3.1 Γνωστές μέθοδοι τροχιάς και εφαρμογές τους

Σε αυτήν την ενότητα, παρέχουμε τις πρώτες προσπάθειες που χρησιμοποιούν δυναμικές τροχίες αναζήτησης, προκειμένου να δημιουργηθούν αξιόπιστες μέθοδοι βελτιστοποίησης. Έτσι, για κάθε μέθοδο, παρουσιάζονται οι βασικές εξισώσεις, η βασική σημειογραφία και η κύρια φιλοσοφία πίσω από την παρουσιαζόμενη μέθοδο.

3.1.1 Μέθοδοι δυναμικών τροχιών ανίχνευσης

Μία από τις πρώτες προσπάθειες στις οποίες ο αναγνώστης συναντά το συνδυασμό διαφορικών εξισώσεων με αριθμητικές τεχνικές, έγινε από τον Branin [44]. Στο συγκεκριμένο άρθρο, ο συγγραφέας αντιμετωπίζει το πρόβλημα της εύρεσης λύσης ενός συστήματος μη γραμμικών εξισώσεων. Ειδικότερα, η προτεινόμενη μέθοδος μπορεί να εντοπίσει πολλαπλές ρίζες και, σε ορισμένες περιπτώσεις, να επιτύχει ολική σύγκλιση. Επιπλέον, αν συμβεί μετασηματισμός, η μέθοδός του μπορεί να εντοπίσει, κάτω από ορισμένες συνθήκες, πολλά ακραία σημεία μιας συνάρτησης. Το προτεινόμενο σχήμα έχει δοκιμαστεί σε ένα γνωστό ηλεκτρικό πρόβλημα, που ονομάζεται μ (tunnel diode problem). Η διαφορική εξίσωση που χρησιμοποιήθηκε στην αρχή της προσπάθειάς του ήταν η ακόλουθη:

$$\frac{dx}{dt} + f(x) = 0; \quad (3.1)$$

όπου $f(x) = 0$ αντιπροσωπεύει το σύστημα των μη γραμμικών εξισώσεων που πρέπει να επιλυθεί. Ωστόσο, η παραπάνω εξίσωση αντικαταστάθηκε από την ακόλουθη διαφορική εξίσωση πρώτης τάξης:

$$\frac{df}{dt} + f(x) = 0 \quad (3.2)$$

καθώς η εξίσωση (3.1) θεωρήθηκε ακατάλληλη όσον αφορά άλλα προβλήματα πέραν του προβλήματος διόδου σήραγγας. Όσον αφορά την αναλυτική λύση:

$$f(x(t)) = f(x(0)) \exp(-t) \quad (3.3)$$

είναι απολύτως αποδεκτό ότι $f(x) = 0$, καθώς το $t \rightarrow \infty$. Εφαρμόζοντας τον κανόνα της αλυσίδας και πολλαπλασιάζοντας από τα αριστερά και τις δύο πλευρές της (3.3) με το αντίστροφο του Ιακωβιανού μητρώου, λαμβάνουμε την ακόλουθη εξίσωση:

$$\begin{aligned} \frac{dx}{dt} &= \left(\frac{\partial f}{\partial x} \right)^{-1} f(x) \\ &= J_{f(x)}^{-1} f(x) \end{aligned} \quad (3.4)$$

Εφαρμόζοντας τη γνωστή μέθοδο του Euler, έχουμε το ακόλουθο επαναληπτικό σχήμα:

$$x_{n+1} = x_n - h_n J_{f(x_n)}^{-1} f(x_n); \quad n = 0; 1; 2; \dots \quad (3.5)$$

όπου h_n είναι το μέγεθος βήματος στην n επανάληψη. Αυτή η μέθοδος είναι γνωστή ως μέθοδος Newton (Newton trajectory method). Αξίζει να σημειωθεί ότι, η ακολουθία των τροχιών που δίνεται από αυτή τη μέθοδο έχει νόημα μόνο στην περίπτωση που το Ιακωβιανό μητρώο της αντικειμενικής συνάρτησης δεν είναι μη-ιδιάζον (non-singular). Στην περίπτωση που το Ιακωβιανό μητρώο είναι ιδιάζον (singular), ο Branin στο άρθρο [44] πρότεινε την αλλαγή προσήμου στην εξίσωση (3.2). Έτσι, η αντίστοιχη εξίσωση γράφεται ως εξής:

$$\frac{df}{dt} - f(x) = 0 \quad (3.6)$$

με λύση:

$$f(x(t)) = f(x(0)) \exp(t) \quad (3.7)$$

Επομένως, όταν η ορίζουσα του Ιακωβιανού μητρώου αλλάζει πρόσημο (αφού περάσει από το μηδέν), το πρόσημο στην εξίσωση (3.6) αντιστρέφεται και η διαδικασία ολοκλήρωσης μπορεί να συνεχιστεί. Σε αυτή την περίπτωση παίρνουμε:

$$\frac{dx}{dt} = J_{f(x)}^{-1} f(x); \quad (3.8)$$

Ο Branin εξέτασε την προτεινόμενη μέθοδο στα προβλήματα: (α) δισδιάστατο και τρισδιάστατο τριγωνομετρικό πρόβλημα, (β) το πρόβλημα του Brent και (γ) το πρόβλημα του Rosenbrock. Επεσήμανε ότι το παραγόμενο μονοπάτι τροχιών, σε πολλές περιπτώσεις, περνά μέσα από όλες τις λύσεις που αναζητούσε.

Οι συγγραφείς του άρθρου [199] πρότειναν μια επαναληπτική διαδικασία για την επίλυση μη γραμμικών εξισώσεων. Η προτεινόμενη μέθοδος συνδέει συστήματα συνήθων διαφορικών εξισώσεων με τις εξισώσεις ενός μη γραμμικού συστήματος. Στη συγκεκριμένη εργασία, αξίζει να σημειώσουμε ότι έχουν χρησιμοποιηθεί διαφορετικές εξισώσεις. Αυτές είναι εμπνευσμένες από τους κλασικούς μηχανισμούς της φύσης και είναι, κυρίως, συνήθεις διαφορικές εξισώσεις δεύτερης τάξης. Ένα πλεονέκτημα της παρεχόμενης μεθόδου είναι ότι μπορεί εύκολα να χειριστεί προβλήματα μη γραμμικών ελαχίστων τετραγώνων (problems of nonlinear least squares). Πρώτον, το ζήτημα της επιτυχούς σύγκλισης και, δεύτερον, η επέκταση των περιοχών σύγκλισης είναι πολύ σημαντικές. Στη βιβλιογραφία [44], έχουν

παρουσιαστεί αρκετές μέθοδοι που στηρίζονται σε συστήματα διαφορικών εξισώσεων για την επίλυση μη γραμμικών αλγεβρικών εξισώσεων. Οι συγγραφείς στο άρθρο [199] χρησιμοποίησαν ένα σύστημα διαφορικών εξισώσεων δεύτερης τάξης για να λύσουν ένα μη γραμμικό σύστημα εξισώσεων $F(x) = 0$, όπου $F = (f_1; f_2; \dots; f_n) : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Θεώρησαν τη συνάρτηση $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$, όπου:

$$G(x) = \sum_{i=1}^n f_i^2(x) \tag{3.9}$$

και την ακόλουθη διαφορική εξίσωση δεύτερης τάξης προκειμένου να βρεθούν οι ρίζες της συνάρτησης G :

$$(t) \frac{d^2 x(t)}{dt^2} + g(t)[a(t)I + (1 - a(t))J^T(x(t))J(x(t))] \frac{dx(t)}{dt} = -r G(x(t)) \tag{3.10}$$

όπου (t) και $g(t)$ είναι πραγματικές θετικές συναρτήσεις, $0 < (t) \leq 1$, δt_0 , I είναι το ταυτοτικό μητρώο και $J(x)$ το Ιακωβιανό μητρώο. Η παραπάνω διαφορική εξίσωση δεύτερης τάξης αντιπροσωπεύει το νόμο του Νεύτωνα.

Στη συνέχεια, παρουσιάζουμε ένα πολύ σημαντικό θεώρημα, όπως διατυπώνεται στο άρθρο [199], το οποίο δηλώνει ότι οι λύσεις της διαφορικής εξίσωσης, κάτω από ορισμένες συνθήκες, οδηγούν στις λύσεις x της συνάρτησης G .

Θεώρημα 1 $\mu \quad G(x) \geq C^2(\mathbb{R}^n) \quad (\quad \quad \quad) \quad G(x)$

1. $(t) \geq C^1([0; +\infty))$, $\frac{d}{dt} \leq 0$, $(t) > \min > 0$ $\mu \quad \min$
 2. $g(t) > 0$ $\frac{g(t)}{(t)} hz; (a(t)I + (1 - a(t))J^T(x)J(x))z > \frac{b}{2} hz; z \in \mathbb{R}^n$ $\mu \quad z = 0$
 3. x^2 « $\mu \quad \mu$ » G .
- , $x(t) \quad x^2 \quad v(t) \quad 0 \quad \mu$:

$$(t) \frac{dv(t)}{dt} + g(t)(a(t)I + (1 - a(t))J^T(x(t))J(x(t)))v(t) = -r G(x(t)) \tag{3.11}$$

$$v(t) = \frac{dx(t)}{dt} \quad , \quad \mu \quad \mu \quad \mu$$

Κάποιος μπορεί εύκολα να συμπεράνει ότι το πρόβλημα επίλυσης $F(x) = 0$ είναι ισοδύναμο με την ελαχιστοποίηση της συνάρτησης $G(x)$. Επιπλέον, αυτό το θεώρημα υπογραμμίζει τη σχέση μεταξύ των διαφορικών εξισώσεων, το πρόβλημα ελαχιστοποίησης, την επίλυση ενός συστήματος μη γραμμικών εξισώσεων και την επίλυση ενός προβλήματος μη γραμμικών ελαχίστων τετραγώνων (παραπέμπουμε τον αναγνώστη στην εργασία [337] για λεπτομερέστερες πληροφορίες επί του συγκεκριμένου θέματος). Οι λύσεις της συνήθους διαφορικής εξίσωσης που αφορούν τις συνθήκες του Cauchy:

$$\begin{aligned} x(0) &= x^0 \\ \frac{dx(0)}{dt} &= v^0 \end{aligned} \tag{3.12}$$

λαμβάνονται χρησιμοποιώντας μια τροποποίηση της πολύ γνωστής μεθόδου του Euler. Τα υπολογιστικά πειράματα που διεξάγονται από τους συγγραφείς του άρθρου έχουν δείξει ότι η προτεινόμενη μέθοδος πλεονεκτεί έναντι της γνωστής μεθόδου των Newton-Raphson που έχει χρησιμοποιηθεί στο άρθρο του Branin [44]. Επιπλέον, η προτεινόμενη μέθοδος δοκιμάστηκε σε πέντε προβλήματα, χρησιμοποιώντας τη συνάρτηση του Rosenbrock και το σύστημα εξισώσεων του Boggs [42].

Στο άρθρο [163], ο Griewank παρουσίασε ορισμένες βασικές ιδιότητες των τροχιών ανίχνευσης που πρέπει να πληρούνται. Παρακάτω, παρουσιάζουμε τα θέματα που σχετίζονται με τη συνάρτηση και την τροχιά που εξετάσε στο άρθρο του.

1. Πρώτα απ' όλα, υποθέτουμε ότι η δεδομένη συνάρτηση μπορεί να εκφραστεί ως ένα σύνολο δύο παραγόντων:

$$F = g + \omega \quad (3.13)$$

όπου g είναι μια συνάρτηση με ένα μοναδικό ελάχιστο, έστω $g^* = g(y^*)$ και έστω ότι το ω είναι μια οριοθετημένη διαταραχή.

2. Υποθέτουμε ότι ο παράγοντας ω είναι μια μικρή ποσότητα σε σχέση με το ποσό της τιμής του g .
3. Μπορούμε να φθάσουμε στο ολικό ελάχιστο της συνάρτησης F , έστω $F^* = F(x^*)$ και όλα τα άλλα τοπικά ελάχιστα της F είναι πλησίον του y^* , με άλλα λόγια ανήκουν στη γειτονιά του y^* .

Λαμβάνοντας υπόψη όλες τις παραπάνω περιπτώσεις, κάθε τεχνική που χρησιμοποιείται για την επίτευξη των ολικών ελαχίστων της F είναι δυνατόν να προσεγγίσει πολλά από τα τοπικά ελάχιστα που βρίσκονται έξω από τη γειτονιά του y^* που μας αφορούν.

Στη συνέχεια, δίνουμε τις ακόλουθες επιθυμητές ιδιότητες μιας τροχιάς αναζήτησης.

1. Η τροχιά δεν μπορεί να συγκλίνει στα ελάχιστα με τιμές μεγαλύτερες από το ϵ , έστω ϵ , που είναι ιδανικά ελαφρώς μεγαλύτερο από το ολικό ελάχιστο $F(x^*)$.
2. Όσο η τιμή της F είναι πολύ μεγαλύτερη από τον στόχο ϵ , η τροχιά δεν επηρεάζεται πολύ από τη διαταραχή ω . Ως αποτέλεσμα, η τροχιά πρέπει να ακολουθήσει κατεύθυνση καθόδου σε σχέση με την αρνητική τιμή της κλίσης του g , $-\nabla g$.
3. Καθώς η τιμή της F προσεγγίζει τον στόχο ϵ , η τροχιά ελαχιστοποιείται επαρκώς και τελικά μειώνεται σε μια τοπική τεχνική βελτιστοποίησης όταν $F \leq \epsilon$.
4. Η τροχιά δεν εξαρτάται εξ ολοκλήρου από το Εσσιανό μητρώο της F , $r^2 F$.
5. Τέλος, η τροχιά δεν επηρεάζεται από μετασχηματισμό στις μεταβλητές που ορίζονται παραπάνω και από τον πολλαπλασιασμό της ποσότητας F από οποιοδήποτε θετικό βαθμωτό στοιχείο.

Επομένως, ο Griewank χρησιμοποίησε την ακόλουθη διαφορική εξίσωση δεύτερης τάξης, η λύση της οποίας παράγει μια τροχιά που πληροί τις παραπάνω συνθήκες. Έτσι, στο άρθρο του [163] επαλήθευσε τις παραπάνω συνθήκες για την εξίσωση:

$$\ddot{x}(t) = -\lambda (I - \dot{x}(t)\dot{x}^T(t)) r F(x(t)) = -\lambda F(x(t)) \quad ; \quad \lambda > 0 \quad (3.14)$$

όπου λ είναι ένας θετικός πραγματικός αριθμός. Παραλείποντας την μεταβλητή t , η εξίσωση (3.14) γράφεται ως εξής:

$$\dot{x} = -\lambda (F(x) - F(x_0)) \quad ; \quad \lambda > 0 \tag{3.15}$$

για κάθε αρχική τιμή $(x_0; \dot{x}_0)$, $F_0 = F(x_0) > 0$ και $kx_0k = 1$. Η παραπάνω εξίσωση προέκυψε από μια παραλλαγή μιας προηγούμενης εργασίας του συγγραφέα [164], ως βασική προϋπόθεση για την επίλυση του ακόλουθου προβλήματος:

$$\min_{x(t) \in X} \int_{t_0}^{t_1} (F(x(t))) \lambda dt \tag{3.16}$$

όπου το X είναι μια συνεχής σειρά διαδρομών μεταξύ δύο σημείων $x(t_0) = x_0; x(t_1) = x_1$ και $kxk = 1$. Στο άρθρο [163] ο συγγραφέας επαληθεύει διεξοδικά την εγκυρότητα των παραπάνω υποθέσεων 1-5. Σε αυτή την εργασία, ωστόσο, δεν πραγματοποίησε την αναλυτική επαλήθευση αυτών των συνθηκών. Διάφορες τροχιές έχουν προέλθει από γνωστούς τύπους της Μηχανικής και Φυσικής, όπως η εξίσωση:

$$m(t)\ddot{x}(t) + v(t)\dot{x}(t) = -\lambda F(x(t)) \tag{3.17}$$

η οποία είναι μια από τις πιο γνωστές διαφορικές εξισώσεις (νόμος του Newton), για ένα σωματίδιο μάζας $m(t)$, για μια αντικειμενική συνάρτηση F , σε μια διαλυτική δύναμη (dissipative force) $-v(t)\dot{x}(t)$. Έτσι, είναι εύκολο να παρατηρήσουμε ότι, αν θέσουμε:

$$m(t) = (F(x(t)))^{-1} \tag{3.18}$$

και

$$v(t) = -\lambda F'(x(t))x(t) \tag{3.19}$$

και αν τις αντικαταστήσουμε στη σχέση (3.5), τότε προκύπτει η σχέση (3.2).

Σε αυτό το σημείο, αξίζει να αναφέρουμε ότι μπορεί να προκύψει το ακόλουθο πρόβλημα. Αν οι παράμετροι $m(t)$ και $v(t)$ είναι θετικές, και κάποιες συνθήκες για την F ισχύουν, τότε οποιαδήποτε τροχιά μπορεί να συγκλίνει σε τοπικό ελάχιστο της F . Ο ακόλουθος τύπος που δόθηκε από τον Griewank στο άρθρο [163], υποδηλώνει ότι υπάρχει μια τάση που επιβραδύνει την τροχιά όταν πληρούνται ανοδικές συνθήκες, όπως $\lambda F'(x) \cdot x > 0$. Ο υπό εξέταση τύπος ήταν:

$$(d/dt)[1/2 \dot{x}x(t)\dot{x}^2] = (\lambda F'(x(t))x(t) + v(t)\dot{x}(t)\dot{x}^2) = m(t) \tag{3.20}$$

Από την άλλη πλευρά, καθώς η τροχιά κινείται προς τα κάτω, θέλουμε να επιταχύνουμε το μονοπάτι προς το ελάχιστο της συνάρτησης, μειώνοντας την επιρροή της τοπικής κλίσης. Ωστόσο, είναι δυνατό η τροχιά να παγιδευτεί σε κάποιο τοπικό ελάχιστο. Έτσι, η τροχιά τοποθετείται σε ένα «βαθούλωμα» καθώς οδηγείται σε ένα λόφο, αντί να πέφτει προς μια «πλαγιά». Επομένως, δεν πραγματοποιείται η επιθυμητή μετάβαση προς το ολικό ελάχιστο της συνάρτησης.

Ο συγγραφέας στην εργασία του [163] διαμορφώνει και αποδεικνύει ένα θεώρημα (Θεώρημα 3.1), το οποίο ενισχύει την άποψη ότι η λύση της διαφορικής εξίσωσης (3.2) μετατρέπεται στην τροχιά της πιο απότομης καθόδου. Αυτό συμβαίνει όταν η τιμή της F προσεγγίζει την τιμή του μ . Αυτό επιτρέπει την ελαχιστοποίηση της τροχιάς όσο χρειάζεται και τελικά, δημιουργεί την ακόλουθη επέκταση:

:

$$\dot{x}(t) = -\lambda F'(x) = -\lambda F'(x) \tag{3.21}$$

Οι προκύπτουσες τροχιές αναζήτησης καθορίζονται με μοναδικό τρόπο από τις αρχικές συνθήκες $(x_0; \dot{x}_0)$ και επιλύουν το σύστημα:

$$\max[0; F(x)] \cdot x + \lambda (f(x) - \bar{f}) \cdot F(x) = 0 \quad (3.22)$$

Τέλος, αν οι λύσεις δεν πλησιάζουν το \bar{f} , οι λύσεις του παραπάνω συστήματος συγκλίνουν είτε σε ένα ελάχιστο της F , είτε σε ένα πιθανό σαγματικό σημείο της F . Ειδικότερα, στο έργο του, προσπάθησε να υπολογίσει τα ελάχιστα μιας αντικειμενικής συνάρτησης με πολλά τοπικά ελάχιστα με χαμηλό κόστος. Προκειμένου να επιτευχθεί αυτός ο στόχος, ο συγγραφέας εξέτασε τις τροχιές που παράγονται από τη διαφορική εξίσωση (3.2), σε σχέση με τις παραμέτρους λ και \bar{f} . Μπορεί κανείς να αναμένει ότι οι συνολικές τροχιές θα είναι αποδεκτές. Ωστόσο, αυτό δε θα μπορούσε να είναι αλήθεια. Επομένως, μόνο οι τροχιές που είναι κοντά στις τιμές στόχου \bar{f} είναι «ζωντανές». Αντίθετα, οι τιμές που είναι πολύ υψηλότερες από την τιμή του \bar{f} απορρίπτονται. Ο συγγραφέας θεώρησε ότι είναι δυνατόν να γίνει αυτό (για να παραχθούν «καλές» τροχιές), αρκεί η σχέση μεταξύ της παραμέτρου λ , και της διάστασης της υπό εξέταση συνάρτησης, να προσδιοριστεί σωστά. Επιπλέον, έδωσε μια ανάλυση της τιμής του \bar{f} .

Το πρόβλημα που περιγράψαμε παραπάνω, δηλαδή ότι οι τροχιές μπορούν να «περιπλανηθούν» στις γειτονιές όπου κατοικούν μη αποδεκτά ακρότατα ή να υπερβούν τις αποδεκτές τιμές, είναι εξαιρετικά σημαντικό. Για το λόγο αυτό, ο συγγραφέας αντιμετώπισε αυτό το ζήτημα κατασκευάζοντας μια διαδικασία, η οποία χρησιμοποιεί τις λύσεις της (3.2) για να διεξάγει ολική αναζήτηση. Έτσι, με τον καθορισμό διαδοχικά χαμηλών τιμών για την παράμετρο στόχου \bar{f} , σε συνδυασμό με τις συμβατικές τεχνικές τοπικής ελαχιστοποίησης, ξεπέρασε την προαναφερθείσα δυσκολία. Μετά από κατάλληλη παραμετροποίηση και μετασχηματισμό της (3.2), ο συγγραφέας σχεδίασε μια μέθοδο ενός βήματος (one step method) τάξης σύγκλισης τρία. Αυτή η καινούρια μέθοδος δοκιμάστηκε σε γνωστές συναρτήσεις, όπως (α) η six-hump συνάρτηση, (β) η camel-back συνάρτηση και (γ) μια συνάρτηση τετραγωνικής μορφής. Τα αποτελέσματα έχουν δείξει ότι η μέθοδος του Griewank είναι αρκετά ανταγωνιστική για να υπολογίσει τα τοπικά ελάχιστα. Το υπολογιστικό κόστος ήταν υψηλό (εκτιμήσεις συναρτησιακών τιμών και κλίσεων), αλλά είναι συγκρίσιμο με τις προτεινόμενες τεχνικές που συναντώνταν μέχρι στιγμής.

Στο έργο του Snyman [363], έχει προταθεί μια πρακτική μέθοδος εντοπισμού τοπικών ελαχίστων, κυρίως για συναρτήσεις για τις οποίες η πρώτη παράγωγος μπορεί εύκολα να εκτιμηθεί. Αυτή η μέθοδος έχει ως στόχο τον εντοπισμό των ελαχίστων συναρτήσεων πολλών μεταβλητών. Από τη μία πλευρά, η προτεινόμενη μέθοδος ήταν μια «απλοϊκή» (naive) προσέγγιση, η οποία χρήζει αρκετών βελτιώσεων. Από την άλλη πλευρά, έχει πολλά πλεονεκτήματα, τα οποία αξίζει να συζητηθούν. Η μέθοδος του Snyman συγκρίθηκε με το πρόγραμμα του Fletcher και τη γνωστή μέθοδο quasi-Newton. Τα αποτελέσματα έχουν δείξει ότι η μέθοδος είναι αξιόπιστη και ισχυρή σε ένα σύνολο τυπικών δοκιμαστικών συναρτήσεων. Συγκεκριμένα, προσέγγισε το απαιτούμενο τοπικό ελάχιστο σε όλες τις δοκιμαστικές περιπτώσεις, σε αντίθεση, για παράδειγμα, με τη μέθοδο quasi-Newton, που σε ορισμένες περιπτώσεις απλώς πλησίαζε κοντά στα ελάχιστα. Όπως υποδείχθηκε παραπάνω, η μέθοδος είναι εννοιολογικά πολύ απλή και το πρόγραμμα κωδικοποίησης σε Fortran είναι επίσης αρκετά εύκολο να κατανοηθεί και να χρησιμοποιηθεί. Επιπλέον, οι απαιτήσεις του χώρου αποθήκευσης, οι οποίες ήταν εξαιρετικά σημαντικές εκείνη την εποχή, ήταν ελάχιστες. Συγκεκριμένα, η μέθοδος απαιτούσε την αποθήκευση πέντε n διαστάσιων διανυσμάτων, αντίθετα με της μεθόδου quasi-Newton, που απαιτούσε την αποθήκευση των προσεγγίσεων του Εσσιανού μητρώου (Hessian matrix).

Λαμβάνοντας υπόψη τη συνολική απόδοση αυτής της μεθόδου, θα μπορούσε κανείς να ισχυριστεί ότι αποδίδει ιδιαίτερα καλά σε συναρτήσεις που παρουσιάζουν απότομες στενές «κοιλιάδες». Επιπλέον, η μέθοδος του Snyman ανταποκρίνεται πολύ καλά σε περιπτώσεις όπου το σημείο εκκίνησης δεν είναι κοντά στα τοπικά ελάχιστα. Αντίθετα, σε περιπτώσεις όπου η αντικειμενική συνάρτηση παρουσιάζει επίπεδα ελάχιστα (flat points), η απόδοσή της δεν ήταν τόσο καλή. Παρ' όλα αυτά, η μέθοδος παραμένει ανταγωνιστική και όχι σημαντικά

λιγότερο αξιόπιστη σε σχέση με της quasi-Newton. Στο αρχικό άρθρο που δημοσιεύθηκε το 1982 [363], ο συγγραφέας αποδεικνύει ότι ο υπολογιστικός χρόνος που απαιτείται από τη μέθοδό του για την επίτευξη σύγκλισης αυξάνεται, στο περίπου, γραμμικά. Σε αυτό το κεφάλαιο, ωστόσο, δεν θα εξετάσουμε λεπτομερώς αυτή την απόδειξη. Εντούτοις, αξίζει να σημειωθεί ότι για τις διεξαχθείσες δοκιμές του άρθρου [363], οι εκτιμήσεις των κλίσεων των αντίστοιχων συναρτήσεων ήταν αρκετά «φθηνές» (υπό την άποψη υπολογιστικού κόστους). Αυτό προφανώς ευνοεί τη δυναμική μέθοδο που παρουσιάστηκε από τον Snyman. Οι συναρτήσεις δοκιμής που χρησιμοποιήθηκαν ήταν: (α) η συνάρτηση parabolic valley ($n = 2; 4; 24$), (β) η συνάρτηση cubic valley, (γ) η συνάρτηση του Beale, (δ) η συνάρτηση του Powell, (ε) η συνάρτηση του Wood, (στ) η ομοιογενής τετραγωνική συνάρτηση (homogeneous quadratic) ($n = 40$) και (ζ) η Oren's power ($n = 20$). Η δυναμική μέθοδος του Snyman συγκρίθηκε με:

(α) Το πρόγραμμα του Fletcher, το οποίο ονομάζεται 'VAO9A', και χρησιμοποιεί το γνωστό σχήμα quasi-Newton [131, 156]

(β) Τη μέθοδο Davidon - Fletcher - Powell (DFP) [181].

Το εξεταζόμενο πρόβλημα βελτιστοποίησης ορίζεται παρακάτω:

$$\min F(x); x = (x_1; x_2; \dots; x_n) \in \mathbb{R}^n; F \in \mathbb{C}^1 \quad (3.23)$$

Υποθέτουμε ότι το $F(x)$ αντιπροσωπεύει τη δυναμική ενέργεια ενός σωματιδίου μοναδιαίας μάζας σε ένα n διάστατο συντηρητικό πεδίο δύναμης, τότε, εάν η συνάρτηση F έχει ένα τοπικό ελάχιστο στο x^* , προκύπτει ότι:

$$F(x) = \int_{x^*}^x a(s) ds + F(x^*) \quad (3.24)$$

όπου $a(s)$ αντιπροσωπεύει μια δύναμη που δρα επί του σωματιδίου στο σημείο s . Η κινητική ενέργεια του σωματιδίου ορίζεται από τη σχέση:

$$T(x) = \frac{1}{2} \sum_{i=1}^n \dot{x}_i^2 = \frac{1}{2} \sum_{j=1}^n \dot{x}_j^2 \quad (3.25)$$

και το ανάπτυγμα Lagrangian δίνεται από τον τύπο:

$$L(x) = T(x) - F(x) = \frac{1}{2} \sum_{i=1}^n \dot{x}_i^2 - \int_{x^*}^x a(s) ds + F(x^*) \quad (3.26)$$

Με την εφαρμογή της αρχής του Hamilton λαμβάνονται οι ακόλουθες εξισώσεις κίνησης:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) - \frac{\partial L}{\partial x_i} = 0 \quad (3.27)$$

Με την πραγματοποίηση των κατάλληλων αντικαταστάσεων λαμβάνουμε την ακόλουθη διαφορική εξίσωση:

$$\ddot{x} = -r F = a \quad (3.28)$$

Η τροχιά $x(t)$ του σωματιδίου δίνεται από τη λύση της εξίσωσης (3.28). Συνοψίζοντας, έχουμε το πρόβλημα 3.28 με τις αρχικές συνθήκες:

$$\begin{cases} x(0) = x_0 \\ \dot{x}(0) = v(0) = v_0 = 0 \end{cases} \quad (3.29)$$

όπου v_0 καθορίζει την αρχική ταχύτητα.

Η εξίσωση (3.28) συνεπάγεται τη διατήρηση της ενέργειας. Έτσι, δx , κατά μήκος της τροχιάς παρατηρούμε ότι:

$$T(x) + F(x) = T(x_0) + F(x_0) = E_0 \quad (3.30)$$

όπου E_0 καθορίζει τη συνολική αρχική ενέργεια. Έτσι, για τη συνάρτηση F έχουμε:

$$F(x) = T(x_0) + F(x_0) - T(x) \quad T(x) = F(x_0) - T(x) \quad (3.31)$$

Επομένως, με το $T(x_0) = 0$ κατά μήκος της διαδρομής των σωματιδίων λαμβάνουμε $F(x) \leq F(x_0)$. Ωστόσο, δεδομένου ότι απουσιάζουν οι δυνάμεις τριβής, μπορεί να προκληθεί το ακόλουθο πρόβλημα: το σωματίδιο θα βρίσκεται σε συνεχή κίνηση, και επομένως, υπάρχει ο κίνδυνος να ξεπεραστεί ο ελαχιστοποιητής x^* της συνάρτησης. Συνοψίζοντας, είναι απαραίτητο να διασφαλιστεί ότι η ενέργεια του σωματιδίου μειώνεται. Ως αποτέλεσμα, η τροχιά x ακολουθεί μια διαδρομή προς τον τοπικό ελαχιστοποιητή x^* . Για να επιτευχθεί αυτός ο στόχος, υπάρχουν δύο γνωστές τεχνικές:

1. Η πρώτη χρησιμοποιεί έναν τεχνητό όρο απόσβεσης:

$$\dot{x} = -r F - ax; \quad a > 0 \quad (3.32)$$

όπου $r > 0$ είναι μια δεδομένη σταθερά απόσβεσης. Η επιτυχία αυτής της στρατηγικής εξαρτάται από την κατάλληλη επιλογή της παραμέτρου r . Επομένως, πρέπει να αντιμετωπίσουμε ένα νέο πρόσθετο πρόβλημα:

Κάποιος μπορεί εύκολα να καταλάβει ότι αυτό δεν είναι ένα απλό πρόβλημα και η λύση δεν είναι τετριμμένη.

2. Η δεύτερη στρατηγική που υιοθετείται στο [363], βασίζεται στην παρακολούθηση της κινητικής ενέργειας του σωματιδίου σε καθορισμένα χρονικά διαστήματα. Ας υποθέσουμε ότι x_k, v_k είναι η θέση και η ταχύτητα του σωματιδίου κατά το χρόνο $t_k = k \cdot \Delta t$. Για οποιαδήποτε χρονικά σημεία t_k και t_{k+1} συνδυάζοντας τη σχέση (3.30) λαμβάνουμε:

$$\begin{aligned} F_k &= F_{k+1} - F_k = F(x_{k+1}) - F(x_k) = \\ &= T(x_0) + F(x_0) - T(x_{k+1}) - (T(x_0) + F(x_0) - T(x_k)) \\ &= T(x_0) + F(x_0) - T(x_{k+1}) - T(x_0) - F(x_0) + T(x_k) \\ &= T(x_{k+1}) - T(x_k) = T_{k+1} - T_k = -\Delta T_k \\ &= -\Delta T_k = -\Delta T(x_k) \end{aligned} \quad (3.33)$$

Προφανώς, για $\Delta T_k > 0$ λαμβάνουμε $F_k < 0$ που απαιτείται.

Ο Snyman στο άρθρο [364], πρότεινε μια βελτίωση για να αυξήσει την απόδοση του αρχικού του αλγορίθμου [363]. Η βελτίωση αυτή αφορούσε κυρίως την επιλογή της παραμέτρου του Δt (time step). Συγκεκριμένα, στον προτεινόμενο αλγόριθμο αυτό το βήμα μειώνεται ή αυξάνεται κατάλληλα με αυτόματο τρόπο. Αυτό οδηγεί σε μια πιο αποτελεσματική εφαρμογή του βασικού δυναμικού αλγορίθμου. Αυτή η αλλαγή είχε ως αποτέλεσμα τη δημιουργία ενός πιο εκλεπτυσμένου αλγορίθμου με πολύ καλύτερη απόδοση συγκριτικά με τον αρχικό αλγόριθμο leap-frog, που χρησιμοποίησε στην εργασία [363]. Σε αυτό το κεφάλαιο δεν παρέχουμε διεξοδική ανάλυση αυτών των αλγορίθμων. Για το λόγο

αυτό, ας μελετήσουμε μαζί αυτά τα δύο άρθρα του Snyman προκειμένου να κατανοήσουμε καλύτερα τους βασικούς τους άξονες. Συνοπτικά, παρουσιάζουμε την κύρια διαφορά στην παράμετρο μ του αλγορίθμου.

Είναι γνωστό ότι αν το μ είναι πολύ μεγάλο, τότε η τροχιά μπορεί να είναι ανακριβής. Από την άλλη πλευρά, η τροχιά μπορεί να μην είναι σημαντικά μακριά από τη λύση, αλλά η καταλληλότερη των περιπτώσεων υποδηλώνει ότι η τροχιά είναι όσο το δυνατόν πιο κοντά στο χ^2 για να επιτευχθεί η σύγκλιση του αλγορίθμου. Μια χρήσιμη ιδέα, προκειμένου να ελεγχθεί αν το μ είναι μεγάλο, είναι να ελέγχουμε τη διαφορά μεταξύ δύο διαδοχικών διανυσμάτων των κλίσεων. Αν αυτά τα διανύσματα είναι κάθετα ή σχηματίζουν γωνία 90 μοιρών ή μεγαλύτερη, τότε το βήμα είναι πολύ μεγάλο. Ωστόσο, ένα βήμα δεν θεωρείται ακατάλληλο κατά την πρώτη παρατήρηση μιας τέτοιας γωνίας. Έτσι, σε κάθε βήμα αξιολογούμε την ποσότητα $\hat{a}_k^T a_{k-1}$, όπου a_k υποδηλώνει αρνητική τιμή της κλίσης της συνάρτησης F στο σημείο x_k . Ομοίως, για την ποσότητα του a_{k-1} . Εάν αυτή η ποσότητα είναι μικρότερη ή ίση με μηδέν για διαδοχικά βήματα, τότε το μ πρέπει να μειωθεί στο μισό. Στη συνέχεια, η διαδικασία επανεκκινείται από το σημείο $(x_k + x_{k+1})/2$ με τιμή ταχύτητας $(v_k + v_{k+1})/4$. Τυπικά, ο αριθμός των διαδοχικών βημάτων που επιλέγονται είναι $m = 3$. Από την άλλη πλευρά, υιοθετείται η ακόλουθη στρατηγική εάν χρειάζεται να αυξηθεί το μ . Προκειμένου να αυξηθεί ο χρόνος, πρέπει να διατηρηθούν οι ακόλουθες δύο ανισότητες: $\hat{a}_{k+1}^T a_k > 0$ και $\|x_k\| < \epsilon$, όπου αντιπροσωπεύει το μέγιστο επιτρεπόμενο μέγεθος βήματος. Επομένως, τα βήματα αλλάζουν σύμφωνα με τη σχέση:

$$\rho = 1 + N ; \quad \epsilon > 0 \quad (3.34)$$

όπου το N αντιπροσωπεύει τα διαδοχικά επιτυχημένα βήματα που έχουν πραγματοποιηθεί. Διαφορετικά, το N επαναφέρεται στην τιμή 1 και δεν αυξάνεται.

Μια από τις πιο γνωστές μεθόδους αυτής της κατηγορίας, είναι η μέθοδος που προτείνεται από τους Snyman και Fatti στο άρθρο [359]. Αυτή η μέθοδος έχει προσελκύσει μεγάλο ενδιαφέρον από πολλούς ερευνητές. Για το λόγο αυτό, έχουν παραχθεί πολλές παραλλαγές αυτής της μεθόδου, τις οποίες θα παρουσιάσουμε εν συντομία σε αυτό το κεφάλαιο. Στη συνέχεια, παρουσιάζονται τα κύρια σημεία της αρχικής μεθόδου. Οι Snyman και Fatti παρήγαγαν μια μέθοδο ολικής ελαχιστοποίησης χρησιμοποιώντας έναν αλγόριθμο πολλαπλών εκκινήσεων (multi-start algorithm). Η επιτυχία αυτής της μεθόδου ήταν ότι με κατάλληλες τροποποιήσεις στις τροχιές αναζήτησης, οι συγγραφείς κατάφεραν να διευρύνουν τις περιοχές σύγκλισης όσον αφορά το ολικό ελάχιστο της αντικειμενικής συνάρτησης. Η διαφορική εξίσωση δεύτερης τάξης που χρησιμοποιήθηκε στην προσέγγισή τους είναι η ακόλουθη:

$$\ddot{x}(t) = -\gamma F(x(t)) \quad (3.35)$$

όπου η λύση x αντιπροσωπεύει την κίνηση ενός σωματιδίου μοναδιαίας μάζας μέσα σε ένα συντηρητικό πεδίο δύναμης, το F αντιπροσωπεύει τη δυναμική ενέργεια αυτού του σωματιδίου και οι σχέσεις $x(0) = x_0$, $\dot{x}(0) = 0$ τις αρχικές συνθήκες. Αν πολλαπλασιάσουμε την παραπάνω εξίσωση με $\dot{x}(t)$ λαμβάνουμε τον τύπο της κινητικής ενέργειας:

$$\frac{1}{2} k \dot{x}(t)^2 + F(x(t)) = F(x_0) \quad (3.36)$$

για $t = 0$. Επιπλέον, εάν πληρούται η ακόλουθη σχέση:

$$-\gamma F' x > 0 \quad (3.37)$$

τότε η κινητική ενέργεια του σωματιδίου αυξάνεται, πράγμα που σημαίνει ότι η τιμή της αντικειμενικής συνάρτησης μειώνεται μέσω της πιο απότομης διαδρομής καθόδου (steepest descent path). Όπως έχουμε ήδη αναφέρει, είναι πιθανό η υπό εξέταση συνάρτηση να έχει πολλά τοπικά ελάχιστα, τα οποία, στην πραγματικότητα, μπορούν να εντοπιστούν στο σύνολο

όπου x_0, Δ_0 είναι οι αρχικές συνθήκες και t το χρονικό βήμα. Όπως έχει αποδειχθεί λεπτομερώς στο άρθρο [363], αυτή η μέθοδος δίνει τη σχέση διατήρησης της ενέργειας. Επιπλέον, η αρχική συνθήκη Δ_0 δίνεται από τη σχέση:

$$\Delta_0 = \frac{r F(x_0) t}{2} \tag{3.43}$$

όπου η πιο συνηθισμένη επιλογή είναι η τιμή $\alpha = 2$. Όσον αφορά το χρονικό βήμα t , επιλέγεται έτσι ώστε να εξασφαλιστεί κάθοδος στο πρώτο βήμα. Αν αυτό δεν συμβεί, τότε απαιτείται μια επιλογή διαφορετικής τιμής βήματος. Συνήθως, το χρονικό βήμα διαιρείται και η τροχιά επανεκκινείται μέχρι να φτάσει σε μια πορεία καθόδου. Η μέθοδος τερματίζεται όταν ισχύει η ακόλουθη σχέση:

$$F(x(t)) - F_n > (F(x_0) - F_n) \tag{3.44}$$

όπου $x(t)$ καθορίζει την τροχιά, F_n είναι η τρέχουσα ελάχιστη τιμή της F και το α αποτιμάται συνήθως ως $\alpha < 1$ και $\alpha \geq 1$. Στο άρθρο [359], η παράμετρος α υποδεικνύεται ως $\alpha = 0.95$. Επιπλέον, αν πληρούται η ακόλουθη σχέση:

$$\frac{1}{2} \sum_{j=1}^n x_j^2 < (1 - \alpha)(F(x_0) - F_n) \tag{3.45}$$

τότε, η τροχιά τερματίζεται σε ένα ανηφορικό μονοπάτι. Επιπλέον, η διαδικασία ελαχιστοποίησης ολοκληρώνεται, εάν για κάθε $x(t)$ πληρούται η ακόλουθη σχέση: $k r F(x_n(t)) k < \epsilon$, όπου ϵ είναι μικρή θετική τιμή. Για να περιγράψουμε τη μέθοδο αυτή, έχουμε αναφέρει ότι ο ολικός στόχος του αλγορίθμου περιλαμβάνει έναν στοχαστικό όρο ο οποίος εμπνεύστηκε από την εργασία [326]. Συγκεκριμένα, αφορά την πιθανότητα της τελευταίας επιτευχθείσας τροχιάς που είναι επιθυμητή, ώστε να επιτευχθεί το ολικό ελάχιστο. Η παρουσίασή μας για τη μέθοδο των Snyman-Fatti ολοκληρώνεται, παρέχοντας το αντίστοιχο θεώρημα [359].

Θεώρημα 3

Έστω k, μ, F, R, p_i, n οι παραμέτρους της μεθόδου, όπου k, μ, F, R, p_i είναι θετικοί αριθμοί και n ολικός αριθμός. Τότε, η πιθανότητα να επιτευχθεί το ελάχιστο είναι:

$$p^? = Pr[R^?] = \max_i p_i \tag{3.46}$$

$$Pr[F = F^?] > q(n; k) = 1 - \frac{(n + 1)!(2n - k)!}{(2n + 1)!(n - k)!} \tag{3.47}$$

Όπου $R^?, p_i$ είναι οι παραμέτρους της μεθόδου, $F^?, p_i$ είναι οι παραμέτρους της μεθόδου, R_i είναι οι παραμέτρους της μεθόδου, $p^?$ είναι οι παραμέτρους της μεθόδου.

Η απόδειξη του θεωρήματος παρουσιάζεται λεπτομερώς στο άρθρο [359], αλλά δε θα μας απασχολήσει στο συγκεκριμένο κεφάλαιο. Όσον αφορά τα πειραματικά αποτελέσματα που έλαβαν οι Snyman και Fatti, αξίζει να σημειωθεί ότι χρησιμοποιήθηκαν εννέα γνωστές συναρτήσεις [97, 163, 314]. Αξιοσημείωτο ήταν ότι η προτεινόμενη μέθοδος των Snyman-Fatti, επέστρεψε με επιτυχία το ολικό ελάχιστο για όλες τις δοκιμαστικές συναρτήσεις που εξετάστηκαν στη μελέτη τους.

$$\begin{aligned} X &= f(x) \in \mathbb{R}^n : x' \in \mathbb{R}^n, x \in \mathbb{R}^n \\ Y &= f(y) \in \mathbb{R}^m : y' \in \mathbb{R}^m, y \in \mathbb{R}^m \end{aligned}$$

$$x' \in \mathbb{R}^n, y' \in \mathbb{R}^m \quad x' < x^h, y' < y^h.$$

Σύμφωνα με τους συγγραφείς του άρθρου, οι παράμετροι $x; y$ ανανεώνονται ανάλογα με τις λύσεις των διαφορικών εξισώσεων. Επιπρόσθετα, έχει γίνει η ακόλουθη παραδοχή: μια min-max λύση της F υπάρχει και είναι μοναδική, καθώς κατάλληλοι περιορισμοί σχετικά με τις μεταβλητές απόφασης έχουν πραγματοποιηθεί [390]. Αξίζει να σημειωθεί ότι οι πληροφορίες της κλίσης, όπου ήταν απαραίτητο να υπολογιστούν, έγιναν με αριθμητικό τρόπο από τους προτεινόμενους αλγορίθμους. Στο πρώτο πρόβλημα, και οι τρεις αλγόριθμοι βρήκαν εύκολα την επιθυμητή λύση. Οι διαφορές που παρατηρήθηκαν στις τροχιές, ωστόσο, υποδηλώνουν την ανάγκη για περαιτέρω μελέτη των συσχετισμών στα αποτελέσματα που ελήφθησαν. Στο δεύτερο πρόβλημα, οι συγγραφείς σημειώνουν ότι η σύγκλιση του αλγορίθμου μπορεί να μην οδηγήσει σε μια καλή λύση. Στο τρίτο πρόβλημα, έδειξαν ότι ο σωστός προσδιορισμός των συνθηκών διαδραματίζει σημαντικό ρόλο. Ο κακός προσδιορισμός οδηγεί σε μη σύγκλιση του αλγορίθμου. Τέλος, μέσω της λύσης του τελευταίου προβλήματος, έδειξαν ότι οι τροχιές μπορούν να κινούνται στα όρια του συνόλου των περιορισμών.

Ένας νέος αλγόριθμος, που ονομάζεται αλγόριθμος DYNAMIC-Q, παρουσιάζεται στο άρθρο [362]. Αυτή η μέθοδος ήταν ένας εύχρηστος και αξιόπιστος αλγόριθμος αντιμετώπισης προβλημάτων βελτιστοποίησης με περιορισμούς (constrained optimization problems) [313]. Είναι ουσιαστικά μια παραλλαγή ενός υπάρχοντος αλγορίθμου [363, 364], που βασίζεται σε δυναμικές τροχιές αναζήτησης, οι οποίες αντιμετώπισαν προβλήματα βελτιστοποίησης χωρίς περιορισμούς. Έτσι, ο προαναφερθείς αλγόριθμος τροποποιήθηκε με τέτοιο τρόπο, ώστε να χειριστεί τους περιορισμούς του προβλήματος. Συγκεκριμένα, εισάγεται μια δυναμική παράμετρος ποινής στον αλγόριθμο, και ως αποτέλεσμα μπορεί να επιλύσει ένα πρόβλημα βελτιστοποίησης με περιορισμούς [313]. Αυτή η νέα μέθοδος έχει εφαρμοστεί σε δομικά προβλήματα βελτιστοποίησης (structural optimization problems), όπως δομές ελάχιστου βάρους δοκών και πλαισίων. Τα προβλήματα βελτιστοποίησης αυτού του είδους είναι πολύ δύσκολο να επιλυθούν, δεδομένου ότι πρέπει να ληφθεί υπόψη ένας μεγάλος αριθμός παραμέτρων, όπως η τάση, η μετατόπιση, το βάρος κλπ. Για το λόγο αυτό, η άμεση εφαρμογή των μεθόδων δυναμικής τροχιάς δε θα ήταν αποτελεσματική. Έτσι, η προτεινόμενη μέθοδος εφαρμόστηκε για την επίλυση υποπροβλημάτων και, ως εκ τούτου, η λύση του αρχικού προβλήματος καθίσταται σημαντικά λιγότερο δαπανηρή. Λεπτομερέστερα, η συνάρτηση ποινής που εμφανίζεται στον αλγόριθμο DYNAMIC-Q διαμορφώνεται ως εξής:

$$P(x) = F(x) + \sum_{i=1}^m a_i G_i^2(x) + \sum_{j=1}^s b_j H_j^2(x) \tag{3.48}$$

όπου $a_i = 0$ εάν $G_i(x) \leq 0$, αλλιώς $a_i = \rho_i$ εάν $G_i(x) > 0$. Σε πολλές περιπτώσεις, για ευκολία λαμβάνεται $\rho_i = b_j = \rho > 0$, όπου ρ δηλώνει ένα μεγάλο αριθμό. Υπό ορισμένες συνθήκες, το ελάχιστο του $P(x)$ συμφωνεί με το ελάχιστο του $F(x)$ σχετικά με το ακόλουθο πρόβλημα ελαχιστοποίησης με περιορισμούς:

Πρόβλημα 3 μ ρ , $F(x)$, $x \in \mathbb{R}^n$ μ $\min F(x)$, ρ :

$$\begin{aligned} G_i(x) &\leq 0, i = 1; 2; \dots; m \\ H_j(x) &= 0, j = 1; 2; \dots; s \end{aligned}$$

$$F; G \quad H \quad \mu \quad \rho \quad (\text{scalar functions}) \quad x.$$

Δεδομένου ότι οι μεγάλες τιμές της παραμέτρου ρ μπορούν να δημιουργήσουν ένα δύσκολο, για χειρισμό, πρόβλημα, προτάθηκε η ακόλουθη στρατηγική. Η παράμετρος ποινής

είναι περισσότερο συμφέρουσα να αυξάνεται σταδιακά, καθώς η τροχιά εξελίσσεται, μέχρι να επιτευχθεί μια προκαθορισμένη τιμή. Επομένως, αρχικά θέτουμε $\epsilon = 0$, ορίζουμε έναν παράγοντα c (ο οποίος είναι συνήθως ένας μικρός αριθμός κοντά στο 1) και επίσης μια μέγιστη τιμή k_{\max} . Έτσι, σε κάθε βήμα, θέτουμε $k = c$ μέχρι να ικανοποιηθεί η ανισότητα $k > k_{\max}$. Στη συνέχεια, θέτουμε $k = k_{\max}$ και συνεχίζουμε τη διαδικασία έως ότου επιτευχθεί σύγκλιση. Οι συντάκτες του άρθρου [362], παρατήρησαν ότι η υιοθέτηση αυτής της στατηγικής παράγει μια καλύτερη τροχιά. Επιπλέον, επιτυγχάνεται μια ταχύτερη σύγκλιση προς την περιοχή του x^* , σε σύγκριση με τη μη σταδιακή αύξηση της παραμέτρου ϵ . Το δύσκολο ζήτημα που δημιουργείται και καθιστά το πρόβλημα υπολογιστικά δαπανηρό, είναι το ακόλουθο. Οι συγγραφείς αντιμετωπίζουν ένα δομικό πρόβλημα βελτιστοποίησης που απαιτεί την εκτίμηση της κλίσης, τόσο για τις αντικειμενικές συναρτήσεις, όσο και για τις συναρτήσεις περιορισμού. Αυτό αυξάνει το υπολογιστικό κόστος του προβλήματος. Μια λύση στο πρόβλημα αυτό δίνεται από μια σειρά τετραγωνικών υπο-προβλημάτων. Συνοπτικά, χρησιμοποιείται ο ακόλουθος τύπος περιορισμού:

$$G_j(x) = G_j(x_k) + \epsilon G_j^>(x_k)(x - x_k) + m_j^k k x - x_k k^2; j = 1; 2; \dots; m \quad (3.49)$$

όπου m προσδιορίζει τον αριθμό των περιορισμών και ο συντελεστής m_j^k πρέπει να υπολογιστεί κατάλληλα για κάθε περιορισμό. Ο τελευταίος τύπος με την αντικειμενική συνάρτηση $F(x)$ αποτελεί το υποπρόβλημα. Το κριτήριο τερματισμού που υιοθετήθηκε στη διαδικασία αυτή ήταν το ακόλουθο:

$$\frac{jF(x_{k+1}^?) - F(x_k^?)j}{F(x_{k+1}^?)^2} \leq \epsilon_r \quad (3.50)$$

όπου $x_{k+1}^?; x_k^?$ δηλώνουν τις λύσεις των τελευταίων δύο υποπροβλημάτων και ϵ_r αντιπροσωπεύει μια προκαθορισμένη μικρή, θετική παράμετρο. Αξίζει να σημειωθεί ότι, αν θέσουμε $m_j^0 = 0$, τότε το πρόβλημα μετατρέπεται σε ένα γραμμικό πρόβλημα προγραμματισμού. Οι συγγραφείς του άρθρου, σε όλες τις πειραματικές δοκιμές που διεξήγαγαν, πρότειναν την τιμή $m_j^0 = 0.01$. Η προτεινόμενη μέθοδος δοκιμάστηκε σε 10 προβλήματα: 5 προβλήματα συρματοπλέγματος και 5 προβλήματα πλαισίου. Επιπλέον, τα προβλήματα ήταν διαφορετικά όσον αφορά τη διάσταση και τον αριθμό των μεταβλητών. Ο αλγόριθμος DYNAMIC-Q τέθηκε σε σύγκριση με τη μέθοδο μ (sequential quadratic programming (SQP) method). Τα αποτελέσματα έδειξαν ότι το προβλεπόμενο πρόγραμμα ήταν ανταγωνιστικό, αξιόπιστο (κυρίως για υπολογιστικά δαπανηρά προβλήματα) και έλυσε τα προβλήματα γρήγορα και αποτελεσματικά.

Στην εργασία [339], οι συγγραφείς παρουσίασαν μια μέθοδο τροχιάς για την επίλυση προβλημάτων βελτιστοποίησης χωρίς περιορισμούς και συγκεκριμένων μη γραμμικών προβλημάτων προγραμματισμού χωρίς περιορισμούς. Αξιοποιώντας κάποιες ιδιότητες των αντικειμενικών συναρτήσεων, και ένα ειδικό σύστημα συνήθων διαφορικών εξισώσεων, παρήγαγαν τροχιές που έδωσαν πολύ καλά αποτελέσματα. Επιπλέον, ένα σημαντικό πλεονέκτημα της προσέγγισής τους είναι η χρήση πληροφοριών της κλίσης της συνάρτησης και όχι η τιμή της ίδιας της αντικειμενικής συνάρτησης. Επιπλέον, αξίζει να επισημανθεί ότι το υπολογιστικό κόστος ήταν πολύ υψηλό, καθώς η προτεινόμενη μέθοδος δεν εκτιμούσε μεγάλης τάξης παραγώγους. Στη συνέχεια, παρουσιάζουμε το πρόβλημα αρχικών τιμών [49, 50], το οποίο χρησιμοποίησαν για την παραγωγή των τροχιών, καθώς και το πρόβλημα που προσέγγισαν. Το μη γραμμικό πρόβλημα χωρίς περιορισμούς ήταν της ακόλουθης μορφής:

$$\begin{aligned} \text{Πρόβλημα 4} \quad & \mu \quad F : \mathbb{R}^n \rightarrow \mathbb{R}, F \in \mathcal{C}^2(\mathbb{R}^2) \quad S_a = \{x : F(x) \leq \\ & F(a) \leq \mu, \delta a \in \mathbb{R}^n. \quad \mu \quad \min_x F(x) \\ & \mu \quad \mu : \end{aligned}$$

$$x(t) = \frac{r F(x(t))}{kr F(x(t))k_2^2 s(t)} \tag{3.51}$$

$$x(0) = x_0$$

$$t > 0, s \in C^1 \quad r F(x(0)) \neq 0.$$

Στη συνέχεια, παρουσιάζουμε μερικά πολύ ενδιαφέροντα θεωρήματα, τα οποία αποδεικνύουν σημαντικές ιδιότητες της μεθόδου που προτείνουν οι συγγραφείς στο άρθρο [339].

Θεώρημα 4 $\mu \quad I = [0; t] \quad \mu \quad \mu \quad \mu \quad \mu$
 $\mu \quad \mu \quad 3.51 \quad \mu \quad \mu \quad \mu \quad t = [0; t] \quad r F(x(t)) = 0. \quad \mu$

Θεώρημα 5 $F(x(t)) = F(x_0) \int_0^t s(\cdot) d = G(t)$

Το Θεώρημα 4, δείχνει την ύπαρξη και τη μοναδικότητα της λύσης του προβλήματος αρχικών τιμών (3.51), ενώ το Θεώρημα 5, υποδηλώνει ότι η συνάρτηση G μπορεί να βρεθεί για διαφορετικές επιλογές του S . Επιπλέον, αν και οι λύσεις του προβλήματος (3.51) είναι άγνωστες, οι συναρτησιακές τιμές της F μπορούν να δοθούν χρησιμοποιώντας τη συνάρτηση G . Αυτό το στοιχείο χρησιμοποιείται από τους συγγραφείς για να υποδείξει ότι μπορεί να πραγματοποιηθεί ο έλεγχος μεγέθους βήματος. Επιπλέον, ο αναγνώστης παραπέμπεται στο άρθρο [339], για να μελετήσει εναλλακτικές διαφορικές εξισώσεις με ισοδύναμα αποτελέσματα πέραν από το πλεονέκτημα της συνάρτησης G . Οι συγγραφείς του άρθρου, μέσω του επόμενου θεωρήματος, έδειξαν ότι η γνωστή μέθοδος απότομης καθόδου με βήμα μεγέθους των Armijo-Goldstein, είναι ισοδύναμη με τη μέθοδο του Euler για $s \leq 1$, χρησιμοποιώντας τη συνάρτηση G για τον έλεγχο μεγέθους του βήματος. Συγκεκριμένα, αν το h υποδηλώνει το μέγεθος του βήματος, τότε μπορεί να επιλεγεί χρησιμοποιώντας τις πληροφορίες της G με τον ακόλουθο τρόπο: Για μια παράμετρο ϵ , όπου $0 < \epsilon < 1$, μπορεί να υπάρχει ένα μέγεθος βήματος $h > 0$, έτσι ώστε:

$$F(x_{new}(h)) - F(x(0)) \leq [G(h) - F(x(0))] \tag{3.52}$$

όπου το $x_{new}(h)$ υπολογίζεται χρησιμοποιώντας τη μέθοδο του Euler που εφαρμόζεται στο πρόβλημα αρχικών τιμών 3.51. Επιπλέον, έδωσαν το ακόλουθο θεώρημα:

Θεώρημα 6 $\mu \quad \mu \quad \text{Euler} \quad \mu \quad \mu \quad \mu$
 (3.51). $\mu \quad \mu \quad \mu \quad \mu \quad 4 \quad s \in C^1, s(t) > 0,$
 $\partial t \geq t \quad \mu \quad \mu \quad (3.52) \quad \mu$

Ο αναγνώστης που ενδιαφέρεται να μελετήσει τις αποδείξεις των παραπάνω θεωρημάτων παραπέμπεται στις εργασίες [339, 406]. Επιπλέον, οι συγγραφείς εξέτασαν την προτεινόμενη μέθοδο με άλλες γνωστές μεθόδους τροχιάς. Επιπλέον, παρουσιάζουν το συνολικό αριθμό επαναλήψεων που απαιτούνται για κάθε πρόβλημα. Συγκεκριμένα, χρησιμοποίησαν μια τροποποιημένη μέθοδο Runge-Kutta με τάξη 4=5 σε πεπλεγμένη μορφή. Τα υπολογιστικά πειράματα που διεξήγαγαν περιλαμβάνουν γνωστές συναρτήσεις όπως: (α) τη συνάρτηση του Rosenbrock, (β) τη συνάρτηση του Wood, (γ) τη συνάρτηση του Powell και (δ) εκθετικές μορφές συναρτήσεων. Τα πειραματικά αποτελέσματα έδειξαν ότι η παρεχόμενη μέθοδος ήταν ανταγωνιστική.

Στο άρθρο [360], παρουσιάζεται μια τροποποίηση του αλγορίθμου των Snyman-Fatti [359], προκειμένου να αντιμετωπιστούν τα μη κυρτά προβλήματα ολικής βελτιστοποίησης με περιορισμούς. Η έκδοση του αλγορίθμου χωρίς περιορισμούς εφαρμόστηκε επιτυχώς [206],

προκειμένου να επιλυθούν τα προβλήματα βελτιστοποίησης που συναντώνται στις δομές. Η βασική ιδέα της προτεινόμενης μεθόδου ήταν: (α) Αρχικά, οι συγγραφείς εφάρμοσαν την παραδοσιακή μέθοδο των Snyman-Fatti (τον αλγόριθμο χωρίς περιορισμούς), για να υπολογίσουν το ολικό ελάχιστο της συνάρτησης ποινής, παρόμοια με τη σχέση (3.48), (β) στο επόμενο βήμα έγινε η αναγνώριση όλων των περιορισμών σχετικά με τη λύση της συνάρτησης ποινής και (γ) εφάρμοσαν τον αλγόριθμο των Snyman-Fatti προκειμένου να ελαχιστοποιηθεί μια συνάρτηση ελαχίστων τετραγώνων. Οι συγγραφείς, για να ελέγξουν την απόδοση της προτεινόμενης μεθόδου, την οποία ονόμασαν Snyman Fatti CONstrained (SFCON) αλγόριθμο, χρησιμοποίησαν ένα σύνολο γνωστών προβλημάτων με περιορισμούς [134]. Τα αποτελέσματα που προέκυψαν έδειξαν ότι η μέθοδος SFCON είναι άκρως αποτελεσματική, δεδομένου ότι εντόπισε τις λύσεις όλων των δοκιμαστικών προβλημάτων (η μέθοδος δοκιμάστηκε σε 10 προβλήματα, εκ των οποίων το ένα είχε 5 ειδικές περιπτώσεις). Η ακρίβεια της μεθόδου ήταν ικανοποιητική, και παρά το γεγονός ότι ο αριθμός των συναρτησιακών υπολογισμών ήταν υψηλός, ο χρόνος της CPU δεν ήταν τόσο μεγάλος.

Μια άλλη παραλλαγή του αλγορίθμου των Snyman-Fatti παρουσιάστηκε στην εργασία [167]. Οι συγγραφείς του άρθρου επεσήμαναν ότι, στην τροποποιημένη έκδοση επιτυγχάνονται σημαντικές βελτιώσεις στην ακρίβεια και την αποτελεσματικότητα. Ειδικότερα, διακρίνουν ορισμένα βασικά στάδια: (α) την του αλγορίθμου, όπου ο αλγόριθμος προσπαθεί να προσεγγίσει μια γειτονιά των ελαχίστων. Έτσι, μπορεί κανείς να ισχυριστεί ότι ο αλγόριθμος συγκλίνει στο ελάχιστο αν και, αρχικά, δεν επιτυγχάνεται η προσέγγιση του ολικού ελαχίστου της αντικειμενικής συνάρτησης, (β) τότε, η του αλγορίθμου δίνει έμφαση στην επίτευξη μεγαλύτερης ακρίβειας. Έτσι, δεδομένου ότι η τροχιά, αρχικά, μετακινείται σε σχετικά χαμηλό ελάχιστο, είναι δυνατό να εντοπίσει το μικρότερο από τα ελάχιστα σημεία. Επιπρόσθετα, η επιλογή παραμέτρων παίζει βασικό ρόλο και απαιτείται δοκιμή των οριοθετημένων περιοχών αλλά και των σχετικών παραβάσεων από τις τροχιές. Επιπλέον, εκτός από τις παραπάνω φάσεις, είναι απαραίτητο να ελέγξουμε και να αλλάξουμε κάποιες συγκεκριμένες παραμέτρους, όπως το χρονικό βήμα t και την παράμετρο που εμφανίζεται στον αρχικό αλγόριθμο. Μια σημαντική λεπτομέρεια είναι, επίσης, το γεγονός ότι ξεπερνάται ένα όριο από μια τροχιά. Σε αυτή την περίπτωση η συνιστώσα του $X(t)$, η οποία παραβιάζει το σχετικό όριο που έχει τεθεί, αναδημιουργείται στην περιοχή ενδιαφέροντος με τυχαίο τρόπο. Οι συγγραφείς εξέτασαν την προτεινόμενη μέθοδο, η οποία ονομάζεται μέθοδος SF-M, με άλλες γνωστές μεθόδους ολικής βελτιστοποίησης [167]. Επιπλέον, η προτεινόμενη μέθοδος δοκιμάστηκε σε προβλήματα βελτιστοποίησης μη τετριμμένων ορθοτροπικών μεμβρανών (problems of nontrivial orthotropic membrane), πλακών (plates) και δομών κελύφους (shell structures). Τα πειραματικά αποτελέσματα έδειξαν ότι η μέθοδος είναι ισχυρή και ακριβής, ενώ η απόδοσή της είναι καλύτερη από τις συγκρινόμενες μεθόδους, ειδικά όταν εξετάζονται συναρτήσεις με μεγάλο αριθμό μεταβλητών ή μεγάλο αριθμό τοπικών ελαχίστων.

Μια ενδιαφέρουσα προσέγγιση παρουσιάζεται επίσης, στο άρθρο [348], όπου οι συγγραφείς του παρουσίασαν έναν αλγόριθμο που ονομάζεται μ μ μ - (nonlinear optimization via external lead (NOVEL) algorithm). Αυτός ο αλγόριθμος επιτυγχάνει ολική βελτιστοποίηση και εφαρμόζεται στη διαδικασία μάθησης ενός τεχνητού νευρωνικού δικτύου. Μέσα από τη μελέτη τους, κατέληξαν στο συμπέρασμα ότι για να οικοδομηθεί μια καλή διαδικασία αναζήτησης, πρέπει να πληρούνται δύο βασικά χαρακτηριστικά: (α) Να είναι σε θέση να χρησιμοποιεί τις πληροφορίες της κλίσης της συνάρτησης για να πραγματοποιήσει μια καλή τοπική αναζήτηση στην περιοχή αναζήτησης και (β) να μπορεί να «ξεφύγει» από τομείς τοπικού ελαχίστου προκειμένου να αναζητηθεί το ολικό ελάχιστο της αντικειμενικής συνάρτησης. Οι τοπικές μέθοδοι ελαχιστοποίησης έχουν δοκιμαστεί αποτελεσματικά σε ορισμένες περιπτώσεις [32, 96]. Τα μειονεκτήματά τους, ωστόσο, όπως η δυσκολία χειρισμού επίπεδων επιφανειών, έχουν οδηγήσει στην ανάγκη για μεθόδους ολικής ελαχιστοποίησης κατά τη διαδικασία μάθησης.

Αυτό, βεβαίως, δε σημαίνει ότι κάποιος μπορεί να «εμπιστευτεί» τυφλά μια μέθοδο ολικής ελαχιστοποίησης, είτε πρόκειται για μια ντετερμινιστική μέθοδο, είτε για μια πιθανοτική μέθοδο. Ο αναγνώστης μπορεί εύκολα να καταλάβει ότι οι μέθοδοι που ανήκουν σε αυτή την κατηγορία έχουν και τα δικά τους μειονεκτήματα. Για παράδειγμα, όταν ο χώρος αναζήτησης είναι πολύ μεγάλος, η σύγκλιση της μεθόδου μπορεί να είναι τρομερά αργή, ή, πιο συγκεκριμένα, οι πληροφορίες της κλίσης, παρότι είναι πολύ χρήσιμες, δεν μπορούν να χρησιμοποιηθούν στην προσομοιωμένη ανόπτηση (simulated annealing). Για το λόγο αυτό, ο αλγόριθμος NOVEL είναι ένα υβριδικό σύστημα που συνδυάζει ολικά και τοπικά χαρακτηριστικά. Λεπτομερέστερα, χρησιμοποιεί μία μέθοδο τροχιάς για να ξεφύγει από μια περιοχή τοπικών ελαχίστων, και τοπικές καθοδικές πορείες, προκειμένου να εντοπίσει αυτές τις περιοχές τοπικών ελαχίστων. Συνοπτικά, ο αλγόριθμος διακρίνεται από τα ακόλουθα χαρακτηριστικά: (α) αναζητά το χώρο της λύσης, (β) εντοπίζει περιοχές όπου είναι δυνατόν να υπάρχουν ελάχιστα σημεία, και (γ) προσδιορίζει τοπικά ελάχιστα. Συγκεκριμένα, όσον αφορά το πρώτο χαρακτηριστικό του αλγορίθμου, η εξερεύνηση του χώρου των λύσεων διεξάγεται με ένα συνεχές ανεξάρτητο ίχνος, το οποίο δεν παγιδεύεται στα τοπικά ελάχιστα. Στη συνέχεια, χρησιμοποιείται η τοπική κλίση, η οποία στηρίζεται στο ίχνος για να αποφύγει περιοχές τοπικού ελαχίστου. Τέλος, επιλέγεται ένα αρχικό σημείο από κάθε υποσχόμενη τοπική περιοχή. Στη συνέχεια, ένας αλγόριθμος καθόδου λαμβάνει αυτά τα σημεία ως αρχικές τιμές και το τοπικό ελάχιστο εντοπίζεται. Συμπερασματικά, ο αλγόριθμος NOVEL χρησιμοποιεί μια διαφορική εξίσωση για τη διεξαγωγή ολικής αναζήτησης και του εντοπισμού περιοχών που περιέχουν τοπικά ελάχιστα. Στη συνέχεια, εφαρμόζεται μια τοπική μέθοδος αναζήτησης για τον υπολογισμό αυτών των σημείων. Οι είσοδοι σε αυτόν τον αλγόριθμο, για παράδειγμα μια μέθοδος συζυγών κλίσεων (conjugate gradient), δίνονται μέσω της μεθόδου δυναμικής τροχιάς. Σε αυτό το σημείο, αξίζει να αναφερθούν πιο αναλυτικά ορισμένα ζητήματα της ολικής φάσης αναζήτησης. Κατά τη διάρκεια της ολικής φάσης αναζήτησης του αλγορίθμου, χρησιμοποιείται μια συνηθισμένη διαφορική εξίσωση της ακόλουθης μορφής:

$$X(t) = P(r \times f(X(t))) + Q(T(t); X(t)) \quad (3.53)$$

όπου t είναι μια αυτόνομη μεταβλητή, το T δηλώνει μια συνάρτηση ίχνους και τα $R; Q$ είναι γενικές μη γραμμικές συναρτήσεις. Η τελευταία εξίσωση αποτελείται από δύο όρους:

- (α) Το $P(r \times f(X))$ αντιπροσωπεύει την έλξη της τροχιάς από την κλίση της συνάρτησης προκειμένου να εντοπιστεί το τοπικό ελάχιστο, ενώ
- (β) $Q(T; X)$ δηλώνει τη συνάρτηση ίχνους η οποία οδηγεί την τροχιά εκτός της περιοχής του τοπικού ελαχίστου.

Αξίζει να δοθεί προσοχή στο ακόλουθο θέμα. Προκειμένου να διερευνηθεί ο χώρος αναζήτησης, οι συγγραφείς υιοθέτησαν την αναζήτηση από «χοντροκομμένες» περιοχές σε περιοχές που είναι πιο «εκλεπτυσμένες». Αυτό έγινε για να αποφευχθεί ο κατακερματισμός του χώρου σε πάρα πολλούς υποτομείς. Ειδικά, για πολλές διαστάσεις, αυτό θα ήταν ανέφικτο. Έτσι, οι συγγραφείς δημιούργησαν μια μη περιοδική αναλυτική συνάρτηση ίχνους [348]. Επιπλέον, χρησιμοποιήθηκαν δύο τρόποι για την επίλυση της διαφορικής εξίσωσης: (α) ένας λύτης που ονομάζεται Livermore solver για συνήθεις διαφορικές εξισώσεις (Livermore solver for ordinary differential equations (LSODE)) και (β) ένας λύτης εξισώσεων πεπερασμένης διαφοράς (finite-difference equation solver). Οι συγγραφείς του άρθρου διενήργησαν εκτεταμένα υπολογιστικά πειράματα για να ελέγξουν την απόδοση του αλγορίθμου τους. Δοκιμάζουν τον αλγόριθμο NOVEL για τα πολύ γνωστά προβλήματα: (α) το two-spiral πρόβλημα, (β) το πρόβλημα sonar, (γ) το πρόβλημα αναγνώρισης φωνηέντων (vowel recognition), (δ) το πρόβλημα 10 ισοτιμιών (10 parity) και (ε) το NetTalk πρόβλημα. Συνέκριναν τη μεθόδό τους με τους ακόλουθους γνωστούς αλγορίθμους ολικής βελτιστοποίησης: (α) προσομοίωση ανόπτησης (simulated annealing), (β) εξελικτικοί αλγόριθμοι (evolutionary

algorithms), (γ) συσχέτιση με πολλαπλές εκκινήσεις (sascade correlation with multistarts), (δ) απότομης κλίσης με πολλαπλές εκκινήσεις (gradient descent with multistarts) και (ε) μέθοδος περικοπής του Newton με πολλαπλές εκκινήσεις (truncated Newton's method with multistarts). Τα αποτελέσματα έχουν δείξει ότι η χρήση λύτη διαφορικών εξισώσεων είναι καλύτερη σε σχέση με την ταχύτητα από τον λύτη LSODE. Ωστόσο, ο λύτης LSODE παρείχε περισσότερες ποιοτικές λύσεις. Συνοπτικά, αυτή η νέα μέθοδος βελτίωσε την εκμάθηση των τεχνητών νευρωνικών δικτύων εμπρόσθιας τροφοδότησης (feed forward neural networks), ενώ λειτούργησε πολύ καλά για τη βελτιστοποίηση γενικών μη γραμμικών συναρτήσεων υψηλής διάστασης. Για περισσότερες λεπτομέρειες σχετικά με την απόδοση του αλγορίθμου NOVEL για κάθε πρόβλημα, ο αναγνώστης παραπέμπεται στην αναφορά [348].

Στο άρθρο [361], έχει προταθεί ένας αλγόριθμος με χαμηλές απαιτήσεις αποθήκευσης για τη διαχείριση προβλημάτων βελτιστοποίησης με περιορισμούς. Αυτός ο αλγόριθμος βασίζεται σε μια υπάρχουσα δυναμική μέθοδο τροχιάς και είναι κατάλληλος για συναρτήσεις πολλών μεταβλητών. Συγκεκριμένα, ο προτεινόμενος αλγόριθμος που ονομάζεται *dynamic-Q*, έχει το πλεονέκτημα να μην απαιτεί πληροφορίες του Εσσιανού μητρώου. Επιπλέον, παράγονται διαδοχικά υποπροβλήματα με την κατασκευή σφαιρικά τετραγωνικών προσεγγίσεων (spherically quadratic approximations). Οι συγγραφείς του άρθρου χρησιμοποίησαν το γνωστό αλγόριθμο *leap-frog* για να επιλύσουν αυτά τα υποπροβλήματα. Η παρεχόμενη μέθοδος σε σύγκριση με τη μέθοδο μ (Sequential Quadratic Programming), παρουσίασε ανταγωνιστικά αποτελέσματα.

Στο άρθρο [226], συναντάμε μια υβριδική μέθοδο ολικής βελτιστοποίησης βασισμένη στον πολύ γνωστό αλγόριθμο πολλαπλών εκκινήσεων [359]. Οι συγγραφείς του άρθρου χρησιμοποίησαν έναν αλγόριθμο από το πεδίο του μ (evolutionary computation), και ειδικότερα έναν αλγόριθμο μ (differential evolution (DE)) [375]. Έτσι, χρησιμοποιώντας μια μεταερευτική τεχνική, δημιουργούν έναν πληθυσμό σημείων για να διεξαγάγουν αναζήτηση στο χώρο των λύσεων. Κατά τη διάρκεια της διαφοροεξελικτικής φάσης, λαμβάνουν χώρα δύο ευρέως χρησιμοποιούμενοι τελεστές, αυτός της μ (mutation) και αυτός του μ (recombination). Η βασική ιδέα ήταν να προσδιοριστούν οι βοηθητικές τροχιές με πιο προηγμένο τρόπο από τον καθορισμένο στο αρχικό άρθρο των Snyman και Fatti. Συγκεκριμένα, η διαδικασία ξεκινά με ένα τυχαία επιλεγμένο σημείο x_0^1 , και χρησιμοποιώντας τη μέθοδο που προτείνεται από τους Snyman και Fatti [359], δημιουργείται η πρώτη τροχιά T_1 . Το αρχικό σημείο εισάγεται σε έναν πληθυσμό P , ο οποίος είναι κενός στην αρχή της διαδικασίας. Η τροχιά T_1 τερματίζεται στο σημείο x_n^1 . Αυτό το σημείο με τη σειρά του, συγχωνεύεται στον πληθυσμό P . Σε αυτό το σημείο του αλγορίθμου ξεκινά η βοηθητική τροχιά x_0^2 , όπως παραπάνω. Όπως είναι γνωστό, ο διαφοροεξελικτικός αλγόριθμος δεν μπορεί να ξεκινήσει ακόμη, δεδομένου ότι ο πληθυσμός P πρέπει να αποτελείται από τουλάχιστον τέσσερα άτομα. Για το λόγο αυτό, λαμβάνει χώρα ο αρχικός σχεδιασμός των βοηθητικών τροχιών, μέχρις ότου ο πληθυσμός P να αποκτήσει τον απαραίτητο αριθμό ατόμων. Επομένως, ο διαφοροεξελικτικός αλγόριθμος αρχίζει να εκτελείται, όταν συγχωνεύονται τα ακόλουθα άτομα στον πληθυσμό:

$$P = \{x_0^1; x_n^1; x_0^2; x_n^2\} \quad (3.54)$$

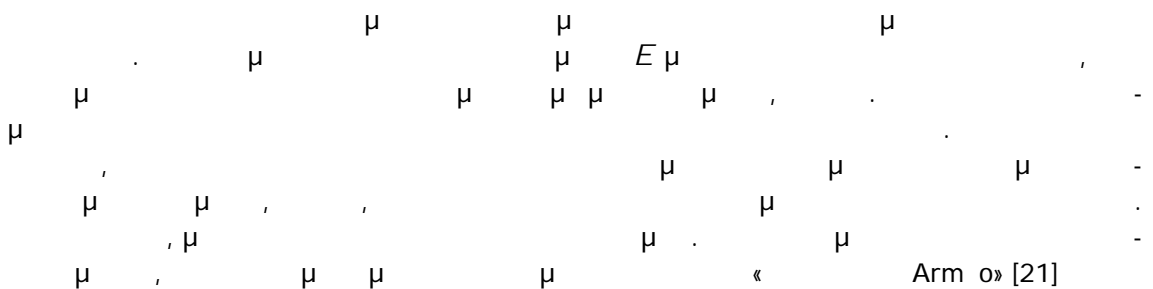
Σε αυτό το στάδιο, εφαρμόζονται οι τελεστές του αλγορίθμου διαφοροεξέλιξης, δηλαδή η μ και ο μ . Ως αποτέλεσμα, γεννιούνται τέσσερα νέα άτομα. Έτσι, ενεργοποιείται η διαδικασία μ και δημιουργείται ένας εντελώς νέος πληθυσμός. Το καλύτερο άτομο του νέου πληθυσμού, έστω P_2 , επιλέγεται για να δημιουργήσει την τροχιά T_3 (θεωρούμε ότι οι τροχιές T_1 και T_2 έχουν ήδη εμφανιστεί). Ακολούθως, το x_0^3 εισάγεται στον πληθυσμό P και τα υπόλοιπα σημεία του P_2 εγκαταλείπονται. Η διαδικασία συνεχίζεται με τον ίδιο τρόπο και παράγονται συνεχώς νέες τροχιές, οι οποίες με παρόμοιο τρόπο συμπεριλαμβάνονται στον πληθυσμό P . Όταν ο πληθυσμός φτάσει στον μέγιστο αριθμό ατόμων που μπορεί να συμπεριλάβει, ανακλύπει ένα ζήτημα. Αυτό αποτελεί πρόβλημα επειδή είναι

δυνατόν να γεννηθεί ένα άτομο το οποίο έχει χαμηλότερη συναρτησιακή τιμή από ένα σημείο που έχει ήδη συμπεριληφθεί στον πληθυσμό. Εάν συμβεί αυτό, (εάν ο πληθυσμός φτάσει στο μέγιστο του μεγέθους που μπορεί), τότε τα υπάρχοντα άτομα ταξινομούνται σύμφωνα με τις συναρτησιακές τιμές τους. Εάν υπάρχει ένα νέο σημείο με καλύτερα χαρακτηριστικά, αντικαθιστά το χειρότερο άτομο. Η διαδικασία που μόλις περιγράψαμε ήταν η κύρια συμβολή των συγγραφέων, προκειμένου να βελτιωθεί η απόδοση της τυπικής μεθόδου των Snyman-Fatti. Οι συγγραφείς του άρθρου [226], διεξήγαγαν υπολογιστικές πειραματικές δοκιμές χρησιμοποιώντας δέκα γνωστές συναρτήσεις διαφορετικών διαστάσεων. Συγκεκριμένα, χρησιμοποίησαν στις δοκιμές τους τις ακόλουθες συναρτήσεις: (α) τη συνάρτηση του Rosenbrock ($D = 2$), (β) τη συνάρτηση Freudenstein-Roth ($D = 2$), (γ) τη συνάρτηση Helical valley ($D = 3$), (δ) η συνάρτηση Levy No. 8 ($D = 3$), (ε) τη συνάρτηση του Wood ($D = 4$), (στ) τη συνάρτηση Watson ($D = 6$), (ζ) τη συνάρτηση Hyper-Ellipsoid ($D = 6$) και (η) τη συνάρτηση Rastrigin ($D = 6; 10; 15$). Αξίζει να σημειωθεί ότι η προτεινόμενη μέθοδος ξεπερνά την τυπική μέθοδο των Snyman-Fatti και σε ορισμένες περιπτώσεις το ποσοστό βελτίωσης έφτασε το 90% ή και περισσότερο. Επιπλέον, υπήρξαν περιπτώσεις όπου η υβριδική μέθοδος επιτυγχάνει σύγκλιση, ενώ ο αρχικός αλγόριθμος των Snyman και Fatti δεν συνέκλινε.

Οι μέθοδοι βελτιστοποίησης που χρησιμοποιούν δυναμικές τροχιές αναζήτησης καθίστανται πιο ενδιαφέρουσες, λόγω του μεγάλου αριθμού εφαρμογών τους. Οι μέθοδοι των τροχιών έχουν εφαρμοστεί επιτυχώς σε πολλές εφαρμογές που σχετίζονται με την μ (computational intelligence). Στη συνέχεια, περιγράφουμε εν συντομία αυτές τις μεθόδους για τη διαδικασία εκπαίδευσης ενός μ (artificial neural network (ANN)). Όπως αναφέρεται στο άρθρο [302], κατά τη διάρκεια εκπαίδευσης ενός τεχνητού νευρωνικού δικτύου, απαιτείται η ελαχιστοποίηση μιας συνάρτησης, γνωστή ως μ (error function). Αυτό καθιστά τη διαδικασία εκπαίδευσης ισοδύναμη με ένα πρόβλημα βελτιστοποίησης. Οι συγγραφείς στην εργασία τους [303], πρότειναν έναν νέο τρόπο ελαχιστοποίησης της συνάρτησης σφάλματος, με βάση τη γνωστή μέθοδο των Snyman-Fatti [359]. Υποθέτοντας ότι η συνάρτηση σφάλματος του δικτύου δηλώνεται με το E , θεωρούν τη διαφορική εξίσωση:

$$\dot{x} = -\nabla E(x) \tag{3.55}$$

όπου το x αντιπροσωπεύει το διάνυσμα των συναπτικών βαρών του τεχνητού νευρωνικού δικτύου. Οι συγγραφείς του άρθρου, μελέτησαν μια συγκεκριμένη κατηγορία νευρωνικών δικτύων, τα μ (multilayer feedforward neural networks (MFNNs)). Έκαναν μια παραλλαγή στην αρχική μέθοδο δυναμικής αναζήτησης των Snyman-Fatti, και εφαρμόζοντάς την στην εκπαίδευση ενός δικτύου MFNN, έλαβαν πολύ καλά αποτελέσματα. Αξίζει να σημειωθεί ότι μέσω της διεξαχθείσας μελέτης οι συγγραφείς κατέληξαν στο συμπέρασμα ότι οποιαδήποτε αριθμητική μέθοδος επίλυσης μιας συνήθους διαφορικής εξίσωσης, μπορεί να εφαρμοστεί στον αλγόριθμο εκπαίδευσης ενός τεχνητού νευρωνικού δικτύου. Η προτεινόμενη μέθοδος διαφέρει από την αρχική μέθοδο των Snyman-Fatti και εφαρμόζει τέσσερα πρόσθετα βήματα. Παρακάτω παρουσιάζουμε συνοπτικά πώς λειτουργεί αυτή η νέα μέθοδος:



$$\mu \quad \mu \quad \mu$$

$$\frac{E(x_{t+1})}{E(x_t)} > \quad (3.56)$$

Για να ελεγχθεί η προαναφερθείσα τροποποιημένη μέθοδος των Snyman-Fatti [303], οι συγγραφείς της εργασίας εκτέλεσαν υπολογιστικά πειράματα χρησιμοποιώντας τρία πολύ γνωστά προβλήματα από το «σύνολο δεδομένων proben1», και συγκεκριμένα τα: (α) Cancer1, (β) Diabetes1 και (γ) Heart1. Οι τοπολογίες (αρχιτεκτονικές) του τεχνητού δικτύου που χρησιμοποιήθηκαν ήταν οι ακόλουθες: (α) 9 4 22, (β) 8 2 2 2 και (γ) 35 8 24. Η προτεινόμενη μέθοδος συγκρίθηκε με γνωστές παραλλαγές του αλγορίθμου εκπαίδευσης οπίσθιας διάδοσης σφάλματος (back propagation (BP) algorithm) και συγκεκριμένα (α) τον αλγόριθμο back propagation with momentum (MBP), (β) τον αλγόριθμο second order momentum (SMBP) οπίσθιας διάδοσης σφάλματος, (γ) τον προσαρμοστικό αλγόριθμο οπίσθιας διάδοσης (adaptive back propagation (ABP) algorithm) (χρησιμοποιώντας το προσαρμοστικό σχήμα του Vogl [402]), (δ) τον αλγόριθμο (parallel tangents method (PARTAN)), (ε) τη μέθοδο βαθμωτών συζυγών κλίσεων (scaled conjugated gradient (SCG)), (στ) τον αλγόριθμο resilient back propagation (RPROP) και (ζ) το βελτιωμένο αλγόριθμο οπίσθιας διάδοσης improved resilient back propagation (iRPROP). Τα πειραματικά αποτελέσματα έχουν δείξει ότι, όσον αφορά το πρόβλημα Cancer1, η προτεινόμενη μέθοδος κατέλαβε την τρίτη θέση ως προς τον αριθμό των συναρτησιακών υπολογισμών που απαιτούνται. Επιπλέον, έχει επιτύχει τη δεύτερη καλύτερη απόδοση για το σφάλμα ταξινόμησης. Όσον αφορά το πρόβλημα Diabetes1, κατέλαβε την τρίτη θέση όσον αφορά τον αριθμό των συναρτησιακών υπολογισμών και το σφάλμα ταξινόμησης. Τέλος, για το πρόβλημα Heart1, έχει επιτύχει τη δεύτερη και τρίτη καλύτερη απόδοση αντίστοιχα. Εν κατακλείδι, αν και η προτεινόμενη τροποποιημένη μέθοδος των Snyman-Fatti δεν ήταν η καλύτερη για το σύνολο των δοκιμαστικών προβλημάτων, ήταν ανταγωνιστική και ήταν στις τρεις πρώτες μεθόδους με τις καλύτερες επιδόσεις σε όλα τα προβλήματα. Στο άρθρο [302], χρησιμοποιήθηκαν αριθμητικές μέθοδοι για την επίλυση προβλημάτων αρχικών τιμών [49, 50] προκειμένου να επιλυθεί το εγγενές πρόβλημα βελτιστοποίησης που λαμβάνει χώρα κατά την εκπαίδευση των τεχνητών νευρωνικών δικτύων. Η ομάδα των μεθόδων που πρότειναν, είναι μια «υποομάδα» των μεθόδων δυναμικής τροχιάς αναζήτησης. Συγκεκριμένα, προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα, επιλύθηκε η ακόλουθη συνήθης διαφορική εξίσωση:

$$\dot{x} = -r E(x) \quad (3.57)$$

όπου το x αντιπροσωπεύει το βάρος του δικτύου και το E καθορίζει τη μ - ενός πολλαπλών στρωμάτων νευρωνικού δικτύου εμπρόσθιας τροφοδότησης (multilayer feedforward neural network (MFNN)). Οι συγγραφείς του άρθρου για την επίλυση του αντίστοιχου προβλήματος αρχικών τιμών:

$$\dot{x} = -r E(x); \quad x(0) = x_0 \quad (3.58)$$

εφάρμοσαν τις αριθμητικές μεθόδους Runge-Kutta. Συγκεκριμένα, εφάρμοσαν τις μεθόδους Runge-Kutta δεύτερης τάξης και δύο σταδίων:

$$x_{n+1} = x_n + a_1 k_1 + a_2 k_2; \quad n = 0; 1; 2; \dots \quad (3.59)$$

όπου:

$$\begin{aligned} k_1 &= h r E(x_n); \\ k_2 &= h r E(x_n + b_2 k_1) = h r E(x_n + h b_2 r E(x_n)) \end{aligned} \quad (3.60)$$

και $h > 0$ είναι ένα δεδομένο μέγεθος βήματος. Εάν οι τιμές των παραμέτρων $a_1; a_2$ και b_2 , πληρούν ένα σύστημα αλγεβρικών εξισώσεων, τότε οι αντίστοιχες μέθοδοι που λαμβάνονται θεωρούνται ως μέθοδοι Runge-Kutta δεύτερης τάξης. Επίσης, είναι γνωστό ότι μπορεί να κατασκευαστεί ένας άπειρος αριθμός μεθόδων Runge-Kutta δεύτερης τάξης [49, 50]. Οι συγγραφείς του άρθρου, ορίζουν τις ακόλουθες τιμές για τις παραμέτρους $a_1; a_2$ και b_2 , προκειμένου να λάβουν τους αντίστοιχους RK1, RK2 και RK3 λύτες:

(α) RK1 μέθοδος: $a_1 = 0; a_2 = 1$ και $b_2 = 1=2$,

(β) RK2 μέθοδος: $a_1 = a_2 = 1=2$ και $b_2 = 1$ και

(γ) RK1 μέθοδος: $a_1 = 1=4$ και $a_2 = b_2 = 2=3$

Οι συγγραφείς εξέτασαν τις προτεινόμενες μεθόδους σε τρία ευρέως γνωστά και χρησιμοποιούμενα προβλήματα από τη βάση δεδομένων proben1. Συγκεκριμένα, στα υπολογιστικά πειράματα που διεξήγαγαν, συμπεριέλαβαν τα προβλήματα: (α) XOR, (β) το πρόβλημα Coder-Decoder και (γ) το πρόβλημα ταξινόμησης Cancer1. Οι αντίστοιχες αρχιτεκτονικές του τεχνητού νευρωνικού δικτύου ήταν: (α) $2 \times 2 \times 1$, (β) $4 \times 2 \times 4$ και (γ) $9 \times 4 \times 4 \times 2$. Τα προτεινόμενα σχήματα συγκρίθηκαν με τις γνωστές μεθόδους της

μ (Backpropagation (BP)). Τα πειραματικά αποτελέσματα έδειξαν ότι οι προτεινόμενες μέθοδοι Runge-Kutta ξεπερνούν την οικογένεια των μεθόδων BP. Συγκεκριμένα, όσον αφορά το πρόβλημα XOR, η μέθοδος RK1 απαιτούσε μικρότερο αριθμό συναρτησιακών υπολογισμών από όλες τις άλλες, ενώ η χαμηλότερη τυπική απόκλιση επιτεύχθηκε με τη μέθοδο RK3. Όσον αφορά το δεύτερο πρόβλημα, όλες οι προτεινόμενες μέθοδοι Runge-Kutta απαιτούσαν λιγότερους συναρτησιακούς υπολογισμούς από όλες τις άλλες μεθόδους και η μόνη ανταγωνιστική μέθοδος ήταν η μ μ μ (adaptive back propagation). Τέλος, όσον αφορά το πρόβλημα Cancer1, η λιγότερο απαιτητική υπολογιστικά μέθοδος ήταν η μέθοδος RK2, ενώ το ελάχιστο σφάλμα ταξινόμησης επιτεύχθηκε από τη μέθοδο RK3.

Οι συγγραφείς συνέχισαν την προσπάθειά τους να αντιμετωπίσουν το πρόβλημα ελαχιστοποίησης που συναντάται στην εκπαίδευση ενός τεχνητού νευρωνικού δικτύου με το άρθρο τους [302]. Σε αυτή την εργασία, προτείνουν μια νέα μέθοδο για δυναμικές τροχιές αναζήτησης χρησιμοποιώντας τον Stoermer [399], σχετικά με το πρόβλημα αρχικών τιμών:

$$\dot{y} = f(x; y(x)); \quad y(x_0) = y_0 \quad y'(x_0) = z_0 \quad (3.61)$$

Συγκεκριμένα, επέλυσαν το ακόλουθο πρόβλημα αρχικών τιμών:

$$\ddot{x} = r f(x) \quad x(0) = x_0 \quad \dot{x}(0) = 0 \quad (3.62)$$

και συνέκριναν το προτεινόμενο σχήμα τους με την οικογένεια μεθόδων BP, όπως έπραξαν και στα προηγούμενα άρθρα τους. Στη συνέχεια, εξετάζουμε συνοπτικά τη λειτουργικότητα του προαναφερθέντος κανόνα και τον τρόπο εφαρμογής του από τους συγγραφείς.

Ο κανόνας του Stoermer περιγράφεται από τους ακόλουθους τύπους:

$$\begin{aligned} y_1 &= y_0 + hfz_0 + \frac{1}{2}hf(x_0; y_0)g \\ y_{n+1} &= 2y_n - y_{n-1} + h^2f(x_0 + nh; y_n); \quad n = 1; 2; \dots; m - 1 \\ z_m &= y_m - \frac{y_{m-1}}{h} + \frac{1}{2}hf(x_0 + H; y_m) \end{aligned} \quad (3.63)$$

όπου το H αντιπροσωπεύει το συνολικό βήμα που πρέπει να χρησιμοποιηθεί σε m υποβήματα. Έτσι, κάθε βήμα είναι μήκους $h = H/m$. Η τελική τιμή Z_m είναι $y(x_0 + H)$. Έχει

δειχθεί πώς μπορούμε να γράψουμε τις σχέσεις 3.63 προκειμένου να μειώσουμε το σφάλμα στρογγυλοποίησης χρησιμοποιώντας τις ποσότητες:

$$y_n = y_{n+1} - y_n \quad (3.64)$$

Έτσι, χρησιμοποιώντας αυτή την προσέγγιση η μέθοδος αποκτά την ακόλουθη μορφή:

$$\begin{aligned} y_0 &= y_1 - y_0 = hfz_0 + \frac{1}{2}hf(x_0; y_0)g; \\ y_1 &= y_0 + y_1; \\ y_n &= y_{n-1} + h^2 f(x_0 + nh; y_n); \quad n = 1; 2; \dots; m-1; \\ y_{n+1} &= y_n + y_n; \quad n = 1; 2; \dots; m-1; \\ z_m &= \frac{m-1}{h} + \frac{1}{2}hf(x_0 + H; y_m) \end{aligned} \quad (3.65)$$

Όπως και στα προηγούμενα έργα που αναλύθηκαν παραπάνω, η προτεινόμενη μέθοδος τέθηκε «αντιμέτωπη» με την οικογένεια μεθόδων BP. Τα προβλήματα που χρησιμοποιήθηκαν ήταν τα εξής: (α) Cancer1, (β) Diabetes1 και (γ) Heart1 με σταθερά σύνολα εκπαίδευσης και δοκιμής. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος πέτυχε λιγότερους συναρτησιακούς υπολογισμούς σε σύγκριση με όλες τις υπόλοιπες, όσον αφορά τα προβλήματα Cancer1 και Diabetes1. Όσον αφορά το σφάλμα ταξινόμησης, η μέθοδος Stoermer κατέλαβε τη δεύτερη θέση, πίσω από την προσαρμοστική BP. Τέλος, όσον αφορά το πρόβλημα Heart1, η προτεινόμενη μέθοδος πέτυχε την πρώτη και τη δεύτερη απόδοση αντίστοιχα.

3.2 Πρόσφατες μέθοδοι δυναμικών τροχιών και εφαρμογές αυτών

Σε αυτήν την ενότητα παρουσιάζονται οι πιο πρόσφατες μέθοδοι βελτιστοποίησης που βασίζονται στις δυναμικές τροχιές αναζήτησης. Συγκεκριμένα, παρέχουμε τις μεθόδους οι οποίες έχουν αναπτυχθεί τα τελευταία δέκα χρόνια.

Σημαντικές είναι οι προσπάθειες που έχουν γίνει χρησιμοποιώντας μεθόδους τροχιάς προκειμένου να αντιμετωπιστούν δύσκολα προβλήματα, όπως η βελτιστοποίηση πολλαπλών αντικειμένων (multi-objective optimization), καθώς και μεγάλης κλίμακας προβλήματα ολικής βελτιστοποίησης [389, 390]. Οι συγγραφείς αυτών των δύο άρθρων, πρότειναν ένα νέο αλγόριθμο, που ονομάζεται μ (multiple trajectory search (MTS) method). Η βασική ιδέα αυτού του αλγορίθμου ήταν ότι, πολλές τροχιές αναζητούν το χώρο των λύσεων, και μέθοδοι βασισμένες στα δέντρα απόφασης πραγματοποιούν τοπικές αναζητήσεις. Ως αποτέλεσμα, ανάλογα με τη μορφή του χώρου των λύσεων, οι τοπικές μέθοδοι θα μπορούσαν να ενεργοποιηθούν, όποτε ήταν απαραίτητο να εντοπιστεί ένα τοπικό ή ολικό ελάχιστο. Η προτεινόμενη μέθοδος συγκρίθηκε κυρίως με τις εξελικτικές μεθόδους [209, 231], και τα υπολογιστικά πειράματα που διεξήγαγαν έχουν δείξει επωφελή αποτελέσματα.

Σε αυτό το σημείο, θα θέλαμε να υπογραμμίσουμε τη συμβολή και τη σημασία των προηγούμενων μεθόδων, οι οποίες καθορίζουν σε μεγάλο βαθμό τις τρέχουσες προσεγγίσεις. Ένα στοιχείο που αποκαλύπτει τη συμβολή τους είναι, τόσο οι βελτιώσεις που έχουν γίνει σε σχέση με αυτούς τους προηγούμενους αλγορίθμους, όσο και οι επαναξιολογήσεις που έγιναν σχετικά με τα πρωτότυπα έργα, προκειμένου να δοκιμαστεί η αποτελεσματικότητα των μεθόδων σε πρόσθετα προβλήματα. Στο άρθρο [366], οι Snyman και Kok επανεκτιμούν τον πολύ γνωστό αλγόριθμο μ (multiple trajectory search (MTS) method).

μ (multiple trajectory search (MTS) method) και τον συγκρίνουν με τον μ (differential evolution

algorithm) των Storn και Price [375] σε ένα εκτεταμένο σύνολο προβλημάτων. Δεδομένου ότι αυτοί οι δύο αλγόριθμοι είναι ευρέως γνωστοί, και ο πρώτος παρουσιάζεται εκτενώς στην Ενότητα 3.1, θα επικεντρωθούμε μόνο σε μερικά πολύ ενδιαφέροντα αποτελέσματα που προκύπτουν από τα πειράματα που διεξήχθησαν από τους συγγραφείς. Αρχικά, η σύγκρισή τους περιλάμβανε 48 γνωστά προβλήματα με συναρτήσεις διαφορετικών διαστάσεων και ευρείας κλίμακας. Συγκεκριμένα, στα πειράματα συμπεριλήφθηκαν συναρτήσεις 2 διαστάσεων έως και συναρτήσεις 20 διαστάσεων με συνεχείς μεταβλητές. Τα πειραματικά αποτελέσματα έδειξαν ότι η μέθοδος των Snyman-Fatti έφθασε σε πολύ καλά αποτελέσματα, καθώς πέτυχε το ζητούμενο στόχο σε 43 από τις 48 περιπτώσεις (απέτυχε να λύσει 5 προβλήματα - τέσσερις συναρτήσεις 9 διαστάσεων και μία συνάρτηση 10 διαστάσεων). Επιπλέον, για περισσότερα από τα μισά από τα προβλήματα, η μέθοδος πέτυχε ποσοστό επιτυχίας 100%. Στις περιπτώσεις όπου το ποσοστό επιτυχίας δεν ήταν ιδιαίτερα υψηλό, οι συγγραφείς προτείνουν ως διόρθωση να αυξηθούν τον ελάχιστο αριθμό επαναλήψεων. Λαμβάνοντας υπόψη την απόδοση των μεθόδων και συγκεκριμένα, το υπολογιστικό κόστος καθεμιάς (ελάχιστος αριθμός συναρτησιακών υπολογισμών), γίνεται εύκολα κατανοητό ότι οι δύο μέθοδοι δεν είναι συγκρίσιμες. Αυτό συμβαίνει επειδή η μέθοδος Snyman-Fatti απαιτεί την εκτίμηση της κλίσης της υπό εξέταση συνάρτησης, ενώ οι αλγόριθμοι διαφοροεξέλιξης από τη φύση τους δεν απαιτούν εκτίμηση της κλίσης. Οι συγγραφείς, ωστόσο, επανέλαβαν μερικά από τα πειράματα, χρησιμοποιώντας την αναλυτική έκφραση της κλίσης. Όπως ήταν αναμενόμενο, ο ρυθμός μείωσης του χρόνου της CPU ήταν αξιοσημείωτος. Επίσης, αξίζει να σημειωθεί ότι υπήρχαν 8 προβλήματα στα οποία η μέθοδος Snyman-Fatti απαιτούσε λιγότερους συναρτησιακούς υπολογισμούς. Οι συγγραφείς, επίσης, εξέτασαν αυτή τη μέθοδο στην περίπτωση που είναι γνωστό το ολικό ελάχιστο. Τα πειραματικά αποτελέσματα έδειξαν ότι η μέθοδος Snyman-Fatti επιτυγχάνει να βρει το ολικό ελάχιστο σε 44 προβλήματα. Τέλος, η φύση της αρχικής μεθόδου των Snyman-Fatti, υποδηλώνει ότι το πρόβλημα της ελαχιστοποίησης της δυναμικής ενέργειας θα προσεγγιστεί επιτυχώς (με αυτή τη μέθοδο). Στην πραγματικότητα, οι συγγραφείς εξέτασαν τη μέθοδο Snyman-Fatti στο γνωστό πρόβλημα [405] και τα αποτελέσματα έδειξαν πάλι ότι η μέθοδος Snyman-Fatti είναι ιδιαίτερα ανταγωνιστική σε σύγκριση με τους αλγορίθμους διαφοροεξέλιξης. Ένα άλλο πλεονέκτημα είναι ότι η μέθοδος Snyman-Fatti μπορεί εύκολα να τροποποιηθεί, προκειμένου να επιλυθούν προβλήματα βελτιστοποίησης με περιορισμούς.

Μια ενδιαφέρουσα εφαρμογή του αλγορίθμου Snyman-Fatti (σε μια περαιτέρω βελτιωμένη εκδοχή του) παρουσιάστηκε στο άρθρο [124]. Σε αυτή την εργασία οι συγγραφείς της εξετάζουν ένα πολύ γνωστό πρόβλημα που συναντάται στις χαλύβδινες κατασκευές και σχετίζεται με τις ενισχυμένες πλάκες σε μια δομή κατασκευής. Ας περιγράψουμε εν συντομία αυτό το πρόβλημα. Σε πολλά προβλήματα δομής [123], όπως στην κατασκευή δαπέδων, πλατφορμών ή στην κατασκευή μιας γέφυρας, χρησιμοποιούνται ενισχυμένες πλάκες. Ένα σημαντικό ζήτημα που είναι εγγενές σε αυτά τα προβλήματα, είναι ο βέλτιστος σχεδιασμός συγκολλημένων τετραγωνικών ενισχυμένων πλακών και φυσικά η ελαχιστοποίηση μιας σειράς παραμέτρων. Συγκεκριμένα, σε αυτό το πρόβλημα έχουμε τους ακόλουθους περιορισμούς:

- (α) Η τάση στην πλάκα βάσης και, (επίσης,) στις ενισχύσεις.
- (β) Η τάση στην εκτροπή των ενισχυτικών ακμών και, (επίσης,) των εσωτερικών ενισχυτικών στοιχείων.

Ως αποτέλεσμα, πρέπει να ελαχιστοποιήσουμε τη συνάρτηση κόστους που περιλαμβάνει τις ακόλουθες παραμέτρους: (α) το υλικό, (β) τη συγκόλληση και (γ) το κόστος βαφής. Η συνάρτηση κόστους προκύπτει σύμφωνα με την ακολουθία της κατασκευής. Επιπλέον, πρέπει να λάβουμε υπόψη τις παραμέτρους σχεδιασμού, όπως: (α) το πάχος της πλάκας βάσης, (β) τις διαστάσεις των άκρων και (γ) τις διαστάσεις και τον αριθμό των εσωτερικών

ενισχυτικών στοιχείων. Αυτοί οι περιορισμοί οδηγούν σε μια συνάρτηση όπου απαιτείται η κατάλληλη βελτιστοποίηση. Για περισσότερες λεπτομέρειες παραπέμπουμε τον αναγνώστη στο άρθρο [124]. Προφανώς, αυτό είναι ένα πρόβλημα βελτιστοποίησης με αρκετούς περιορισμούς. Επομένως, λαμβάνοντας ένα κατάλληλο μετασχηματισμό, το αρχικό πρόβλημα τροποποιείται κατάλληλα σε ένα πρόβλημα ελαχιστοποίησης χωρίς περιορισμούς, προκειμένου να εφαρμοστεί η μέθοδος Snyman-Fatti για την εύρεση των λύσεων. Οι συγγραφείς του άρθρου, συνέκριναν τη μέθοδο Snyman-Fatti με τον γνωστό αλγόριθμο βελτιστοποίησης με χρήση σμήνους σωματιδίων (particle swarm optimization (PSO) algorithm). Τα αποτελέσματα που προέκυψαν, έδειξαν ότι η μέθοδος Snyman-Fatti είναι ανταγωνιστική, σθεναρή και αποτελεσματικά ακριβής για το βέλτιστο σχεδιασμό που απαιτούσε το έργο.

Στη συνέχεια, παρουσιάζουμε μια άλλη εφαρμογή που σχετίζεται με τις δυναμικές τροχιές αναζήτησης και τα τεχνητά νευρωνικά δίκτυα. Συγκεκριμένα, στο άρθρο [39], οι συγγραφείς του μελέτησαν ένα συγκεκριμένο είδος συναρτήσεων που ονομάζονται μη κυρτές κλιμακωτές συναρτήσεις (non-convex scalar functions). Ιδιαίτερα, η προσέγγισή τους βασίζεται στη συνήθη διαφορική εξίσωση:

$$\dot{x} + g(x) + r f(x) = 0 \quad (3.66)$$

όπου $g > 0$ είναι ένα θετικό βαθμωτό στοιχείο, και $f \in C^1$ είναι μια αντικειμενική συνάρτηση με ένα μοναδικό ολικό ελάχιστο που σχετίζεται με έναν αλγόριθμο νευρωνικού δικτύου, γνωστός ως μ (continuous-time algorithm) [37]. Η βασική ιδέα ήταν η κατασκευή μιας νέας μεθόδου συνεχούς χρόνου με βάση τη γνωστή μέθοδο

(conjugate gradient). Χρησιμοποιώντας αυτό το νέο σχήμα, οι τροχιές που παράγονται από τη λύση της συνήθους διαφορικής εξίσωσης, μπορούν να «ξεφύγουν» από το τοπικό ελάχιστο. Ο προτεινόμενος αλγόριθμος, που ονομάζεται μ (controlled conjugate gradient (CCG) network), περιγράφεται από το ακόλουθο σύστημα συνήθων διαφορικών εξισώσεων:

$$\begin{aligned} \dot{x} &= a(x; u)u \\ \dot{u} &= -r f(x) - b(x; u)u \end{aligned} \quad (3.67)$$

όπου a, b αντιπροσωπεύουν συγκεκριμένες συναρτήσεις. Αυτή η μέθοδος βασίζεται στη Lyapunov (control Lyapunov function (CLF)) η οποία περιγράφεται αναλυτικά στο άρθρο [38]. Η προτεινόμενη μέθοδος συγκρίθηκε με το μ (heavy ball with friction (HBF) network) [311]. Τα πειράματα που διεξάγονται από τους συγγραφείς του άρθρου περιλαμβάνουν τις ακόλουθες δοκιμαστικές συναρτήσεις: (α) τις συναρτήσεις κορυφών MATLAB's peaks functions, (β) τη συνάρτηση του Rastrigin, (γ) τη συνάρτηση του Ackley και (δ) τη συνάρτηση του Griewank. Η πλειονότητα των αλγορίθμων CCG συγκλίνουν ταχύτερα στο ελάχιστο από τον αλγόριθμο HBF. Επιπλέον, οι συγγραφείς επεσήμαναν ότι εάν η προσαρμογή των παραμέτρων γίνει πολύ προσεκτικά, τότε η οικογένεια μεθόδων CCG θα ξεφεύγει από ένα τοπικό ελάχιστο. Παρά το γεγονός ότι τα αποτελέσματα ήταν πολλά υποσχόμενα, οι συγγραφείς επεσήμαναν ότι χρειάζονται περισσότερα υπολογιστικά πειράματα. Για το λόγο αυτό, πρότειναν να δοκιμαστούν τα δίκτυα CCG χρησιμοποιώντας άλλες γνωστές μεθόδους τροχιάς, όπως αυτές των άρθρων [226] και [359].

Οι εφαρμογές μεθόδων που χρησιμοποιούν δυναμικές τροχιές δεν περιορίζονται στο πεδίο της Φυσικής, των Μαθηματικών ή της Μηχανικής, όπως έχουμε ήδη αναφέρει στην εισαγωγή του παρόντος κεφαλαίου. Στο άρθρο [28], οι συγγραφείς της εργασίας, παρείχαν έναν αλγόριθμο ανοιχτού κώδικα για την «άπληστη» αναζήτηση (greedy search) προκειμένου να αντιμετωπιστεί η μεγιστοποίηση του κέρδους στις πληροφορίες σχετικά με τον επιθυμητό στόχο. Πρώτον, έλαβαν την κατάλληλη εξίσωση κίνησης, έτσι, ώστε οι τροχιές να προχω-

Πίνακας 3.1: Σύνοψη περιγραφή, εφαρμοσιμότητα και αριθμός αναφορών σχετικά με τις μεθόδους δυναμικής τροχιάς που παρουσιάστηκαν στην Ενότητα 3.1 και Ενότητα 3.2. Όπου «ΣΠΔ» υποδηλώνει το σύνολο των αναφορών, ενώ «ΑΔΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.

Τίτλος	Πρώτης τάξης ΣΔΕ	Δευτερης τάξης ΣΔΕ	ΣΠΔ	ΑΔΕ	Σύντομη περιγραφή
“Widely Convergent Methods for Finding Multiple Solutions of Simultaneous Nonlinear Equations”	(ναι)	(όχι)	414	8.80	Στη συγκεκριμένη εργασία, ο συγγραφέας αντιμετωπίζει το πρόβλημα της εύρεσης λύσης ενός συστήματος μη γραμμικών εξισώσεων. Ειδικότερα, η προτεινόμενη μέθοδος μπορεί να προσδιορίσει πολλαπλές ρίζες και, σε ορισμένες περιπτώσεις, να επιτύχει ολική σύγκλιση [44].
“Generalized descent for global optimization”	(όχι)	(ναι)	379	9.97	Ο Griewank έχει παρουσιάσει κάποιες βασικές ιδιότητες της τροχιάς που πρέπει να τηχθούν. Επιπλέον, διατυπώνει και αποδεικνύει ένα θεώρημα, το οποίο ενισχύει την άποψη ότι η λύση της διαφορικής εξίσωσης που εξετάζεται μετατρέπεται στην τροχιά της πιο απότομης καθόδου [163]. Προτείνεται μια πρακτική μέθοδος εντοπισμού των τοπικών ελαχίστων, κυρίως για μια συνάρτηση που η πρώτη παράγωγος μπορεί εύκολα να εκτιμηθεί. Αυτή η μέθοδος στοχεύει στον εντοπισμό των ελαχίστων για συναρτήσεις πολλαπλών μεταβλητών. Η μέθοδος του Shyman αντιστοικίζεται πολύ καλά σε περιπτώσεις όπου το σημείο εκκίνησης δεν είναι κοντά στα τοπικά ελάχιστα [363].
“A new and dynamic method for unconstrained minimization”	(όχι)	(ναι)	172	4.64	Προτείνεται μια συγκεκριμένη εργασία που παρουσιάζεται με τη μορφή ενοχλήσεων για να ελεγχθεί η απόδοση της μεθόδου. Η μέθοδος του Shyman αντιστοικίζεται πολύ καλά σε περιπτώσεις όπου το σημείο εκκίνησης δεν είναι κοντά στα τοπικά ελάχιστα [363].
“A multi-start global minimization algorithm with dynamic search trajectories”	(όχι)	(ναι)	149	4.65	Οι συγγραφείς του άρθρου παρήγαγαν μια μέθοδο ολικής ελαχιστοποίησης χρησιμοποιώντας έναν αλγόριθμο πολλαπλών εκκινήσεων. Η επιτυχία αυτής της μεθόδου ήταν ότι, με τις κατάλληλες τροποποιήσεις στις τροχιές αναζήτησης, κατάφεραν να διεκρινούν τις περιοχές σύγκλισης όσον αφορά το ολικό ελάχιστο της αντικειμενικής συνάρτησης [359].
“Evolutionary operators in global optimization with dynamic search trajectories”	(όχι)	(ναι)	13	0.80	Στη συγκεκριμένη εργασία παρουσιάζεται μια υβριδική μέθοδος ολικής βελτιστοποίησης που βασίζεται στον πολύ γνωστό αλγόριθμο πολλαπλών εκκινήσεων [359]. Οι συγγραφείς του άρθρου χρησιμοποίησαν έναν αλγόριθμο από το πεδίο του εξελικτικού υπολογισμού, ειδικότερα έναν αλγόριθμο [226].
“Multiple trajectory search for large scale global optimization”	(όχι)	(ναι)	186	16.90	Οι συγγραφείς του άρθρου πρότειναν έναν νέο αλγόριθμο, ο οποίος ονομάζεται μ. Η βασική ιδέα αυτού του αλγορίθμου ήταν ότι πολλές τροχιές διεξάγονται ανεξάρτητα στο χώρο των λύσεων και πραγματοποιούνται επίσης, μέθοδοι τοπικής αναζήτησης με βάση τα δέντρα [390].
“The Dynamic-Q optimization method: An alternative to SQP?”	(όχι)	(ναι)	104	6.12	Ένας αλγόριθμος με χαμηλές απαιτήσεις αποθήκευσης για το χειρισμό προβλημάτων βελτιστοποίησης με περιορισμούς προτείνεται στο συγκεκριμένο άρθρο. [361]
“A reassessment of the Shyman-Fatti dynamic search trajectory method for unconstrained global optimization”	(όχι)	(ναι)	8	0.80	Οι Shyman και Kok επανεκτιμούν τον γνωστό μ. Η μεθοδολογία του Shyman-Fatti [359] και τον συγκρίνουν με την ευρέως χρησιμοποιούμενη μέθοδο διαφοροεξέλιξης των Storti και Price [375] σε μια σειρά προβλημάτων [366].
“Global minimum cost design of a welded square stiffened plate supported at four corners”	(ναι)	(όχι)	14	1.55	Αυτή η εργασία εξετάζει ένα πολύ γνωστό πρόβλημα, αυτό των μ σχετικά με τις [124].

Η παραπάνω μέθοδος θεωρείται ως η πιο απλή μέθοδος Runge-Kutta, και μάλιστα πρώτης τάξης ως προς την ακρίβεια. Το σφάλμα σε ένα μόνο βήμα είναι ανάλογο με το τετράγωνο του μεγέθους βήματος $h > 0$ (το σφάλμα για ένα μοναδικό βήμα συμπεριφέρεται σαν το $O(h^2)$). Σε αυτή την περίπτωση, ο αντίστοιχος πίνακας Butcher είναι ο ακόλουθος [49]:

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

Είναι προφανές ότι αν εφαρμόσουμε τη μέθοδο του Euler (3.69) στο ακόλουθο αυτόνομο πρόβλημα αρχικών τιμών:

$$\dot{x} = r F(x); \quad x(0) = x_0; \quad (3.70)$$

λαμβάνουμε την παραδοσιακή μέθοδο του Cauchy για τη βελτιστοποίηση μιας αντικειμενικής συνάρτησης $F : \mathbb{R}^N \rightarrow \mathbb{R}$:

$$x^{k+1} = x^k - h r F(x^k); \quad k = 0; 1; 2; \dots; \quad (3.71)$$

Μπορούμε να λάβουμε ακριβέστερα αποτελέσματα για την επίλυση του προβλήματος 3.68, χρησιμοποιώντας μεθόδους Runge-Kutta, και συγκεκριμένα τάξης ακρίβειας δύο, για τις οποίες το σφάλμα για ένα βήμα συμπεριφέρεται σαν το $O(h^3)$. Για παράδειγμα, μπορούμε να χρησιμοποιήσουμε τις ακόλουθες γνωστές και ευρέως χρησιμοποιούμενες μεθόδους Runge-Kutta:

- (α) Τη μέθοδο Euler-Cauchy, η οποία σχετίζεται με την τετραγωνική μέθοδο μέσου σημείου (midpoint quadrature method),
- (β) Τη μέθοδο Heun, η οποία σχετίζεται με τετραγωνικό τύπο του τραπεζίου και
- (γ) Τη μέθοδο Ralston, η οποία παρουσιάζει το μικρότερο τοπικό σφάλμα αποκοπής.

Συγκεκριμένα, η μέθοδος Euler-Cauchy (ή άμεση ή μη πεπλεγμένη μέθοδος μέσου σημείου) για την επίλυση του αυτόνομου προβλήματος αρχικών τιμών (3.68) δίνεται από τη σχέση:

$$y_{n+1} = y_n + hf \left(y_n + \frac{1}{2} hf(y_n) \right); \quad h = x_{n+1} - x_n; \quad (3.72)$$

και ο αντίστοιχος πίνακας του Butcher είναι ο εξής:

$$\begin{array}{c|cc} 0 & & \\ 1=2 & 1=2 & \\ \hline & 0 & 1 \end{array}$$

Επιπλέον, η μέθοδος του Heun δίνεται από:

$$y_{n+1} = y_n + \frac{1}{2} hf(y_n) + \frac{1}{2} hf(y_n + hf(y_n)); \quad h = x_{n+1} - x_n; \quad (3.73)$$

και ο αντίστοιχος πίνακας του Butcher είναι ο εξής:

$$\begin{array}{c|cc} 0 & & \\ 1 & 1 & \\ \hline & 1=2 & 1=2 \end{array}$$

Επίσης, η μέθοδος Ralston δίνεται από:

$$y_{n+1} = y_n + \frac{1}{4}hf(y_n) + \frac{3}{4}hf(y_n + \frac{2}{3}hf(y_n)) ; h = x_{n+1} - x_n; \quad (3.74)$$

και ο αντίστοιχος πίνακας του Butcher είναι ο εξής:

$$\begin{array}{c|c} 0 & \\ \hline 2=3 & 2=3 \\ \hline & 1=4 \quad 3=4 \end{array}$$

Γενικά, για τις αυτόνομες μεθόδους Runge-Kutta δεύτερης τάξης (second-order) και δύο σταδίων (two-stage), ο πίνακας του Butcher μπορεί να γραφτεί ως εξής [49]:

$$\begin{array}{c|c} 0 & \\ \hline c_2 & c_2 \\ \hline & 1 \quad 1=(2c_2) \quad 1=(2c_2) \end{array}$$

και για κάθε πραγματική πεπερασμένη τιμή $c_2 > 0$ οι αντίστοιχες μέθοδοι έχουν τη μορφή:

$$y_{n+1} = y_n + \frac{1}{2c_2} hf(y_n) + \frac{1}{2c_2} hf(y_n + c_2 hf(y_n)) ; h = x_{n+1} - x_n; \quad (3.75)$$

Για παράδειγμα, χρησιμοποιώντας τον παραπάνω τύπο και τις τιμές $c_2 = 1=2; c_2 = 1$ και $c_2 = 2=3$ λαμβάνουμε αντίστοιχα τις μεθόδους (3.72), (3.73) και (3.74). Είναι προφανές ότι, αν εφαρμόσουμε την παραπάνω προσέγγιση στο πρόβλημα αρχικών τιμών (3.70), λαμβάνουμε την ακόλουθη οικογένεια μεθόδων άπειρου πλήθους για την βελτιστοποίηση της αντικειμενικής συνάρτησης $F : \mathbb{R}^N \rightarrow \mathbb{R}$:

$$x^{k+1} = x^k + \frac{1}{2c_2} hr F(x^k) + \frac{1}{2c_2} hr F(x^k + c_2 hr F(x^k)) ; k = 0; 1; 2; \dots; \quad (3.76)$$

Παρατήρηση 3

μ μ μ μ
μ μ Runge-Kutta μ

Παρατήρηση 4

μ μ μ μ μ Runge-Kutta

Οι παραπάνω μέθοδοι εφαρμόστηκαν με επιτυχία σε διάφορες δοκιμαστικές συναρτήσεις και η εμπειρία μας είναι ότι οι μέθοδοι έχουν καλή συμπεριφορά και είναι αξιόπιστες. Τα αποτελέσματα που ελήφθησαν είναι συγκρίσιμα με εκείνα της μεθόδου του Cauchy (μέθοδος απότομης καθόδου) και των μεθόδων συζυγών κλίσεων. Ας υπογραμμίσουμε εδώ ότι, γενικά, οι συγκρίσεις διαφορετικών μεθόδων βελτιστοποίησης είναι ένα δύσκολο ζήτημα. Επίσης, λόγω του ότι «δεν υπάρχει δωρεάν γεύμα στη βελτιστοποίηση», δεν υπάρχει ένας μόνον αλγόριθμος που να λειτουργεί καλά σε όλα τα προβλήματα, και αν βελτιωθεί ένας αλγόριθμος για ένα συγκεκριμένο πρόβλημα, δεν θα αποδώσει καλά σε άλλα προβλήματα. Έτσι, κατά μέσο όρο για όλα τα προβλήματα βελτιστοποίησης, χωρίς επαναδειγματοληψία, όλοι οι αλγόριθμοι βελτιστοποίησης αποδίδουν εξίσου καλά [424] (για μια πρόσφατη ανασκόπηση ο αναγνώστης μπορεί να μελετήσει την εργασία [2]). Επιπλέον, είναι γνωστό ότι, για σχεδόν οποιοδήποτε ζεύγος αλγορίθμων και διάφορα μέτρα απόδοσης αλγορίθμου, όπως ο χρόνος εκτέλεσης ή η ποιότητα των λύσεων, κάθε αλγόριθμος θα έχει καλύτερη απόδοση από άλλους για συγκεκριμένες εισόδους [332].

Στη συνέχεια, παρουσιάζουμε μία στρατηγική για την ανάπτυξη αλγορίθμων ολικής σύγκλισης, η οποία είναι, επίσης, εφαρμόσιμη για τις παραπάνω προτεινόμενες μεθόδους. Οι αλγόριθμοι ολικής σύγκλισης έχουν την ιδιότητα ότι η ακολουθία των επαναλήψεων συγκλίνει σε έναν τοπικό ελαχιστοποιητή της αντικειμενικής συνάρτησης, ξεκινώντας από σχεδόν οποιοδήποτε αρχικό σημείο [92].

Χωρίς βλάβη της γενικότητας θα ξαναγράψουμε τις παραπάνω προτεινόμενες μεθόδους 3.76 στο ακόλουθο επαναληπτικό σχήμα:

$$x^{k+1} = x^k + \Delta x^k; \quad k = 0; 1; \dots; \quad (3.77)$$

όπου, στην k οστή επανάληψη, το $\Delta x^k > 0$ καθορίζει το μέγεθος βήματος και το $\Delta x^k \neq 0$ την κατεύθυνση αναζήτησης.

Το θεωρητικό αποτέλεσμα που παρουσιάζεται παρακάτω, μάς επιτρέπει να εξοπλίσουμε τους αλγορίθμους με μια στρατηγική για την προσαρμογή της κατεύθυνσης αναζήτησης σε μια κατεύθυνση καθόδου. Έτσι, εξασφαλίζεται μια μείωση των τιμών της συνάρτησης σε κάθε επανάληψη και η σύγκλιση σε έναν τοπικό ελαχιστοποιητή της αντικειμενικής συνάρτησης λαμβάνεται από απομακρυσμένα αρχικά σημεία.

Θεώρημα 7 $F : \mathbb{R}^n \rightarrow \mathbb{R}$ μ $L > 0$, μ Lipschitz, μ

$$x^{k+1} = x^k + \Delta x^k; \quad k = 0; 1; \dots; \quad (3.78)$$

$$M^k = \text{diag} \{ \mu_1^k; \mu_2^k; \dots; \mu_n^k \}; \quad \mu_1^k, \mu_2^k, \dots, \mu_n^k \in \mathbb{R}; \quad M^k \neq 0. \quad (3.79)$$

$$\Delta x^k > 0 \quad \text{Wolfe:} \quad (3.79)$$

$$F(x^k + \Delta x^k) - F(x^k) \leq \alpha \Delta x^k \nabla F(x^k)^T (M^k \Delta x^k); \quad (3.80)$$

$$\Delta x^k \nabla F(x^k + \Delta x^k)^T (M^k \Delta x^k) \geq \beta \Delta x^k \nabla F(x^k)^T (M^k \Delta x^k); \quad (3.81)$$

$$0 < \alpha < \beta < 1. \quad \mu \quad \nabla F(x^k) \neq 0, \quad \mu \quad (3.78)$$

Απόδειξη Ακολουθώντας την απόδειξη του Θεωρήματος 6 που συναντάται στο άρθρο [404], λόγω της σχέσης (3.79), η ακολουθία των σημείων $\nabla F(x^k)$ του σχήματος (3.78) ακολουθεί μια κατεύθυνση καθόδου. Επιπλέον, η συνθήκη του Zoutendijk [447]:

$$\sum_{k=1}^{\infty} \cos^2 \theta_k \Delta x^k \nabla F(x^k)^T (M^k \Delta x^k) < 1; \quad (3.82)$$

όπου

$$\cos \theta_k = \frac{\Delta x^k \nabla F(x^k)^T (M^k \Delta x^k)}{\Delta x^k \nabla F(x^k)^T (M^k \Delta x^k)}; \quad (3.83)$$

πληρούται [414, 415, 447]. Στην περίπτωση μας η σχέση (3.83) γίνεται:

$$\cos \theta_k = \frac{r F(x^k) \cdot (M^k \cdot k)}{kr F(x^k)kkM^k k} > 0; \tag{3.84}$$

και έτσι, $\lim_{k \rightarrow \infty} kr F(x^k)k = 0$, που σημαίνει ότι η ακολουθία των κλίσεων της συνάρτησης συγκλίνει στο μηδέν και η ακολουθία των σημείων $f(x^k)g_{k=0}^1$ συγκλίνει καθολικά σε έναν τοπικό ελαχιστοποιητή. Έτσι, το θεώρημα αποδεικνύεται.

Παρατήρηση 5

$$\begin{pmatrix} \mu & & & & \\ & \mu & & & \\ & & \ddots & & \\ & & & \mu & \\ & & & & \mu \end{pmatrix} \quad (n \times n), \quad \mu = \begin{matrix} \mu_1, \mu_2, \dots, \mu_i, \dots, \mu_{i+1}, \dots, \mu_n, \\ \mu \end{matrix} \quad n \times \mu$$

$$M^k = \text{diag} \{ \mu_1^k; \mu_2^k; \dots; \mu_n^k \} \tag{3.79}$$

Παρατήρηση 6

$$\begin{pmatrix} \mu & & & & \\ & \mu & & & \\ & & \ddots & & \\ & & & \mu & \\ & & & & \mu \end{pmatrix} \quad F \quad \mu$$

Μολονότι οι μονοτονικές στρατηγικές σύγκλισης παρέχουν έναν αποδοτικό και αποτελεσματικό τρόπο για να εξασφαλίζεται ότι η συνάρτηση σφάλματος μειώνεται επαρκώς, αυτές έχουν το μειονέκτημα ότι καμία πληροφορία, η οποία ενδέχεται να επιταχύνει τη σύγκλιση, δεν αποθηκεύεται, και ως εκ τούτου δεν χρησιμοποιείται [127]. Για να μετριάσουμε αυτή την κατάσταση, μπορούμε να χρησιμοποιήσουμε μια μ μ - (nonmonotone convergence strategy) που αξιοποιεί τις συσσωρευμένες πληροφορίες σε σχέση με τις πιο πρόσφατες τιμές της αντικειμενικής συνάρτησης.

Στο άρθρο [166], έχει αποδειχθεί ότι η συνθήκη που υποδηλώνει μια μονοτονική μείωση της συναρτησιακής τιμής $F(x^k)$ μπορεί να «χαλαρώσει» και έτσι, να καθοριστεί καθολική σύγκλιση. Για το σκοπό αυτό, οι συγγραφείς του άρθρου [166], καθόρισαν ένα κριτήριο αποδοχής για το μέγεθος του βήματος που μπορεί να θεωρηθεί ως γενίκευση του κανόνα του Armijo [21] και απέδειξαν το ακόλουθο θεώρημα:

Θεώρημα 8

$$f(x^k)g \quad \mu$$

$$x^{k+1} = x^k + d^k; \quad d^k \neq 0; \tag{3.85}$$

$$\alpha > 0, \quad \beta \in (0; 1), \quad \gamma \in (0; 1) \quad W \mu \quad \mu$$

$$(i) \quad \alpha = f(x) \in F(x^0)g \quad \mu$$

$$(ii) \quad \mu \quad q_1; q_2 \quad :$$

$$r F(x^k) \cdot d^k \in q_1 kr F(x^k)k^2; \tag{3.86}$$

$$kd^k \in q_2 kr F(x^k)k; \tag{3.87}$$

Πίνακας 3.2: Συγκρίσεις - Αποτελέσματα

Function	χ_0	hstep	Cauchy-Heun			Cauchy				
			$f(x_{k+1})$	FEval	GEval	$f(x_{k+1})$	FEval	GEval		
z^3 1	[0;1; 1]	0.1	4:78e	05	36	72	4:78e	05	37	37
	[0;5;0;5]	0.1	5:58e	05	30	60	5:58e	05	31	31
	[1;1]	0.1	4:56e	05	36	72	4:56e	05	37	37
Rosenbrock 2D	[0;5;0;5]	0.25	1:52e	03	67	134	1:55e	03	45	45
	[0;2]	0.1	7:82e	05	33	66	1:56e	03	100	100
	[0;2]	0.25	1:38e	03	77	154	1:29e	03	43	43
Rotated Ellipse 2D	[2;2]	0.25	1:05e	03	55	110	7:18e	04	31	31
	[2;2]	0.5	1:01e	03	28	56	5:96e	04	13	13
	[1;1]	0.25	2:96e	04	19	38	1:12e	04	7	7
	[1;1]	0.5	1:12e	04	13	26	1:12e	04	14	14
Sphere	[1;1]	0.1	5:93e	04	67	134	5:52e	04	39	39
	[0;5; 1]	0.1	5:65e	04	65	130	5:39e	04	38	38
	[2;2]	0.1	5:42e	04	74	148	5:79e	04	42	42
Booth	[0;0]	0.1	5:93e	04	67	134	6:17e	05	52	52
	[0;2]	0.1	5:73e	05	48	96	5:73e	05	49	49
	[3; 3]	0.1	6:11e	04	80	160	8:68e	05	52	52
Matyas	[1;1]	0.5	3:10e	02	244	488	3:01e	02	235	235
	[1;0;5]	0.75	2:97e	02	115	230	3:07e	02	110	110
	[2; 2]	0.75	1:01e	03	18	36	1:49e	04	5	5
Himmelblau's	[2;1]	0.02	3:25e	05	34	68	1:30e	05	20	20
	[1;2]	0.02	1:22e	05	19	38	1:22e	05	20	20
	[5; 5]	0.02	5:59e	06	27	54	1:14e	05	26	26
	[2; 2]	0.01	3:46e	05	41	82	2:37e	05	25	25
Three-hump camel	[2;2]	0.25	7:80e	04	27	54	5:32e	04	16	16
	[1; 1]	0.1	7:42e	04	74	148	6:67e	04	38	38
	[0;5;2;5]	0.25	6:38e	04	36	72	5:56e	04	18	18

Πίνακας 3.3: Συγκρίσεις - Αποτελέσματα

Function	x_0	hstep	Cauchy-Heun			Fletcher-Reeves				
			$f(x_{k+1})$	FEval	GEval	$f(x_{k+1})$	FEval	GEval		
z^3 1	[0;1; 1]	0.1	4:78e	05	36	72	NaN			
	[0;5;0;5]	0.1	5:58e	05	30	60	Inf			
	[1;1]	0.1	4:56e	05	36	72	Inf			
Rosenbrock 2D	[0;5;0;5]	0.25	1:52e	03	67	134	NaN			
	[0;2]	0.1	7:82e	05	33	66	1:44e	03	58	116
	[0;2]	0.25	1:38e	03	77	154	NaN			
Rotated Ellipse 2D	[2;2]	0.25	1:05e	03	55	110	1:02e	03	44	88
	[2;2]	0.5	1:01e	03	28	56	1:52e	04	9	18
	[1;1]	0.25	2:96e	04	19	38	9:00e	04	7	14
	[1;1]	0.5	1:12e	04	13	26	NaN			
Sphere	[1;1]	0.1	5:93e	04	67	134	4:80e	04	22	44
	[0;5; 1]	0.1	5:65e	04	65	130	3:00e	04	22	44
	[2;2]	0.1	5:42e	04	74	148	3:80e	04	24	48
Booth	[0;0]	0.1	5:93e	04	67	134	NaN			
	[0;2]	0.1	5:73e	05	48	96	NaN			
	[3; 3]	0.1	6:11e	04	80	160	NaN			
Matyas	[1;1]	0.5	3:05e	02	244	488	3:05e	02	119	238
	[1;0;5]	0.75	2:97e	02	115	230	2:92e	02	58	116
	[2; 2]	0.75	1:01e	03	18	36	1:20e	04	39	78
Himmelblau's	[2;1]	0.02	3:25e	05	34	68	2:66e	06	11	22
	[1;2]	0.02	1:22e	05	19	38	Inf			
	[5; 5]	0.02	5:59e	06	27	54	5:55e	06	19	38
	[2; 2]	0.01	3:46e	05	41	82	4:96e	06	89	178
Three-hump camel	[2;2]	0.25	7:80e	04	27	54	1:43e	04	11	22
	[1; 1]	0.1	7:42e	04	74	148	7:75e	04	20	40
	[0;5;2;5]	0.25	6:38e	04	36	72	2:38e	04	12	24

Πίνακας 3.4: Συγκρίσεις - Αποτελέσματα

Function	x_0	hstep	Euler-Cauchy			Cauchy		
			$f(x_{k+1})$	FEval	GEval	$f(x_{k+1})$	FEval	GEval
z^3 1	[0;1; 1]	0.01	4:8964e 09	56	112	6:36e 05	59	59
	[0.5;0.5]	0.01	6:2640e 09	75	150	6:24e 05	78	78
	[1;1]	0.01	5:7016e 09	41	82	4:82e 05	46	46
Rosenbrock 2D	[0.5;0.5]	0.1	1:4173e 07	115	230	1:46e 03	118	118
	[2; 1]	0.05	1:5352e 07	256	512	1:5088e 07	255	255
	[1.5;2.5]	0.1	1:5014e 07	141	282	1:5754e 07	144	144
Rotated Ellipse 2D	[2;2]	0.1	1:2203e 07	78	156	1:01e 03	83	83
	[2;2]	0.5	8:6025e 08	9	18	5:96e 04	13	13
	[1;1]	0.25	2:7940e 09	3	6	1:12e 04	7	7
	[3;3]	0.25	2:5146e 08	3	6	6:2864e 09	8	8
Sphere	[1;1]	0.25	4:3012e 08	9	18	2:9802e 08	13	13
	[0.5; 1]	0.25	2:6883e 08	9	18	1:8626e 08	13	13
	[2;2]	0.1	5:0383e 08	38	76	5:79e 04	42	42
	[4;3.5]	0.25	1:2015e 08	10	20	2:6310e 08	15	15
Booth	[2;1]	0.05	5:7741e 08	72	144	5:5441e 08	76	76
	[3; 3]	0.06	5:7614e 08	74	148	5:4127e 08	79	79
	[0;0]	0.02	5:9600e 088	208	416	6:0818e 08	212	212
Matyas	[1;1]	0.75	3:1085e 06	153	306	2:98e 02	156	156
	[1;0.5]	0.5	3:0976e 06	164	328	3:55e 03	166	166
	[2; 2]	0.75	3:7253e 09	3	6	1:49e 04	5	5
Himmelblau's	[2;1]	0.01	2:6781e 09	34	68	3:90e 05	38	38
	[2; 2]	0.01	1:3691e 09	24	48	2:37e 05	25	25
	[5; 5]	0.01	6:3398e 010	9	18	3:04e 06	12	12
	[1;2]	0.01	8:2877e 011	12	24	8:01e 05	16	16
Three-hump camel	[1;1]	0.1	1:1877e 08	15	30	6:6662e 08	38	38
	[2;2]	0.07	7:5381e 08	79	158	7:3201e 08	76	76
	[0.5;2.5]	0.25	2:0049e 08	18	36	5:56e 04	18	18

Πίνακας 3.5: Συγκρίσεις - Αποτελέσματα

Function	x_0	hstep	Euler-Cauchy			Fletcher-Reeves		
			$f(x_{k+1})$	FEval	GEval	$f(x_{k+1})$	FEval	GEval
z^3	[0;1; 1]	0.01	4:8964e 09	56	112	3:43e 05	34	68
	[0.5;0.5]	0.01	6:2640e 09	75	150	5:27e 05	43	86
	[1;1]	0.01	5:7016e 09	41	82	5:46e 05	26	52
Rosenbrock 2D	[0.5;0.5]	0.1	1:4173e 07	115	230	1:38e 03	62	124
	[2; 1]	0.05	1:5352e 07	256	512	1:5565	131	262
	[1.5;2.5]	0.1	1:5014e 07	141	282	1:2991e 07	75	150
Rotated Ellipse 2D	[2;2]	0.1	1:2203e 07	78	156	1:02e 03	44	88
	[2;2]	0.5	8:6025e 08	9	18	1:52e 04	9	18
	[1;1]	0.25	2:7940e 09	3	6	9:00e 04	5	10
	[3;3]	0.25	2:5146e 08	3	6	1:1175e 09	6	12
Sphere	[1;1]	0.25	4:3012e 08	9	18	7:5771e 09	9	18
	[0.5; 1]	0.25	2:6883e 08	9	18	4:2455e 08	8	16
	[2;2]	0.1	5:0383e 08	38	76	3:80e 04	24	48
	[4;3.5]	0.25	1:2015e 08	10	20	1:1869e 08	10	20
Booth	[2;1]	0.05	5:7741e 08	72	144	6:1474e 08	41	82
	[3; 3]	0.06	5:7614e 08	74	148	NaN		
	[0;0]	0.02	5:9600e 088	208	416	5:4295e 08	111	222
Matyas	[1;1]	0.75	3:1085e 06	153	306	2:86e 02	80	160
	[1;0.5]	0.5	3:0976e 06	164	328	3:55e 03	87	174
	[2; 2]	0.75	3:7253e 09	3	6	1:20e 04	39	78
Himmelblau's	[2;1]	0.01	2:6781e 09	34	68	1:85e 05	23	46
	[2; 2]	0.01	1:3691e 09	24	48	4:96e 06	89	178
	[5; 5]	0.01	6:3398e 010	9	18	4:19e 06	9	18
	[1;2]	0.01	8:2877e 011	12	24	3:44e 06	10	20
Three-hump camel	[1;1]	0.1	1:1877e 08	15	30	7:7481e 08	20	40
	[2;2]	0.07	7:5381e 08	79	158	5:3874e 08	40	80
	[0.5;2.5]	0.25	2:0049e 08	18	36	2:38e 04	12	24

Πίνακας 3.6: Συγκρίσεις - Αποτελέσματα

Function	χ_0	hstep	OptRunge-Kuttam2			Cauchy		
			$f(\chi_{k+1})$	FEval	GEval	$f(\chi_{k+1})$	FEval	GEval
z^3 1	[0;1; 1]	0.02	3:4485e 09	25	50	3:5495e 09	28	28
	[0.5;0.5]	0.06	3:6343e 09	25	50	1:7661e 09	10	10
	[1;1]	0.06	<i>NaN</i>			8:4901e 011	11	11
Rosenbrock 2D	[0.5;0.5]	0.1	1:4176e 07	115	230	1:46e 03	118	118
	[2; 1]	0.06	1:5615e 07	212	424	1:5407e 07	212	212
	[0;2]	0.09	1:4674e 07	116	232	1:4710e 07	111	111
Rotated Ellipse 2D	[2;2]	0.5	8:6025e 08	9	18	5:96e 04	13	13
	[0.5;0.5]	0.25	6:9849e 010	3	6	2:7940e 09	7	7
	[3;3]	0.25	2:5146e 08	3	6	6:2864e 09	8	8
	[1;1]	0.25	2:7940e 09	3	6	1:12e 04	7	7
Sphere	[1;1]	0.25	4:3012e 08	9	18	2:9802e 08	13	13
	[0.5; 1]	0.25	2:6883e 08	9	18	1:8626e 08	13	13
	[0.5;0.5]	0.25	1:0753e 08	9	18	2:9802e 08	12	12
	[4;4]	0.25	1:3609e 08	11	22	2:9802e 08	15	15
Booth	[1.5;2]	0.05	5:3488e 08	76	152	5:3697e 08	80	80
	[3; 3]	0.05	5:4569e 08	91	182	5:2446e 08	96	96
	[0;0]	0.07	<i>MAXITER</i>			5:0449e 08	58	58
Matyas	[1;1]	0.5	3:0891e 06	232	464	3:0083e 06	235	235
	[2;2]	0.5	8:6025e 06	9	18	5:9605e 08	13	13
	[3;3]	0.9	2:9160e 06	157	314	3:1103e 06	159	159
Himmelblau's	[4;4]	0.01	1:0373e 010	12	24	1:0003e 09	11	11
	[2;2]	0.01	4:2480e 09	30	60	2:9319e 09	35	35
	[3;3]	0.01	3:8341e 09	28	56	4:1105e 09	34	34
	[1;2]	0.01	8:2877e 011	12	24	8:01e 05	16	16
Three-hump camel	[0.5;0.5]	0.1	7:3526e 08	35	70	6:5250e 08	39	39
	[3;3]	0.1	<i>NaN</i>			<i>NaN</i>		
	[1.5;0.5]	0.25	2:7798e 08	22	44	<i>MAXITER</i>		

Πίνακας 3.7: Συγκρίσεις - Αποτελέσματα

Function	χ_0	hstep	OptRunge-Kuttam2			Fletcher-Reeves		
			$f(\chi_{k+1})$	FEval	GEval	$f(\chi_{k+1})$	FEval	GEval
z^3 1	[0;1; 1]	0.02	3:4485e 09	25	50	5:2445e 09	17	34
	[0.5;0.5]	0.06	3:6343e 09	25	50	NaN		
	[1;1]	0.06	NaN			NaN		
Rosenbrock 2D	[0.5;0.5]	0.1	1:4176e 07	115	230	1:38e 03	62	124
	[2; 1]	0.06	1:5615e 07	212	424	1:4221e 07	110	220
	[0;2]	0.09	1:4674e 07	116	232	1:4230e 07	64	128
Rotated Ellipse 2D	[2;2]	0.5	8:6025e 08	9	18	1:52e 04	9	18
	[0.5;0.5]	0.25	6:9849e 010	3	6	2:2511e 09	5	10
	[3;3]	0.25	2:5146e 08	3	6	1:1175e 09	6	12
	[1;1]	0.25	2:7940e 09	3	6	9:00e 04	6	12
Sphere	[1;1]	0.25	4:3012e 08	9	18	7:5771e 09	9	18
	[0.5; 1]	0.25	2:6883e 08	9	18	4:2455e 08	8	16
	[0.5;0.5]	0.25	1:0753e 08	9	18	1:6982e 08	9	18
	[4;4]	0.25	1:3609e 08	11	22	1:2444e 08	11	22
Booth	[1.5;2]	0.05	5:3488e 08	76	152	5:1491e 08	43	86
	[3; 3]	0.05	5:4569e 08	91	182	6:0561e 08	51	102
	[0;0]	0.07	MAXITER			NaN		
Matyas	[1;1]	0.5	3:0891e 06	232	464	3:0471e 06	119	238
	[2;2]	0.5	8:6025e 06	9	18	1:5154e 08	9	18
	[3;3]	0.9	2:9160e 06	157	314	2:8548	82	164
Himmelblau's	[4;4]	0.01	1:0373e 010	12	24	6:2693e 010	8	16
	[2;2]	0.01	4:2480e 09	30	60	2:4232e 09	21	42
	[3;3]	0.01	3:8341e 09	28	56	4:4366e 09	20	40
	[1;2]	0.01	8:2877e 011	12	24	3:4366e 010	10	20
Three-hump camel	[0.5;0.5]	0.1	7:3526e 08	35	70	7:7940e 08	22	44
	[3;3]	0.1	NaN			NaN		
	[1.5;0.5]	0.25	2:7798e 08	22	44	NaN		

3.4 Σύνοψη και μελλοντικές ερευνητικές επιδιώξεις

Σε αυτό το κεφάλαιο παρουσιάζεται μια λεπτομερής έρευνα των παραδοσιακών και ευρέως χρησιμοποιούμενων μεθόδων δυναμικής τροχιάς αναζήτησης για ολική βελτιστοποίηση. Επίσης, παρουσιάζονται οι πιο πρόσφατες από αυτές τις μεθόδους και οι εφαρμογές τους. Επιπλέον, προκειμένου να ενημερωθεί ο αναγνώστης σχετικά με την εφαρμοσιμότητα και το ενδιαφέρον που συγκεντρώθηκε σχετικά με αυτές τις μεθόδους, παρέχεται μια ανάλυση των αναφορών, τις οποίες έλαβαν οι πιο γνωστές και περισσότερο αναφερόμενες μέθοδοι. Με αυτό τον τρόπο γίνεται αντιληπτό, το ενδιαφέρον που συγκέντρωσε κάθε μία από αυτές τις μεθόδους. Ακόμα, έγινε λόγος για τη δυνατότητα εφαρμογής και συνδυασμού κάθε μεθόδου με άλλες μεθόδους είτε κλασικής βελτιστοποίησης είτε με μεθόδους που ανήκουν στον τομέα εξελικτικού υπολογισμού.

Επιπλέον, αναπτύσσονται νέες οικογένειες μεθόδων δυναμικής τροχιάς αναζήτησης απείρου πλήθους για ολική βελτιστοποίηση, που βασίζονται στην αριθμητική επίλυση αυτόνομων συνήθων διαφορικών εξισώσεων. Επίσης, παρέχεται μια στρατηγική για την ανάπτυξη μεθόδων ολικής σύγκλισης, η οποία δύναται να εφαρμοστεί για τις προτεινόμενες οικογένειες μεθόδων και ακόμα, αποδεικνύεται το αντίστοιχο θεώρημα σύγκλισης.

Τέλος, δίνονται τα θεωρητικά αποτελέσματα για την επίτευξη μη-μονοτονικών μεθόδων σύγκλισης, που αξιοποιούν τις συσσωρευμένες πληροφορίες σε σχέση με τις πιο πρόσφατες τιμές της αντικειμενικής συνάρτησης. Σε μια μελλοντική προσπάθεια, θα δοθούν αναλυτικά και εκτενώς, συγκριτικά αριθμητικά αποτελέσματα σχετικά με την εξέταση των παραπάνω μεθόδων και την απόδοσή τους. Επιπλέον, θα μελετήσουμε τις επιδόσεις τους υπό την επίδραση στρατηγικών ολικής και μη μονοτονικής σύγκλισης.

Τέλος, ανοίγονται δρόμοι για την εφαρμογή των προσεγγίσεών μας σε συνήθεις διαφορικές εξισώσεις μεγαλύτερης τάξης. Επιπλέον, εκτός από τις μεθόδους Runge-Kutta δεύτερης τάξης που παρουσιάζονται σε αυτό το κεφάλαιο, θα διερευνηθούν διάφορες άλλες μέθοδοι Runge-Kutta για την αριθμητική επίλυση αυτόνομων προβλημάτων αρχικών τιμών, και θα παρουσιαστούν οι αντίστοιχες μέθοδοι δυναμικών τροχιών αναζήτησης για ολική βελτιστοποίηση.

Η παρούσα μεθοδολογία δύναται να επεκταθεί περαιτέρω, και συγκεκριμένα, στη χρήση αριθμητικών μεθόδων επίλυσης συνήθων διαφορικών εξισώσεων πολλαπλού βήματος, όπως αυτές της οικογένειας Adams-Bashforth και Adams-Moulton. Είναι πολύ σημαντική η θεωρητική ανάλυση, η εφαρμογή τους σε προβλήματα αρχικών τιμών διαφορετικών τάξεων, καθώς και η σύγκρισή τους με ευρέως καθιερωμένες μεθόδους δυναμικών τροχιών, όπως των Snyman-Fatti, αναφορικά με γνωστά προβλήματα βελτιστοποίησης και γνωστές δοκιμαστικές συναρτήσεις διαφορετικών διαστάσεων.

Ολοκληρώνοντας τη σύνοψη του κεφαλαίου, θα θέλαμε να τονίσουμε τις προεκτάσεις που μπορεί να λάβει η παραπάνω μεθοδολογία. Οι απειρία αυτών των μεθόδων δύναται να εφαρμοστεί σε διαφορετικά επιστημονικά πεδία, στα οποία αναζητώνται ελάχιστα ή μέγιστα μιας αντικειμενικής συνάρτησης με ή χωρίς περιορισμούς. Είναι πολύ χρήσιμο να χαρτογραφήσουμε ορισμένες από αυτές τις εφαρμογές (θα αναφερθούμε στο επόμενο κεφάλαιο στο πρόβλημα της πολυ-αντικειμενικής βελτιστοποίησης), αναφέροντας ενδεικτικά την εκπαίδευση τεχνητών νευρωνικών δικτύων, τη βελτιστοποίηση σχήματος σε υλικοτεχνικές δομές, την εύρεση ενός ακραίου σημείου στην επιφάνεια ενός αχαρτογράφητου πλανήτη κ.ά.

Μέρος III

Μηχανική Μάθηση

Προεπεξεργασία δεδομένων στην Εξόρυξη Δεδομένων

Αρχή ήμισυ παντός.

— (427-347 . . .)

Μια τεράστια ποικιλία ζητημάτων επηρεάζει την επιτυχία της μ () (Data Mining (DM)) σε ένα συγκεκριμένο πρόβλημα. Δύο πρωταρχικά και άκρως σημαντικά ζητήματα είναι η (representation) και η (quality) του συνόλου δεδομένων. Συγκεκριμένα, εάν στο σύνολο δεδομένων παρουσιάζονται πολλές (redundant) και (unrelated) ή (noisy) και (unreliable) πληροφορίες, τότε η (Knowledge Discovery (KD)) γίνεται μια πολύ δύσκολη διεργασία. Είναι ευρέως γνωστό ότι τα βήματα της προεπεξεργασίας δεδομένων απαιτούν σημαντικό χρόνο επεξεργασίας στις διάφορες εργασίες της () (Machine Learning (ML)). Έτσι, θα ήταν τρομερά βοηθητικό και επωφελές, εάν μπορούσαν να υπάρξουν διαφορετικοί αλγόριθμοι προεπεξεργασίας με την ίδια αξιόπιστη και αποτελεσματική απόδοση σε όλα τα σύνολα δεδομένων, αλλά κάτι τέτοιο είναι αδύνατο να συμβεί. Για το σκοπό αυτό, στο παρόν κεφάλαιο γίνεται η παρουσίαση των πιο γνωστών και ευρέως χρησιμοποιούμενων σύγχρονων αλγορίθμων για κάθε βήμα προεπεξεργασίας δεδομένων στο πλαίσιο της μ () (Predictive Data Mining (PDM)) [11].

4.1 Μια σύντομη ανασκόπηση

Η μ () (Data preprocessing) ή συχνά και μ (Data preparation) έχει πάντα σημαντική επίδραση στην ικανότητα (generalization) ενός εποπτευόμενου αλγορίθμου (). Λαμβάνοντας υπόψη ότι γνωστές και ευρέως χρησιμοποιούμενες μέθοδοι MM εμπλέκονται συχνά στην μ (), κανείς μπορεί εύκολα να αντιληφθεί τη σημασία της προεπεξεργασίας δεδομένων στην ΕΔ. Γενικά, ο τομέας της MM ασχολείται με την πρόβλεψη ενός αποτελέσματος δεδομένου ενός συγκεκριμένου συνόλου δεδομένων [413]. Πολλές μέθοδοι MM μπορούν να διατυπωθούν ως τυπικά πιθανοτικά μοντέλα. Έτσι, υπό αυτή την έννοια, η MM ομοιάζει σημαντικά με την Στατιστική, αλλά διαφέρει στο γεγονός ότι δεν μεριμνά για τις εκτιμήσεις των παραμέτρων (παρά μόνο για προβλέψεις), και επικεντρώνεται στην υπολογιστική αποτελεσματικότητα [413]. Η ΕΔ είναι ένα πεδίο που βασίζεται σε διάφορες προσεγγίσεις της MM και της Στατιστικής, και επιτυγχάνεται από έναν ειδικό αναφορικά με ένα συγκεκριμένο σύνολο δεδομένων και ένα επιθυμητό αποτέλεσμα ως στόχο [413]. Σε πολλές περιπτώσεις, το σύνολο δεδομένων είναι τεράστιο, περίπλοκο ή είναι πιθανό να παρουσιάζει ιδιαίτερα προβλήματα, όπως όταν υπάρχουν περισσότερα χαρακτηριστικά από τις περιπτώσεις. Γενικά, ο στόχος είναι, είτε να δημιουργηθούν ορισμένες προκαταρκτικές γνώσεις, οι οποίες σχετίζονται με ένα πεδίο εφαρμογής, όπου υπάρχει μικρή εκ των προτέρων

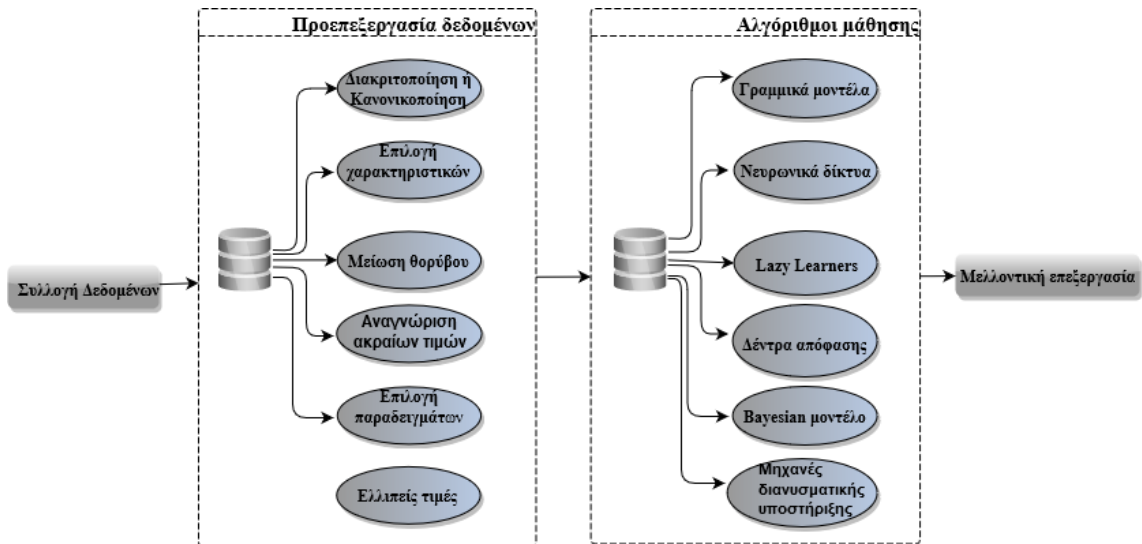
γνώση (a priori) είτε να είναι σε θέση να προβλέψει με ακρίβεια μελλοντικές παρατηρήσεις.

Σε πολλά προβλήματα, το σύνολο δεδομένων που πρέπει να διαχειριστούμε περιέχει μ , τα οποία καθιστούν την εξάλειψη των θορυβωδών περιπτώσεων απαραίτητη και μία από τις πιο δύσκολες και σημαντικές διαδικασίες στη ΜΜ [394]. Ένα άλλο δύσκολο ζήτημα είναι η διάκριση μεταξύ inliers και μ μ μεταξύ των δεδομένων. Αυτές οι τιμές (inliers) είναι εσφαλμένες τιμές δεδομένων που μπορούν να βρεθούν στο εσωτερικό μιας στατιστικής κατανομής και ο εντοπισμός και η διόρθωσή τους αποτελεί ένα πολύ δύσκολο έργο. Έτσι, παρόλο που ο μ μ μ (multivariable data cleaning) συνιστά μια περίπλοκη διαδικασία, είναι σε πολλές περιπτώσεις μια απαραίτητη, αποτελεσματική και αποδοτική ενέργεια [117].

Ένα σύννηθες πρόβλημα, που πρέπει να αντιμετωπιστεί με τη χρήση της λεγόμενης μ μ (data preparation), είναι η διαχείριση των ελλειπών δεδομένων [122, 299]. Σε πολλές περιπτώσεις, ποικίλα σύνολα δεδομένων μπορεί να περιέχουν από κοινού μ (numerical) και μ (categorical) δεδομένα, κάτι που απαιτεί ιδιαίτερη διαχείριση. Από την άλλη πλευρά, είναι γνωστό ότι πολλοί αλγόριθμοι, όπως οι μ () (Logic Learning Algorithms (LLA)), μπορούν να χειριστούν καλύτερα (ή να επιδείξουν καλύτερη απόδοση) δεδομένα μόνο με κατηγορηματικά παραδείγματα. Στην περίπτωση που κάτι τέτοιο λαμβάνει μέρος σε ένα σύνολο δεδομένων, η μ (discretization) των αριθμητικών δεδομένων αποτελεί ένα πολύ σημαντικό ζήτημα στην προεπεξεργασία [113, 133]. Μια πολύ χρήσιμη προσέγγιση για το χειρισμό του παραπάνω προβλήματος, είναι η μ μ μ (grouping of the categorical data). Ως εκ τούτου, το αρχικό σύνολο δεδομένων μετασχηματίζεται από κατηγορηματικό σε αριθμητικό. Από άλλη άποψη, ένα σύνολο δεδομένων με πολλά διαφορετικά είδη δεδομένων ή πολλές τιμές, είναι δύσκολο να διαχειριστεί. Αυτό εγείρει ερωτήματα όπως, «

μ μ () μ Rule learners;» Ως φυσική συνέπεια, μπορεί να υπάρχει σοβαρός κίνδυνος υπερεκτίμησης μεγάλου μέρους των δεδομένων, γεγονός που καθιστά τη μ μια διαδικασία δύσκολη αλλά ταυτόχρονα απαραίτητη. Μια γνωστή και ευρέως χρησιμοποιούμενη τεχνική, μέσω της οποίας καλύπτεται αυτή η φυσική δυσκολία μάθησης από πολύ μεγάλα σύνολα δεδομένων, είναι η μ μ από το αρχικά μεγάλο σύνολο [412]. Αφετέρου, το ερώτημα παραμένει και είναι συναφές, « μ μ ;» Ακόμα, διάφορες τεχνικές έχουν προταθεί για την επίλυση ενός επίσης, σημαντικού προβλήματος, αυτού των μ μ (imbalanced data-classes) [31, 119], το οποίο εξακολουθεί να παραμένει ένα πρόβλημα που απαιτεί περαιτέρω μελέτη.

Τα μ μ με τα οποία θα ασχοληθούμε εκτενώς στο συγκεκριμένο κεφάλαιο παρουσιάζονται στο Σχήμα 4.1. Συγκεκριμένα, εμφανίζεται η ροή που ακολουθείται πριν από την εφαρμογή των μ μ σε μια συγκεκριμένη εργασία Εξόρυξης Δεδομένων. Αρχικά, συλλέγουμε τις πληροφορίες ή λαμβάνουμε σύνολα δεδομένων από μια πηγή. Στη συνέχεια, τα βήματα προεπεξεργασίας εφαρμόζονται κατάλληλα προκειμένου να καθαριστεί το δείγμα ώστε να είναι έτοιμο προς χρήση. Ακολούθως, τα δεδομένα δίνονται ως είσοδος στους αλγόριθμους μάθησης και, τέλος, εφαρμόζονται με τη σειρά τους για την επίλυση ενός συγκεκριμένου προβλήματος (π.χ. μ (classification), μ (regression), μ (pattern recognition), μ (clustering) κ.ά.). Όλα ή μερικά από τα βήματα ΠΔ μπορεί να είναι χρήσιμα σε όλους τους αλγόριθμους που παρουσιάζονται στο Σχήμα 4.1. Αξίζει να τονιστεί μια λεπτομέρεια, που παρουσιάζεται και αφορά το βήμα της μ (discretization) ή της μ (normalization) των δεδομένων. Η μόνη, ενδεχομένως, διαφοροποίηση που μπορεί να συναντήσει ο αναγνώστης, έχει να κάνει με αλγόριθμους που διαχειρίζονται



Σχήμα 4.1: Μοντελοποίηση Προγνωστικής διαδικασίας

μόνο σύνολα δεδομένων, που αποτελούνται από αριθμούς, όπως οι μ (Support Vector Machines (SVM)) ή τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) (Artificial Neural Networks (ANN)). Σε αυτή την περίπτωση, πρέπει να προχωρήσουμε σε κανονικοποίηση των δεδομένων και όχι σε διακριτοποίηση. Σε κάθε άλλη περίπτωση, οποιοδήποτε βήμα προεπεξεργασίας μπορεί να εφαρμοστεί σε οποιονδήποτε αλγόριθμο. Όσον αφορά το πεδίο εφαρμογών, δηλαδή τα προβλήματα στα οποία μπορούν να εφαρμοστούν τα στάδια προεπεξεργασίας, περιλαμβάνουν, αλλά δεν περιορίζονται, προβλήματα όπως η μ και μ . Ωστόσο, δεδομένου ότι εστιάζουμε τη μελέτη μας σε προγνωστικές διεργασίες της ΕΔ, αυτό το κεφάλαιο εστιάζει στις μεθόδους που εφαρμόζονται για την αντιμετώπιση προβλημάτων ταξινόμησης και παλινδρόμησης.

Μία από τις βασικές προκλήσεις που συναντάμε στις προσεγγίσεις της ΕΔ και σχετίζεται με την απόδοση των αλγορίθμων μάθησης, είναι το αποτέλεσμα να είναι μ (dense) και εύκολα μ (clarified). Σημαντικό ρόλο στην εκπλήρωση αυτού του στόχου παίζει η μ του συνόλου δεδομένων. Έτσι, ένα σύνολο δεδομένων με πάρα πολλά χαρακτηριστικά ή χαρακτηριστικά με συσχετισμούς, δεν θα ήταν καλό να συμπεριληφθεί στη διαδικασία μάθησης, δεδομένου ότι αυτού του είδους τα δεδομένα δεν προσφέρουν χρήσιμες πληροφορίες [171]. Έτσι, προκειμένου να χρησιμοποιηθούν μόνο τα πιο μ (informative) δεδομένα, είναι απαραίτητο να επιλέξουμε χαρακτηριστικά, ώστε να μειωθούν οι πληροφορίες που είναι περιττές και δεν σχετίζονται με τον στόχο της μελέτης μας. Επιπλέον, η μ (instance selection) και μ (feature selection) είναι επίσης χρήσιμη στην μ μ μ (distributed Data Mining) και την μ μ μ (knowledge discovery from distributed databases) [78, 356].

Μια εναλλακτική προσέγγιση αναφορικά με τη χρήση πληροφοριακών δεδομένων είναι η προσέγγιση μ μ (feature weighting), η οποία βασίζεται σε κάποιες συνθήκες (π.χ. μετρικές απόστασης (distance metrics)) [284, 444]. Για το σκοπό αυτό, το αρχικό σύνολο δεδομένων δύναται να χωριστεί σε σταθμισμένα χαρακτηριστικά και να εμπλακούν, τελικώς, στη διαδικασία μάθησης, τα χαρακτηριστικά με το μεγαλύτερο βάρος. Επιπλέον, σε πολλές περιπτώσεις είναι πιο χρήσιμο να κατασκευαστούν χαρακτηριστικά μετασχηματίζοντας το πρωτότυπο σύνολο δεδομένων, αντί για την εύρεση ενός καλού υποσυνόλου χαρακτηριστικών [250, 401]. Όπως έχουμε ήδη αναφέρει, στο παρόν κεφάλαιο θα παρουσιάσουμε ζητήματα που αφορούν την ΠΔ. Για περισσότερες και λεπτομερέστερες

αναφορές για το θέμα της ΠΔ, ο αναγνώστης παραπέμπεται στα άρθρα [316] και [148]. Συγκεκριμένα, εστιάζουμε στους πιο γνωστούς και ευρέως χρησιμοποιούμενους, σύγχρονους αλγόριθμους για κάθε βήμα της προεπεξεργασίας δεδομένων στον τομέα της ΠΕΔ.

Η επόμενη ενότητα, καλύπτει το ζήτημα εντοπισμού του (noise) και των μ (outliers). Το θέμα της επεξεργασίας άγνωστων τιμών χαρακτηριστικών περιγράφεται στην Ενότητα 4.3. Στη συνέχεια, στην Ενότητα 4.4, παρουσιάζονται οι διαδικασίες της κανονικοποίησης και διακριτοποίησης των δεδομένων. Η Ενότητα 4.5 και η Ενότητα 4.6, καλύπτουν αντιστοίχως τα ζητήματα επιλογής παραδειγμάτων και χαρακτηριστικών. Το Κεφάλαιο 4, τελειώνει στην Ενότητα 4.7, με μια συζήτηση και μερικές τελικές παρατηρήσεις. Επιπλέον, στο τέλος κάθε παραγράφου, παρέχουμε έναν πίνακα με τις αναφορές¹ που έχουν λάβει οι πιο γνωστές και ευρέως χρησιμοποιούμενες μέθοδοι κάθε βαθμίδας προεπεξεργασίας δεδομένων. Κατά την άποψή μας, αυτό φανερώνει τη σημασία και τη συμβολή κάθε μεθόδου, καθώς και τον αντίκτυπό της στην επιστημονική κοινότητα. Ακόμα, ο αναγνώστης μπορεί να μελετήσει μια σύντομη περιγραφή κάθε μεθόδου, η οποία περιλαμβάνεται σε κάθε πίνακα.

4.2 Εντοπισμός θορύβου και ακραίων τιμών

Ένα από τα κύρια προβλήματα, που πρέπει να αντιμετωπιστούν, είναι ότι μια πλειάδα αλγορίθμων, όπως οι μ (instance-based learners), είναι ευαίσθητοι στην παρουσία (noise). Αυτό, έχει ως αποτέλεσμα την μ (misclassification) η οποία βασίζεται, για παράδειγμα, στον λάθος πλησιέστερο γείτονα. Αυτό συμβαίνει, επειδή τα μ (similarity measures) μπορούν εύκολα να παραποιηθούν από δεδομένα με θόρυβο στις τιμές τους. Μια γνωστή και κοινώς αποδεκτή τεχνική προεπεξεργασίας για την εύρεση και αφαίρεση των θορυβωδών παραδειγμάτων από το σύνολο δεδομένων, είναι η μέθοδος μ (noise filters). Ως αποτέλεσμα, δημιουργείται ένα νέο, βελτιωμένο σύνολο δεδομένων χωρίς θόρυβο, και ως εκ τούτου, το σύνολο δεδομένων που παράγεται μπορεί να χρησιμοποιηθεί ως είσοδος σε έναν αλγόριθμο ΕΔ.

Μια απλή προσέγγιση φιλτραρίσματος είναι ο καθαρισμός δεδομένων μ (variable-by-variable). Στην προσέγγιση αυτή, οι τιμές που θεωρούνται ως «ύποπτες» τιμές, σύμφωνα με συγκεκριμένα κριτήρια, απορρίπτονται ή διορθώνονται. Έτσι, αυτές οι τιμές μπορούν να συμπεριληφθούν στο σύνολο δεδομένων, που θα δοθεί με τη σειρά του ως είσοδος σε έναν αλγόριθμο ΕΔ. Τα κριτήρια που μπορεί να συμπεριληφθούν σε αυτή την προσέγγιση, μεταξύ άλλων, μπορεί να είναι: η γνώμη ενός εμπειρογνώμονα, ο οποίος αξιολογεί τα ύποπτα δεδομένα, είτε ως μ (errors), είτε ως μ (false labeled), ή σε άλλες περιπτώσεις ένας ταξινομητής ο οποίος προβλέπει αυτές τις τιμές ως «ασαφή» (unclear) δεδομένα.

Όπως έχουμε ήδη αναφέρει στην Ενότητα 4.1, εκτός από την επίδραση των θορυβωδών τιμών, είναι δυνατό να εμφανιστούν στο σύνολο δεδομένων τιμές, οι οποίες χαρακτηρίζονται ως (outliers). Στο άρθρο [3], ο συγγραφέας περιγράφει μια μέθοδο για τη δημιουργία μιας ταξινομημένης λίστας τιμών, με εκείνες τις τιμές που είναι υψηλότερα στη λίστα να αναγνωρίζονται ως . Στη συνέχεια αυτής της ενότητας, περιγράφουμε μεθόδους που χρησιμοποιούνται για τον εντοπισμό των ακραίων τιμών, όπως αυτές κατηγοριοποιούνται με τον ακόλουθο τρόπο [64]:

- | | | |
|-----|-------|-----------------------------|
| (α) | μ | (statistics-based methods), |
| (β) | μ | (distance-based methods), |
| (γ) | μ | (density-based methods). |

¹πηγή: google scholar, σύμφωνα με μετρήσεις που έγιναν 1/2019

Μια βασική μέθοδος που βασίζεται στα στατιστικά στοιχεία παρουσιάζεται στο άρθρο [195]. Συγκεκριμένα, δοθέντος ενός στατιστικού μοντέλου ή μιας κατανομής (π.χ. Gaussian κατανομή ή κανονική κατανομή) για το αρχικό σύνολο δεδομένων, οι ακραίες τιμές μπορεί να ανιχνευθούν χρησιμοποιώντας μερικές στατιστικές δοκιμές. Επιπλέον, πολλές μέθοδοι, όπως και η τεχνική που προτάθηκε στο άρθρο [51], εντοπίζουν τις ακραίες αυτές τιμές και ταυτόχρονα εκτιμούν τη μέση τιμή (mean), τη διακύμανση (variance) και τις τιμές που είναι ακραίες.

Στην περίπτωση των μ μ (large datasets), οι συγγραφείς της εργασίας [17] πρότειναν έναν αλγόριθμο ανίχνευσης ακραίων τιμών που (distance-based algorithm) και είναι σε θέση να προσδιορίσει τις m κορυφαίες ακραίες τιμές του συνόλου δεδομένων, όπου m είναι ένας δεδομένος ακέραιος αριθμός. Ο αλγόριθμος που πρότειναν, ονομάζεται HilOut, υπολογίζει το βάρος ενός σημείου ως το άθροισμα της απόστασης που το διαχωρίζει από τους k (k Nearest Neighbors (k-NN)). Ως εκ τούτου, τα σημεία με το υψηλότερο βάρος αναγνωρίζονται ως ακραίες τιμές. Μια παρόμοια προσέγγιση για την ανίχνευση ακραίων τιμών σχετίζεται με το γρήγορο αλγόριθμο, που ονομάζεται RBRP, ο οποίος παρουσιάστηκε στο άρθρο [154]. Αναλυτικότερα, οι συγγραφείς του συγκεκριμένου άρθρου αντιμετώπισαν το πρόβλημα της ανίχνευσης ακραίων τιμών εστιάζοντας σε μεγάλης διάστασης σύνολα δεδομένων. Τα πειράματα που διεξήγαγαν έδειξαν ότι η μέθοδος τους είναι πιο αποδοτική, όσον αφορά τον υπολογιστικό χρόνο, σε σύγκριση με άλλες προσεγγίσεις της ίδιας κατηγορίας. Ένας αλγόριθμος ΕΔ, που ονομάζεται OMABD, ο οποίος βασίζεται στην μ (dissimilarity) των δεδομένων, προτάθηκε πρόσφατα στην εργασία [;]. Σύμφωνα με αυτόν τον αλγόριθμο, η ανίχνευση των ακραίων τιμών βασίζεται σε μια σύγκριση δύο τιμών: (i) η πρώτη τιμή καθορίζει τον βαθμό ανομοιοτητας, ο οποίος κατασκευάζεται από την ανομοιογένεια κάθε αντικειμένου, που λαμβάνει μέρος στο σύνολο δεδομένων και (ii) η δεύτερη, καθορίζει ένα κατώφλι ανομοιοτητας. Στην εργασία [130] προτάθηκε ένας αλγόριθμος υπολογιστικά αποδοτικότερος, για την ανίχνευση των ακραίων τιμών ενός συνόλου δεδομένων. Αυτός ο αλγόριθμος είναι σε θέση να χειριστεί αποτελεσματικά μεγάλης διαστασιμότητας σύνολα δεδομένων χρησιμοποιώντας τη γνωστή μέθοδο (Principal Component Analysis (PCA)). Τέλος, αξίζει να αναφερθούμε στη μέθοδο που έχει προταθεί στην εργασία [65], η οποία υλοποιεί την (neighborhood outlier identification). Έτσι, εμπνευσμένοι από το γνωστό και ευρέως χρησιμοποιούμενο αλγόριθμο των k Κοντινότερων Γειτόνων (k Nearest Neighbors (k-NN)), οι συγγραφείς της δημιούργησαν ένα νέο αλγόριθμο που ονομάζεται μ μ (neighborhood based outlier detection algorithm). Αυτός ο αλγόριθμος, σε σύγκριση με τις κλασικές μεθόδους αποστάσεων, όπως ο αλγόριθμος των κοντινότερων γειτόνων, οι τεχνικές βασισμένες στην απόσταση (distance-based techniques) και η μέθοδος ανίχνευσης ακραίων τιμών βασισμένη στα αναπαραγωγικά νευρωνικά δίκτυα (replicator neural network), υπερέβη τις παραπάνω μεθόδους σε απόδοση, στην περίπτωση των μικτών συνόλων δεδομένων (σύνολα δεδομένων με διαφορετικού τύπου δεδομένα).

Στην ανίχνευση των ακραίων τιμών με βάση την (density), ένα ακραίο παράδειγμα αναγνωρίζεται λαμβάνοντας υπόψη την πυκνότητα του περιβάλλοντος χώρου. Παρά το γεγονός ότι οι μέθοδοι ανίχνευσης ακραίων τιμών με βάση την πυκνότητα έχουν ένα μεγάλο αριθμό πλεονεκτημάτων, η υπολογιστική πολυπλοκότητα που τις χαρακτηρίζει παραμένει ένα ισχυρό και πολύ ανησυχητικό πρόβλημα για την εφαρμογή τους. Μια προσέγγιση για τη μείωση της υπολογιστικής πολυπλοκότητας αυτών των μεθόδων έχει προταθεί στην εργασία [213]. Συγκεκριμένα, οι συγγραφείς της συνδυάζουν δύο γνωστές τεχνικές: το k μ (k Dimensional Tree), που είναι μια γνωστή δομή δεδομένων, και τον αλγόριθμο των k μ , προκειμένου να βελτιώσουν την απόδοση του αλγορίθμου (Local Outlier Factor (LOF)). Τα πειραματικά αποτελέσματα που έχουν σημειώσει σε τρία γνωστά σύνολα δεδομένων του Πανεπιστημίου

της Καλιφόρνια-University of California at Irvine Machine Learning Repository (UCI), έδειξαν ότι η προσέγγισή τους είναι καλύτερη από τον αρχικό αλγόριθμο LOF, όσον αφορά τον υπολογιστικό χρόνο από τον αρχικό αλγόριθμο LOF.

Από την άλλη, στο άρθρο [238] προτάθηκε μια από τις γνωστότερες προσεγγίσεις της κατηγορίας, στην οποία δε γίνεται χρήση της παραμέτρου της πυκνότητας. Η μέθοδός τους, που ονομάζεται iForest, συγκρίθηκε με αρκετές γνωστές μεθόδους, όπως τον αλγόριθμο των (Random Forest), τον αλγόριθμο που χρησιμοποιεί τον

, μία κατηγορία μ και τον αλγόριθμο ORCA. Τα πειραματικά αποτελέσματα έδειξαν ότι ο αλγόριθμος iForest ξεπερνά όλες τις παραπάνω μεθόδους, όσον αφορά τον υπολογιστικό χρόνο και ακόμα, χρειάζεται μικρότερες απαιτήσεις, από άποψη μνήμης, από τις προαναφερθείσες μεθόδους. Συγκεκριμένα, αυτή η μέθοδος επικεντρώνεται τυχαία σε ένα συγκεκριμένο χαρακτηριστικό. Επιλέγει τυχαία μια τιμή μεταξύ της μέγιστης και της ελάχιστης τιμής του συγκεκριμένου χαρακτηριστικού, προκειμένου να «απομονώσει» παραδείγματα. Αυτή η διαδικασία μπορεί να επαναληφθεί με τη χρήση δομής (δεδομένων) δέντρου. Ο απαιτούμενος αριθμός μ (partitions), προκειμένου να εφαρμοστεί η παραπάνω διαδικασία απομόνωσης σε ένα σύνολο δεδομένων, είναι ο ίδιος με το μήκος της διαδρομής από την κορυφή του δέντρου προς τους τελικούς κόμβους, που βρίσκονται στα φύλλα του δέντρου. Όπως διαπίστωσαν, σε μια τέτοια δενδρική δομή δεδομένων, εάν το μήκος της διαδρομής για συγκεκριμένα παραδείγματα είναι μικρότερο από το αναμενόμενο, τότε είναι πολύ πιθανό τα παραδείγματα αυτά να είναι μ (anomalies). Οι

(editing techniques) εστιάζουν ειδικά στην απομάκρυνση του θορύβου ή των ανωμαλιών στο σύνολο εκπαίδευσης [345]. Επομένως, η αξιολόγηση κάθε επιμέρους παραδείγματος στο σύνολο δεδομένων, ανεξάρτητα από το εάν πρέπει να αφαιρεθεί ή όχι, είναι απαραίτητη και πολύ σημαντική. Οι προσεγγίσεις για την ανίχνευση ακραίων τιμών ή ανωμαλιών και το φιλτράρισμα θορύβου στο σύνολο δεδομένων μπορούν να κατηγοριοποιηθούν ως εξής [15]: (i) εποπτευόμενες ή επιτηρούμενες, (ii) ημι-εποπτευόμενες ή ημι-επιτηρούμενες και (iii) χωρίς εποπτεία ή μη επιτηρούμενες. Οι εποπτευόμενες και ημι-εποπτευόμενες μέθοδοι, ετικετοποιούν τα δεδομένα, προκειμένου να δημιουργηθεί ένα μοντέλο που ξεχωρίζει τις ακραίες τιμές από τις τιμές των inliers. Από την άλλη πλευρά, οι μη εποπτευόμενες προσεγγίσεις δεν απαιτούν κανένα ετικετοποιημένο αντικείμενο και ανιχνεύουν τις ακραίες τιμές, ως σημεία που είναι αρκετά διαφορετικά από τα υπόλοιπα σημεία του συνόλου δεδομένων.

Ένα άλλο σημαντικό ζήτημα, που σχετίζεται με το θόρυβο στα σύνολα δεδομένων, είναι η περίπτωση όπου ο θόρυβος επηρεάζει τα χαρακτηριστικά εξόδου. Είναι ευρέως γνωστό, ότι τα θοδουβώδη δεδομένα επηρεάζουν τα χαρακτηριστικά εισόδου και εξόδου σε έναν αλγόριθμο. Στο άρθρο [48] έχει εισαχθεί μια μ μ (Ensemble noise Filter approach), η οποία βελτιώνει το φιλτράρισμα με τη χρήση ενός συνόλου μαθητών (learners). Έτσι, διεξήγαγαν υπολογιστικά πειράματα, που περιελάμβαναν μεμονωμένους ταξινομητές και ένα σύνολο ταξινομητών, ελέγχοντας τη μέθοδο φιλτραρίσματος που στηρίζεται σε σύνολα ταξινομητών σε διάφορα σύνολα δεδομένων. Τα πειραματικά αποτελέσματα έδειξαν ότι η ακρίβεια ταξινόμησης ήταν καλύτερη, όταν χρησιμοποιήθηκε η προσέγγιση φιλτραρίσματος των πολλαπλών μαθητών. Επιπλέον, στην εργασία [90], παρουσιάστηκε μια πολύ χρήσιμη μελέτη για την αξιολόγηση διαφορετικών τεχνικών μ (noise reduction), σχετικά με τη μορφή των παραδειγμάτων στα οποία επικεντρώνονται οι τεχνικές. Τα αποτελέσματα που προέκυψαν από την μελέτη [336], έδειξαν ότι υπάρχει αξιοσημείωτη σχέση μεταξύ των μετρικών πολυπλοκότητας (complexity metrics) ενός συνόλου δεδομένων και της αποδοτικότητας και αποτελεσματικότητας διαφορετικών μεθόδων φιλτραρίσματος θορύβου, που σχετίζονται με τους ταξινομητές πλησιέστερων γειτονικών.

Η ιδέα συνδυασμού διαφορετικών ταξινομητών με διαφορετικά χαρακτηριστικά υποστηρίχθηκε στην εργασία [335]. Συγκεκριμένα, προτάθηκε μια μέθοδος που συνδυάζει την απόφαση πολλών ταξινομητών, προκειμένου να γίνει ανίχνευση θορύβου, ενωποιώντας έτσι,

τα διαφορετικά χαρακτηριστικά ταξινομητών μέσω ενός επαναληπτικού σχήματος. Έτσι, το σχήμα φιλτραρίσματος που εισήχθη, αναγνωρίζει το θόρυβο αποφεύγοντας την ανίχνευση θορυβωδών παραδειγμάτων σε κάθε νέα επανάληψη της διαδικασίας. Επίσης, στο άρθρο [144], οι συγγραφείς του εξέτασαν τον τρόπο με τον οποίο μπορεί να βελτιωθεί η ετικετοποιημένη ανίχνευση θορύβου μέσα από τη χρήση ενός μ μ (ensemble of noise filtering schemes). Στην προσέγγιση αυτή, καταγράφεται μια προσπάθεια να περιοριστούν τα θορυβώδη δεδομένα στο τελικό σύνολο δεδομένων, καθώς και να αφαιρεθούν τα άχρηστα δεδομένα (δεδομένα με «κακή» πληροφορία). Έτσι, με τη δημιουργία μ - (meta-features) από κατεστραμμένα ή φθαρμένα σύνολα δεδομένων, έχουν παραγάγει ένα μ μ - μ (meta-learning model) που προβλέπει τα θορυβώδη δεδομένα σε ένα νέο σύνολο δεδομένων.

Επιπρόσθετα, έχει αποδειχθεί [112] ότι μερικές βασικές αρχές που χρησιμοποιούνται στις ΜΔΥ έχουν την ιδιαιτερότητα να συλλαμβάνουν τα μη-σημασμένα παραδείγματα ως διανύσματα υποστήριξης. Λόγω του γεγονότος ότι δεν μπορεί να υπάρξει μία μέθοδος που να ξεπερνά όλες τις άλλες σε ένα συγκεκριμένο πρόβλημα αναγνώρισης θορύβου, η μελέτη που παρουσιάστηκε στο άρθρο [144] έχει μεγάλη σημασία, καθώς εκθέτει μια συγκρητική μελέτη για την αναμενόμενη απόδοση συγκεκριμένων μεθόδων φιλτραρίσματος θορύβου αναφορικά με συγκεκριμένες εργασίες της ΕΔ.

Ο Πίνακας 4.1 παρουσιάζει μία σύντομη περιγραφή των πιο γνωστών και ευρέως χρησιμοποιούμενων μεθόδων, που σχετίζονται με τα ζητήματα ανίχνευσης θορύβου και ακραίων τιμών. Ακόμα, ο αναγνώστης μπορεί να βρει πληροφορίες για την εφαρμογή αυτών των μεθόδων. Επιπλέον, παρουσιάζονται πληροφορίες σχετικά με τις αναφορές που έχουν λάβει² από την επιστημονική κοινότητα, γεγονός που δίνει μια κατεύθυνση για την απήχηση και τη σημαντικότητα κάθε προσέγγισης.

²πηγή: google scholar, σύμφωνα με μετρήσεις που έγιναν 1/2019

Τίτλος	Ανίχνευση θορύβου	Ανίχνευση ακραίων τιμών	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Identifying mislabeled training data'	(ναι)	(όχι)	578	30.42	Σκοπός της συγκεκριμένης εργασίας είναι η αναγνώριση και η εξάλειψη των εσφαλμένα ετικετοποιημένων παραδειγμάτων εκπαίδευσης. Για να επιτευχθεί αυτό, οι συγγραφείς της εργασίας πρότειναν μια μ μ (ensemble of classifiers) για να φιλτράρουν τα θορυβώδη παραδείγματα των δεδομένων εκπαίδευσης και να βελτιώσουν τη διαδικασία ταξινόμησης [48].
'Outlier mining in large high-dimensional data sets'	(όχι)	(ναι)	368	28.31	Η προτεινόμενη μέθοδος βασίζεται σε αποστάσεις, προκειμένου να υπολογιστούν και να διαχωριστούν οι ακραίες τιμές από το υπόλοιπο σύνολο δεδομένων. Οι συντάκτες της εργασίας παρείχαν έναν αλγόριθμο, που υπολογίζει τα υψηλότερα n δεδομένα του συνόλου που αποτελούσαν μ (outliers), όπου n είναι ένας δεδομένος αέρατος αριθμός [17].
'Review of outlier detection'	(όχι)	(ναι)	31	2.58	Υποθέτοντας ότι το σύνολο δεδομένων υπακούει σε μια συγκεκριμένη κατανομή, οι συγγραφείς του άρθρου υπολογίζουν τις ακραίες τιμές του συνόλου δεδομένων σύμφωνα με συγκεκριμένα στατιστικά κριτήρια [195].
'Fast mining of distance-based outliers in high-dimensional datasets'	(όχι)	(ναι)	215	21.50	Αυτή η προσέγγιση είναι παρόμοια με τη μέθοδο [17] για την ανίχνευση των ακραίων τιμών με βάση τις αποστάσεις. Ο προτεινόμενος αλγόριθμος εστιάζει σε μεγάλα σύνολα δεδομένων και παρέχει ένα γρήγορο προσδιορισμό των πλησιέστερων γειτόνων [154].

Τίτλος	Ανίχνευση θορύβου	Ανίχνευση ακραίων τιμών	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Outlier identification in high dimensions'	(όχι)	(ναι)	315	31.50	Η προτεινόμενη μέθοδος χρησιμοποιεί βασικές πτυχές της για τον εντοπισμό των ακραίων τιμών. Η προτεινόμενη προσέγγιση είναι κατάλληλη για τη χρήση συνόλων δεδομένων μεγάλης διάστασης και απαιτεί λιγότερο υπολογιστικό χρόνο, όσον αφορά άλλες παρόμοιες προσεγγίσεις [130].
'Neighborhood outlier detection'	(όχι)	(ναι)	60	7.50	Προτείνεται ένας αλγόριθμος για την ανίχνευση των ακραίων τιμών, ο οποίος βασίζεται στη γειτονιά των σημείων. Συγκεκριμένα, οι συντάκτες της εργασίας πρότειναν μια νέα μέτρηση, που χρησιμοποιεί τη γειτονιά του κάθε παραδείγματος για να προσδιορίσει εάν το παράδειγμα είναι ακραία τιμή ή όχι [65].
'A comparison of outlier detection algorithms for ITS data'	(όχι)	(ναι)	65	8.125	Οι συγγραφείς του συγκεκριμένου άρθρου παρείχαν τρεις κύριους τύπους αλγορίθμων ανίχνευσης ακραίων τιμών: έναν βασισμένο σε , έναν βασισμένο στις και έναν βασισμένο στην . Επίσης, παρουσίασαν μια σύγκριση αυτών των μεθόδων σε μ μ (intelligent transportation systems) [64].
'Noise reduction for instance-based learning with a local maximal margin approach'	(ναι)	(όχι)	48	6.00	Προτείνεται μια προσέγγιση για τη μείωση του θορύβου που είναι ανταγωνιστική, όσον αφορά το γνωστό ταξινομητή k NN. Αυτή η νέα προσέγγιση, που ονομάζεται μ , είναι μια τροποποίηση του ευρέως χρησιμοποιούμενου ταξινομητή SVM και ανταποκρίνεται πραγματικά καλά σε μεγάλα και θορυβώδη σύνολα δεδομένων [345].

Τίτλος	Ανίχνευση θορύβου	Ανίχνευση ακραίων τιμών	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Outlier detection in large data sets'	(όχι)	(ναι)	42	6.00	Σε αυτή την προσέγγιση, οι συγγραφείς της παρείχαν ένα μοντέλο, που εκτός από την ανίχνευση ακραίων τιμών, υπολογίζει επιπλέον τη μέση τιμή και τη διακύμανση [51].
'Fast outlier detection for very large log data'	(όχι)	(ναι)	37	5.29	Η δομή δεδομένων K D Tree και ο γνωστός ταξινομητής K NN συνδυάζονται, προκειμένου να μειωθεί το υπολογιστικό κόστος της ανίχνευσης ακραίων τιμών με βάση τον παράγοντα της πυκνότητας [213].
'Profiling instances in noise reduction'	(ναι)	(όχι)	14	2.33	Η σύγκριση μεταξύ μεθόδων μείωσης του θορύβου, σχετικά με τους μ (instance based learners) παρουσιάζεται στη συγκεκριμένη εργασία. Επιπλέον, οι συγγραφείς της μελετούν τις περιπτώσεις αφαίρεσης παραδειγμάτων, και ειδικότερα τον τύπο αυτών των παραδειγμάτων [90].
'An introduction to outlier analysis'	(όχι)	(ναι)	42	8.40	Εφαρμόζονται τα βασικά μοντέλα ανίχνευσης ακραίων τιμών, καθώς και τα ευρέως χρησιμοποιούμενα μοντέλα μετα-μάθησης για τον εντοπισμό των ακραίων τιμών [3].
'Predicting noise filtering efficacy with data complexity measures for nearest neighbor classification'	(ναι)	(όχι)	59	11.80	Παρουσιάζεται μια μέθοδος που επιλέγει το κατάλληλο φίλτρο θορύβου για ένα δεδομένο σύνολο θορυβωδών δεδομένων. Η μέθοδος βασίζεται σε μέτρα πολυπλοκότητας δεδομένων και βελτιώνει τα αποτελέσματα του γνωστού ταξινομητή k NN [336].
'Exploiting domain knowledge to detect outliers'	(όχι)	(ναι)	9	2.25	Το κύριο θέμα αυτής της εργασίας είναι η σύγκριση αλγορίθμων που αντιμετωπίζουν το πρόβλημα αναγνώρισης ακραίων τιμών. Η κύρια διαφορά τους είναι ο απαιτούμενος χρόνος εκτέλεσης [15].

Τίτλος	Ανίχνευση θορύβου	Ανίχνευση ακραίων τιμών	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Active cleaning of label noise'	(ναι)	(όχι)	13	6.50	Προτείνεται μια μέθοδος που βασίζεται στην προσέγγιση των SVM για τον εντοπισμό των δεδομένων, τα οποία μπορεί να είναι ακραίες τιμές. Τα δεδομένα που προσδιορίζονται ως «ύποπτα» καταργούνται από το σύνολο δεδομένων [112].
'Noise detection in the meta-learning level'	(ναι)	(όχι)	12	6.00	Οι συγγραφείς του άρθρου, μέσω της δημιουργίας μεταχαρακτηριστικών, έδωσαν μια μέθοδο που προβλέπει τα θορυβώδη δεδομένα [144].
'Ensembles of label noise filters: a ranking approach'	(ναι)	(όχι)	10	5.00	Αυτή η μελέτη εξετάζει τη δυνατότητα βελτίωσης της διαδικασίας ανίχνευσης θορύβου μέσω ενός συνόλου μεθόδων φιλτραρίσματος θορύβου. Για το σκοπό αυτό, οι συγγραφείς της διενήργησαν εκτεταμένα υπολογιστικά πειράματα [145].
'INFFC: An iterative class noise filter based on the fusion of classifiers with noise sensitivity control'	(ναι)	(όχι)	28	14.00	Προτείνεται ένα επαναληπτικό σχήμα που συνδυάζει διάφορους ταξινομητές για την ανίχνευση θορύβου. Το ποσό των παραδειγμάτων που μπορεί να θεωρηθεί θορυβώδες δεν είναι σταθερό και εξαρτάται από τον χρήστη [335].

Πίνακας 4.1: Συνοπτική περιγραφή, δυνατότητα ανίχνευσης θορύβου και ακραίων τιμών, αριθμός των αναφορών που σχετίζονται με τα θέματα ανίχνευσης θορύβου και ακραίων τιμών που παρουσιάζονται στην Ενότητα 4.2. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.

4.3 Ελλιπείς τιμές χαρακτηριστικών

Το πρόβλημα των μ (missing feature values), που σχετίζονται με (όχι απαραίτητα) μεγάλα σύνολα δεδομένων, είναι ένα από τα κύρια προβλήματα που πρέπει να αντιμετωπιστούν κατά την ΠΔ. Τα περισσότερα σύνολα δεδομένων σε πραγματικές εφαρμογές, είναι σύνολα δεδομένων που περιέχουν ελλιπείς πληροφορίες, δηλαδή λείπει ένα μικρό ή μεγάλο μέρος του συνόλου δεδομένων. Υπάρχουν ταξινομητές που παρουσιάζουν σθεναρή συμπεριφορά σε σύνολα δεδομένων με ελλιπούς τιμές χαρακτηριστικών, όπως ο αλγόριθμος Naive Bayes και ως εκ τούτου, το τελικό αποτέλεσμα της απόφασης δεν επηρεάζεται από τις τιμές που λείπουν. Από την άλλη πλευρά, υπάρχουν ταξινομητές, όπως τα μ και ο αλγόριθμος των k , που απαιτούν προσεκτική διαχείριση, όταν στο σύνολο δεδομένων παρουσιάζονται ελλιπείς πληροφορίες.

Το πρόβλημα που περιγράφηκε παραπάνω, αποτέλεσε βασική μέριμνα της επιστημονικής κοινότητας και έχει σημειωθεί σημαντική πρόοδος από πολλούς ερευνητές στον τομέα αυτό τα τελευταία χρόνια [437]. Ωστόσο, υπάρχουν σημαντικά ζητήματα που απαιτούν προσοχή και θέτουν ερωτήματα, τα οποία καλούνται να απαντήσουν οι εμπειρογνώμονες. Το πιο σημαντικό ερώτημα προέρχεται από την πηγή αυτής της «ατέλειας» ή αλλιώς από τη μη πληρότητα που παρουσιάζεται σε ένα σύνολο δεδομένων. Πιο συγκεκριμένα, «

μ , μ » καθώς και « μ » ενώ στην πραγματικότητα αυτά τα δεδομένα είναι χρήσιμα και απαραίτητα για μια καλύτερη πρόβλεψη από τον αλγόριθμο. Γενικά, η τυχαία έλλειψη δεδομένων σημαίνει ότι η τάση για έλλειψη ενός σημείου (δεδομένου) δεν σχετίζεται κατ' ανάγκη με τα ελλιπόντα στοιχεία, αλλά σχετίζεται με ορισμένα από τα παρατηρούμενα δεδομένα. Σε αυτήν την περίπτωση, είναι ασφαλές να αφαιρέσουμε τα δεδομένα με ελλιπούς τιμές ανάλογα με τις εμφανίσεις τους. Από την άλλη πλευρά, στην άλλη περίπτωση, (δηλαδή στην μη τυχαία έλλειψη κάποιων δεδομένων), με την αφαίρεση στιγμιοτύπων με ελλιπούς τιμές θα μπορούσε να προκληθεί μια μ (bias) στο μοντέλο μάθησης. Έτσι, θα μπορούσαμε να χρησιμοποιήσουμε τον μ μ (imputation) δεδομένων σε αυτή την περίπτωση, αλλά αυτό δε δίνει πάντα καλύτερα αποτελέσματα [437].

Υπάρχουν αρκετές μέθοδοι και τεχνικές για τη διαχείριση ελλιπούς τιμών σε ένα σύνολο δεδομένων, τα οποία μπορούν να διακριθούν ως ακολούθως [70, 202]:

- (α) μ μ μ : Συμπληρώνουμε τις ελλιπούς τιμές με την τιμή που εμφανίζεται πιο συχνά στο δοθέν σύνολο δεδομένων.
- (β) μ μ μ : Ομοίως με την παραπάνω περίπτωση, εκτός από το ότι οι ελλιπούς τιμές συμπληρώνονται από τις τιμές γνωρισμάτων που ανήκουν στην ίδια κλάση.
- (γ) μ μ : Συμπληρώστε τις τιμές χαρακτηριστικών που λείπουν με τη μέση τιμή των διαθέσιμων χαρακτηριστικών του συνόλου δεδομένων. Εναλλακτικά, αντί να χρησιμοποιήσουμε «γενικά» τη μέση τιμή των διαθέσιμων χαρακτηριστικών, μπορεί να χρησιμοποιηθεί ο μέσος όρος των δειγμάτων που ανήκουν στην ίδια κλάση.
- (δ) μ μ μ μ : Αναπτύσσουμε ένα μοντέλο παλινδρόμησης για την περίπτωση των αριθμητικών χαρακτηριστικών, συμπληρώνοντας τα κενά του συνόλου δεδομένων, λαμβάνοντας το αποτέλεσμα της απόφασης του μοντέλου που χτίστηκε, χρησιμοποιώντας το σύνολο των γνωστών χαρακτηριστικών ως προβλέψεις. Ομοίως, αναπτύσσουμε το μοντέλο ταξινόμησης για την

κατηγορική περίπτωση [188, 353].

- (ε) μ : Στην περίπτωση αυτή, τα άγνωστα στοιχεία του συνόλου δεδομένων θεωρούνται ως πλήρεις νέες τιμές για τα χαρακτηριστικά που περιέχουν ελλιπούς τιμές.

Επιπλέον, επισημαίνουμε ότι στο άρθρο [121], διεξήχθη μια περιεκτική πειραματική μελέτη, η οποία βασίζεται σε μ μ μ (imputation methods), προκειμένου να αντιμετωπιστεί το πρόβλημα των ελλιπούς τιμών σε δεκαπέντε διακριτά σύνολα δεδομένων. Τα υπολογιστικά πειράματα έδειξαν ότι κατά μέσο όρο ο καταμερισμός βελτιώνει την επικείμενη ταξινόμηση. Διάφοροι άλλοι αλγόριθμοι έχουν σχεδιαστεί για να καταλογιστούν (impute) οι τιμές στόχου που λείπουν, όπως παραλλαγές του αλγορίθμου k NN, ο αλγόριθμος μ (Shell Neighbors Imputation (SNI)) [435] και η μ μ μ μ [317]. Όσον αφορά την προσέγγιση [435], ο συντάκτης του άρθρου πρότεινε μια παραλλαγή του k NNI αλγορίθμου. Σε αυτή την περίπτωση, μόνο δύο γείτονες συμμετέχουν στην επιλογή των πιο γειτονικών από τις ελλιπούς τιμές δεδομένων και συγκεκριμένα, το αριστερό και το δεξί στοιχείο. Αυτά τα σημεία υπολογίζονται χρησιμοποιώντας τον ορισμό του (shell neighbor). Αυτό αποτελεί την κύρια διαφορά με τον αλγόριθμο k NNI που παίρνει τους k γείτονες, όπου το k είναι σταθερό. Τα πειραματικά αποτελέσματα σε μικτά σύνολα δεδομένων, έδειξαν ότι ο αλγόριθμος SNI ξεπερνά σε επιδόσεις τον αλγόριθμο k NNI σε σχέση με την ταξινόμηση και την ακρίβεια καταλογισμού. Όσον αφορά την προσέγγιση [317], οι συγγραφείς της μέσω υπολογιστικών πειραμάτων που πραγματοποίησαν, συνέκριναν τη μ μ (kernel-based method) που πρότειναν με διάφορες μη-παραμετρικές μεθόδους. Τα αποτελέσματα έδειξαν ότι ο αλγόριθμος μ - μ (non-Parametrical Optimization (POP)) ξεπερνά τις άλλες μεθόδους που στηρίζονται σε στατιστικές παραμέτρους, όπως η μέση τιμή, η συνάρτηση κατανομής και τα ποσοτικά δεδομένα μετά την έλλειψη δεδομένων και την αποδοτικότητα καταλογισμού.

Πρόσφατα προτάθηκε μια προσέγγιση καταλογισμού [436] που ονομάζεται NIIA. Αυτή η μέθοδος καταλογίζει τα δεδομένα που λείπουν, χρησιμοποιώντας πληροφορίες από ελλιπόντα παραδείγματα. Πρόκειται για ένα επαναληπτικό σχήμα υπολογισμού, που έχει σχεδιαστεί για να καταλογίζει με προσοχή τις τιμές στόχου που λείπουν. Η μέθοδος καταλογίζει κάθε τιμή που λείπει διαδοχικά, μέχρι τη σύγκλιση της.

Στην εργασία [248], οι συγγραφείς της προσπάθησαν να λύσουν το πρόβλημα των ελλιπούς τιμών μέσω της χρήσης διαφορετικών μεθόδων καταλογισμού. Συγκεκριμένα, έχουν επικεντρωθεί σε μια εργασία ταξινόμησης και τα πειραματικά αποτελέσματά τους έδειξαν ότι, οι συγκεκριμένες μέθοδοι καταλογισμού των ελλιπών τιμών παρέχουν καλύτερα αποτελέσματα, όσον αφορά την ακρίβεια προβλέψεων.

Οι συγγραφείς του άρθρου [245], παρουσίασαν για πρώτη φορά μια λύση στο πρόβλημα των ελλιπούς τιμών συνδυάζοντας τεχνικές εξελικτικού υπολογισμού, όπως

μ (Genetic Algorithms (GA)), για τον καταλογισμό των δεδομένων. Συγκεκριμένα, οι συγγραφείς της πρότειναν έναν μ - μ (multi-objective GA), ο οποίος ονομάζεται MOGAImp και είναι κατάλληλος για σύνολα δεδομένων μικτών χαρακτηριστικών.

Ένα πολύ δύσκολο και χρονοβόρο πρόβλημα αποτελεί η επιλογή του βέλτιστου συνδυασμού μεταξύ μεθόδων ταξινόμησης και καταλογισμού. Ένα νέο, εναλλακτικό σχήμα που προτείνει έναν τέτοιο συνδυασμό προσαρμοστικά, έχει προταθεί στην εργασία [354]. Για να δημιουργήσουν αυτό το νέο σχήμα και να βρουν το σωστό ζεύγος μεταξύ μεθόδου καταλογισμού και ταξινόμησης, οι συγγραφείς της διενήργησαν ένα σύνολο πειραματικών δοκιμών. Μια σημαντική διαφορά σχετικά με τις προηγούμενες προσπάθειες ήταν ότι, οι συντάκτες της χρησιμοποίησαν μετα-δεδομένα για να καλύψουν τα ελλιπή δεδομένα. Έτσι, απέφυγαν

τον εξαιρετικά υψηλό χρόνο εκτέλεσης. Επιπλέον, μια άλλη προσέγγιση, που καλύπτει τα ελλιπή δεδομένα μέσω προσαρμοστικού υπολογισμού βασισμένου στη

(belief function theory), έχει παρουσιαστεί στο άρθρο [244]. Συγκεκριμένα, οι συγγραφείς της αντιμετωπίζουν το πρόβλημα της ταξινόμησης ελλιπών μοτίβων, καλύπτοντας τις ελλιπούς τιμές μέσω των μεθόδων k -NN και μ -SOM (Self Organizing Map (SOM)). Αυτό συμβαίνει μόνο στις περιπτώσεις, όπου οι διαθέσιμες πληροφορίες δεν επαρκούν για την ταξινόμηση ενός δεδομένου αντικειμένου.

Ο Πίνακας 4.2 παρουσιάζει μια σύντομη περιγραφή και πληροφορίες σχετικά με τις αναφορές που έλαβαν οι πιο γνωστές και ευρέως χρησιμοποιούμενες μέθοδοι, που σχετίζονται με τις ελλιπείς τιμές χαρακτηριστικών.

Τίτλος	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'A SVM regression based approach to filling in missing values'	60	4.61	Χρησιμοποιείται ένα μοντέλο παλινδρόμησης Μηχανών Διανυσματικής υποστήριξης για την πρόβλεψη των ελλειπουσών τιμών για διακριτά συνόλα δεδομένων [188].
'Impact of imputation of missing values on classification error for discrete data'	214	21.4	Οι συγγραφείς του άρθρου εστίασαν τη μελέτη τους σε διακριτά δεδομένα, προκειμένου να εκτιμήσουν την επίδραση πολλών προσεγγίσεων καταλογισμού στη διαδικασία ταξινόμησης. Η μελέτη τους περιλαμβάνει εκτεταμένα υπολογιστικά πειράματα, ώστε να γίνει σύγκριση μεταξύ γνωστών μεθόδων καταλογισμού [121].
'Missing value imputation based on data clustering'	69	6.9	Παρέχεται μια μέθοδος συσταδοποίησης, σχετικά με την πλήρωση των ελλειπόντων δεδομένων. Προκειμένου να επιτευχθεί αυτό, οι συγγραφείς της εργασίας χρησιμοποίησαν χαρακτηριστικά των οποίων οι τιμές χαρακτηριστικών δεν λείπουν και είναι παρόμοια με τα στοιχεία που εμφανίζουν ελλειπούσες τιμές [437].
'POP algorithm: Kernel-based imputation to heat missing values in knowledge discovery from datasets'	41	4.55	Προτείνεται μια μ μ (kernel-based method) και συγκρίνεται με διαφορετικές μη παραμετρικές προσεγγίσεις [317].
'Missing data imputation using statistical and machine learning methods in a real breast cancer problem'	175	21.875	Γίνεται σύγκριση μεταξύ δύο κατηγοριών: συγκρίνονται μέθοδοι Στατιστικής και Μηχανικής Μάθησης. Οι μέθοδοι δοκιμάστηκαν σε ένα πραγματικό πρόβλημα, αυτό του καρκίνου του μαστού [202].
'Missing value imputation on missing completely at random data using multilayer perceptron'	84	12	Παρέχεται μια σύγκριση μεταξύ τριών γνωστών μεθόδων καταλογισμού και ενός αυτοματοποιημένου μοντέλου που βασίζεται σε ΤΝΔ. Τα πειράματα που έκαναν οι συγγραφείς, έδειξαν ότι το μοντέλο του μ Perceptron (Multilayer Perceptron) δίνει καλύτερα αποτελέσματα [353].
'Missing data imputation by utilizing information within incomplete instances'	75	10.71	Ο προτεινόμενος αλγόριθμος είναι ένα επαναληπτικό σχήμα, το οποίο καταλογίζει τις τιμές στόχους που λείπουν με επαναληπτικό τρόπο. Για να γίνει αυτό, ο αλγόριθμος αξιοποιεί τις χρήσιμες πληροφορίες εντός των ελλειπών δεδομένων [436].

Τίτλος	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Shell-neighbor method and its application in missing data imputation'	103	14.71	Μια παραλλαγή του αλγορίθμου k NN για την πλήρωση των ελλιπόντων δεδομένων προτείνεται στη συγκεκριμένη εργασία. Προκειμένου να επιτευχθεί αυτός ο στόχος, ο συγγραφέας της προσπάθισης να βελτιώσει τις προηγούμενες προσπάθειες λαμβάνοντας τα πιο γειτονικά δεδομένα (μόνο αριστερά και δεξιά), σχετικά με τις ελλιπούσες τιμές-στόχους [435].
'On the choice of the best imputation methods for missing values considering three groups of classification methods'	129	21.50	Το πρόβλημα των ελλιπουσών τιμών αντιμετωπίζεται μέσω της χρήσης διαφορετικών προσεγγίσεων καταλογισμού. Αυτή η μελέτη επικεντρώνεται στο πρόβλημα ταξινόμησης [248].
'Missing data in medical datasets: Impute, delete or classify?'	49	9.80	Το προτεινόμενο σχήμα είναι κατάλληλο για να επιλέξετε ποια από τα ελλιπή στοιχεία θα συμπληρωθούν και ποια όχι. Για τη διαχείριση αυτού του στόχου, οι συγγραφείς του άρθρου χρησιμοποίησαν ένα στατιστικό ταξινομητή και ασαφή μοντελοποίηση, ώστε το αποτέλεσμα να είναι πιο ακριβές [70].
'Multi-objective genetic algorithm for missing data imputation'	23	7.67	Σε αυτή τη μελέτη, για πρώτη φορά αντιμετωπίζεται το πρόβλημα του καταλογισμού των ελλιπών δεδομένων χρησιμοποιώντας έναν εξελικτικό αλγόριθμο [245].
'Adaptive imputation of missing values for incomplete pattern classification'	84	42	Εξετάζεται το πρόβλημα της ταξινόμησης ελλιπών μοτίβων που καλύπτουν τις ελλιπούσες τιμές μέσω των μεθόδων k NN και SOM [244].
'Adaptive pairing of classifier and imputation methods based on the characteristics of missing values in data sets'	5	2.5	Στην προσέγγιση αυτή χρησιμοποιήθηκαν μετα-δεδομένα για να συμπληρωθούν τα ελλιπή δεδομένα. Ένα προσαρμοστικό μοντέλο επιλέγει το κατάλληλο ζεύγος ταξινομητών για το σκοπό αυτό [354]

Πίνακας 4.2: Συνοπτική περιγραφή και αριθμός των αναφορών που σχετίζονται με τα θέματα ελλιπουσών τιμών χαρακτηριστικών που παρουσιάζονται στην Ενότητα 4.3. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.

4.4 Κανονικοποίηση και διακριτοποίηση

Σε πολλά σύνολα δεδομένων τα οποία καλούμαστε να διαχειριστούμε, υπάρχουν πολύ συχνά μεγάλες διαφορές μεταξύ των τιμών που σημειώνουν τα χαρακτηριστικά τους, όπως η μέγιστη και η ελάχιστη τιμή, π.χ. 0.001 και 10.000. Γενικά, το ζήτημα αυτό δεν είναι επιθυμητό και απαιτεί προσεκτική παρέμβαση, προκειμένου να γίνει ένας κλιμακωτός μετασχηματισμός κατά τέτοιο τρόπο, ώστε όλες οι τιμές χαρακτηριστικών να είναι κατάλληλες και αποδεκτές. Αυτή η διαδικασία είναι γνωστή ως μ (feature scaling) ή μ (data normalization) (στην περίπτωση της προεπεξεργασίας δεδομένων) και είναι απαραίτητη και πολύ σημαντική για διάφορους ταξινομητές, όπως τα Τεχνητά Νευρωνικά Δίκτυα, οι Μηχανές Διανυσματικής Υποστήριξης, ο αλγόριθμος των k Κοντινότερων Γειτόνων, καθώς και οι μ (fuzzy classifiers), οι οποίοι δεν μπορούν να αποδώσουν καλά, αν προκύπτουν μεγάλες διαφορές μεταξύ των τιμών των χαρακτηριστικών στο σύνολο δεδομένων.

Οι πιο κοινές μέθοδοι για αυτό το πεδίο εφαρμογής είναι οι εξής:

- (α) $\min \max$ $(-\mu)$ μ
 $\mu [0;1]$. Ο τύπος που παρέχει νέες τιμές χαρακτηριστικών είναι ο παρακάτω:

$$v^{\beta} = \frac{v - \min_value}{\max_value - \min_value};$$

- (β) $\min \max$ $(-\mu)$ μ
 $\mu [a;b]$. Ο τύπος που παρέχει νέες τιμές χαρακτηριστικών είναι ο ακόλουθος:

$$v^{\beta} = \frac{v - \min_value}{\max_value - \min_value} (\text{newmax_value} - \text{newmin_value}) + \text{newmin_value};$$

- (γ) Z (Z score) (standardization). Ο τύπος που δίνει τη νέα τυποποιημένη τιμή χαρακτηριστικών είναι:

$$v^{\beta} = \frac{v - \text{mean}}{\text{stand_dev}};$$

- (δ) μ μ μ (unit length scaling):

$$v^{\beta} = \frac{v}{\|v\|};$$

όπου ' v ' είναι η αρχική τιμή χαρακτηριστικού, ενώ ' v^{β} ' είναι η νέα, κανονικοποιημένη τιμή δεδομένων, ' \min_value ' και ' \max_value ' είναι αντίστοιχα οι ελάχιστες και οι μέγιστες τιμές χαρακτηριστικών πριν από τη διαδικασία κανονικοποίησης. Στη δεύτερη μέθοδο, το διάστημα $[a; b]$ υποδηλώνει μια συγκεκριμένη περιοχή, στην οποία θέλουμε να μετατρέψουμε το αρχικό σύνολο δεδομένων. Έτσι, η νέα μέγιστη τιμή ' newmax_value ' και η νέα ελάχιστη τιμή ' newmin_value ' υποδεικνύουν την επιθυμητή νέα μέγιστη και νέα ελάχιστη τιμή του κανονικοποιημένου συνόλου δεδομένων (a και b είναι οι τιμές που λαμβάνουν αντίστοιχα). Επιπλέον, στην τρίτη μέθοδο, η μέση τιμή ' mean ' υποδηλώνει τη μέση τιμή του διανύσματος

χαρακτηριστικών και *'stand_dev'* αντιπροσωπεύει την τυπική απόκλιση. Ουσιαστικά, αυτή η μέθοδος είναι ο αριθμός τυπικών αποκλίσεων, με τις οποίες η τιμή μιας παρατήρησης ή ενός δεδομένου είναι πάνω από τη μέση τιμή από αυτό που παρατηρείται ή μετράται. Τέλος, το *'k k'* υποδεικνύει μια απόσταση (norm), π.χ. το Ευκλείδειο μήκος του διανύσματος χαρακτηριστικών ή την απόσταση Manhattan. Εν κατακλείδι, η επιλογή μιας από τις παραπάνω μεθόδους κανονικοποίησης εξαρτάται σαφώς από το σύνολο δεδομένων που θέλουμε να κανονικοποιήσουμε (ή να μετασχηματίσουμε).

Ένα άλλο σημαντικό ζήτημα, που πρέπει να αντιμετωπιστεί, είναι αυτό των συνεχών τιμών, που μπορεί να λαμβάνουν τα χαρακτηριστικά ενός συνόλου δεδομένων. Συνήθως, τα χαρακτηριστικά του δείγματος μπορούν να αποκτήσουν συνεχείς τιμές ή τιμές από ένα ευρύ φάσμα. Κάτι τέτοιο μπορεί να συνθέσει μια χρονοβόρα διαδικασία χειρισμού ή, σε πολλές περιπτώσεις, μπορεί να καταστήσει ανεπαρκή τη διαχείριση δεδομένων για τις διαδικασίες που χρησιμοποιούνται στον τομέα της ΠΕΔ. Έτσι, το κύριο αποτέλεσμα αυτών και μία λύση στο παραπάνω πρόβλημα είναι η διαδικασία (discretization).

Η (discretization process) μπορεί στην πραγματικότητα να είναι πολύ χρήσιμη σε ένα σύνολο ταξινομητών, όπως τα (Decision Trees) ή οι Bayesian μ (Bayesian classifiers). Από την άλλη πλευρά, η επιστημονική κοινότητα πρέπει να επιλύσει αρκετά άλλα ζητήματα σχετικά με αυτή τη διαδικασία, όπως τον εντοπισμό των καταλληλότερων άκρων ενός διαστήματος ή την αρτιότητα για τη διακριτοποίηση ενός αριθμητικού εύρους τιμών. Αυτά τα θέματα είναι ανοικτά προς επίλυση και απαιτούν εκτεταμένη μελέτη. Για περισσότερες λεπτομέρειες σχετικά με τις διαδικασίες διακριτικοποίησης, παραπέμπουμε τον αναγνώστη στις ακόλουθες μελέτες [225, 240, 243].

Η κύρια διαφορά μεταξύ των μ μ (supervised discretization methods) και των μ μ (unsupervised discretization methods) είναι ότι, οι μέθοδοι της πρώτης κατηγορίας διακριτοποιούν τα χαρακτηριστικά του συνόλου δεδομένων σύμφωνα με την κατηγορία στην οποία ανήκουν, ενώ της δεύτερης κατηγορίας δεν το κάνουν. Από την άλλη πλευρά, η διάκριση των μεθόδων σε μεθόδους «πάνω-κάτω» (top-down) και «κάτω-πάνω» (bottom-up) αφορά την αρχικοποίηση του αρχικού διαστήματος. Στις μεθόδους «πάνω-κάτω», η εκκίνηση πραγματοποιείται με το διάστημα αρχικοποίησης και σταδιακά προχωράμε σε μικρότερα. Από την άλλη πλευρά, στις μεθόδους «κάτω-πάνω» λαμβάνει χώρα η εκκίνηση των διαστημάτων μίας μόνο τιμής και ακολούθως, συντελείται η συγχώνευση των παρακείμενων διαστημάτων. Επιπλέον, η ποιότητα των διαδικασιών διακριτικοποίησης μετράται χρησιμοποιώντας δύο βασικά κριτήρια: (α) την μ (classification performance) και (β) τον μ μ (number of discretization intervals).

Δύο από τις πιο γνωστές και ευρέως χρησιμοποιούμενες μεθόδους διακριτικοποίησης είναι οι παρακάτω [102]: (α) αυτές του μ (equal size) και (β) αυτές της (equal frequency). Και οι δύο μέθοδοι είναι μη εποπτευόμενης μάθησης. Οι μέθοδοι της πρώτης κατηγορίας, όπως υποδηλώνεται από το όνομά τους, βρίσκουν τη μέγιστη και την ελάχιστη τιμή των χαρακτηριστικών και στη συνέχεια, διαιρούν το εύρος που βρέθηκε σε k ίσα διαστήματα. Αυτές της δεύτερης κατηγορίας, βρίσκουν τον αριθμό των τιμών του χαρακτηριστικού και στη συνέχεια, τον διαχωρίζουν σε διαστήματα που περιέχουν τον ίδιο αριθμό παραδειγμάτων.

Ένα πολύ σημαντικό ζήτημα, όσον αφορά τις μεθόδους «πάνω-κάτω» και «κάτω-πάνω», είναι ότι απαιτούν τη συμμετοχή του χρήστη σε ένα σύνολο παραμέτρων, όπως η τροποποίηση των κριτηρίων για την παύση των διαδικασιών διακριτικοποίησης. Ένας γνωστός αλγόριθμος διακριτικοποίησης «κάτω-πάνω», ονομάζεται KMeans και παρουσιάστηκε στο άρθρο [43]. Το πλεονέκτημα αυτής της μεθόδου σε σχέση με τις άλλες μεθόδους της ίδιας κατηγορίας, και ιδιαίτερα τους αλγορίθμους ChiSplit και ChiMerge, είναι ότι χρησιμοποιεί ως κριτήριο παύσης το χ^2 στατιστικό τεστ (chi-square statistical test). Ως αποτέλεσμα, η μέθοδος αυτή βελτιστοποιεί ένα (global criterion) και όχι ένα (local crite-

gion). Επιπλέον, δεν απαιτεί να καθορίζονται από τον χρήστη οποιεσδήποτε παράμετροι. Ακόμα, δεν υστερεί όσον αφορά την πολυπλοκότητα χρόνου, (όπως αποδεικνύεται από τα πειραματικά αποτελέσματα), σύμφωνα με άλλες μεθόδους που βελτιστοποιούν τοπικά κριτήρια. Αυτό καθιστά τον Khiops πολύ χρήσιμο και εύκολο στο χειρισμό του. Στο άρθρο [160] παρουσιάστηκε μια άλλη γνωστή μέθοδος της ίδιας οικογένειας, που ονομάζεται Aneva, η οποία κατορθώνει να παράγει τον ελάχιστο αριθμό διαστημάτων. Συγκεκριμένα, αυτή η μέθοδος έχει συγκριθεί με το γνωστό αλγόριθμο, που ονομάζεται CAIM και επίσης, με άλλες γνωστές προσεγγίσεις γενετικού υπολογισμού. Τα πειραματικά αποτελέσματα δείχνουν ότι ο αλγόριθμος Aneva παράγει πάντοτε το μικρότερο αριθμό διακριτών διαστημάτων και ακόμη, και σε περιπτώσεις με μεγάλο αριθμό κλάσεων, δεν υστερεί σε όρους πολυπλοκότητας. Όσον αφορά τις επιδόσεις του Aneva σε σχέση με τις γενετικές προσεγγίσεις, η απόδοση αυτών των αλγορίθμων είναι ανταγωνιστική και παραπλήσια.

Στην εργασία [388], οι συγγραφείς της πρότειναν έναν αλγόριθμο βαθμιαίας, εποπτευόμενης, «πάνω-κάτω» διακριτοποίησης μέσω του οποίου επιτυγχάνεται καλύτερη ακρίβεια ταξινόμησης. Αυτός ο αλγόριθμος βασίζεται σε ένα συντελεστή «Εκτακτης Ανάγκης» (class-attribute contingency coefficient) και συνδυάζεται με μια άπληστη μέθοδο (greedy method), που έχει επίσης παρουσιάσει καλά αποτελέσματα χρόνου εκτέλεσης. Όπως έχουμε ήδη αναφέρει, η μέτρηση της ποιότητας των αλγορίθμων διακριτοποίησης είναι ένα από τα θέματα που χρήζει μεγάλης προσοχής. Τα υπολογιστικά πειράματα που έχουν διεξαχθεί στη μελέτη [203], έχουν δείξει ότι, πιο διακριτοποιημένα διαστήματα, συνήθως, ισοδυναμούν με λιγότερα σφάλματα ταξινόμησης και ως εκ τούτου, μειώνουν το κόστος της διαδικασίας διακριτικοποίησης δεδομένων.

Στο άρθρο [200] μελετήθηκαν και αναλύθηκαν μέθοδοι διακριτοποίησης βασισμένες (α) στα σφάλματα (error-based), (β) στην εντροπία (entropy-based) και (γ) στο κόστος (cost-based). Οι μέθοδοι της πρώτης κατηγορίας, ομαδοποιούν τις τιμές σε ένα διάστημα με στόχο την ελαχιστοποίηση των σφαλμάτων στο σύνολο εκπαίδευσης, για να επιτευχθεί η βέλτιστη διακριτοποίηση. Από την άλλη πλευρά, στις μεθόδους που βασίζονται στην εντροπία, χρησιμοποιείται η τιμή εντροπίας ενός σημείου για να βρεθεί η $\min \max$ τιμή ενός διαστήματος. Όπως οι μέθοδοι «πάνω-κάτω», οι μέθοδοι διακριτοποίησης με βάση την εντροπία, διαιρούν το αρχικό διάστημα σε μικρότερα, επιλέγοντας τις τιμές που ελαχιστοποιούν την εντροπία, έως ότου τα διαστήματα γίνουν βέλτιστα ή μέχρι να πληρούνται συγκεκριμένα κριτήρια [224]. Αυτή η κατηγορία χρησιμοποιείται ευρέως και η έννοια της εντροπίας έχει αναλυθεί, επίσης, στις μελέτες [170] και [84], προκειμένου να αξιοποιηθούν οι εξαρτήσεις μεταξύ των χαρακτηριστικών, καθώς και να διακριτικοποιήσουν την τάξη των δεδομένων. Από την άλλη πλευρά, στις μεθόδους που βασίζονται σε σφάλματα της τρίτης κατηγορίας, το σφάλμα λαμβάνεται υπόψη ως κόστος εσφαλμένης ταξινόμησης.

Οι Bayesian ταξινομητές είναι από τους πιο γνωστούς ταξινομητές στην ΕΔ και ως εκ τούτου, οι ερευνητές έχουν μεγάλη συμμετοχή σε αυτή την κατηγορία αλγορίθμων. Συγκεκριμένα, για να βελτιωθεί ο βαθμός ακρίβειας στην πρόβλεψη του αφελούς ταξινομητή Bayes (naive Bayes classifier), ειδικότερα στη μείωση της διακύμανσης και της μεροληψίας, στο άρθρο [433] έχει προταθεί ένα σχήμα που συνδυάζει την (proportional discretization) και τη

(fixed frequency

discretization). Μια άλλη προσπάθεια βελτίωσης της ακρίβειας του μ Bayes έχει γίνει στην εργασία [426]. Ο συγγραφέας της, στην προσπάθειά του να παρουσιάσει μία σχέση που να εξηγεί το επίπεδο εξάρτησης μεταξύ της κλάσης και του συνεχούς χαρακτηριστικού, πρότεινε μια μέθοδο που χρησιμοποιεί ένα μη παραμετρικό μέτρο. Με αυτό το μέτρο ο συγγραφέας, τελικά, πέτυχε τη βελτίωση του ταξινομητή. Επιπλέον, μέσω της εκτεταμένης πειραματικής έρευνας που διεξήχθη στο άρθρο [265], έχει ληφθεί μια σημαντική παρατήρηση σχετικά με τη διαδικασία διακριτικοποίησης και την επίδρασή της, τόσο στους αφελείς όσο και στους ημι-αφελείς Bayesian ταξινομητών (semi-naive Bayesian classifiers). Οι συγγραφείς του άρθρου έχουν δείξει ότι, παρά τη χρονοβόρα διαδικασία της

διακριτοποίησης σε σχέση με τον χρόνο εκπαίδευσης του ταξινομητή, η όλη διαδικασία συχνά ολοκληρώνεται ταχύτερα από την αρχική διαδικασία ταξινόμησης του Bayes ταξινομητή.

Μια μέθοδος που διεξάγει αυτόματα τη διαδικασία διακριτοποίησης, προτάθηκε πρόσφατα [23] και στηρίζεται σε δύο παράγοντες: (α) τη (dispersion) καθώς και (β) την (uncertainty) του εύρους των δεδομένων. Τα πειράματα που διεξάγονται από τους συγγραφείς του άρθρου, τα οποία βασίστηκαν σε δεδομένα του πραγματικού κόσμου, έχουν δείξει ότι, ο προτεινόμενος αλγόριθμος πλησιάζει στην απόδοση του μικρότερου αριθμού διακριτών διαστημάτων και ο χρόνος διακριτοποίησης που απαιτεί είναι μικρότερος από άλλους αλγορίθμους της ίδιας κατηγορίας. Επιπλέον, στο άρθρο [235], οι συγγραφείς του πρότειναν το (Class-Attribute Coherence Maximization (CACM) criterion) και τους αντίστοιχους αλγορίθμους διακριτοποίησης: (α) τον CACM αλγόριθμο και (β) τον αποδοτικό-CACM αλγόριθμο. Τα εκτεταμένα υπολογιστικά πειράματα που διεξήγαγαν, δείχνουν ότι η μεθόδός τους ξεπερνά τις επιδόσεις άλλων γνωστών ταξινομητών, όπως τον αλγόριθμο RBF-SVM και C4.5. Επιπλέον, η παραλλαγή που έδωσαν ήταν ταχύτερη σε σύγκριση με τη μέθοδο Fast-CAIM.

Επιπρόσθετα, στη μελέτη που διεξήχθη πρόσφατα [53] έχει προταθεί μια βελτιωμένη έκδοση του αλγορίθμου CAIM. Ο νέος αλγόριθμος, που ονομάστηκε ur-CAIM, επιτυγχάνει πιο ευέλικτη διακριτοποίηση, παρουσιάζει περισσότερα ποιοτικά υποδιαστήματα και έχει καλύτερα αποτελέσματα ταξινόμησης για μη ισορροπημένα δεδομένα (unbalanced data). Επιπλέον, έχει παρουσιαστεί μια πιο ολοκληρωμένη και λεπτομερέστερη έρευνα σχετικά με τις μεθόδους διακριτοποίησης [433], καθώς και μια ταξινόμηση και εμπειρική ανάλυση αυτών των μεθόδων [149]. Ο αναγνώστης μπορεί, επίσης, να ανατρέξει σε μια σύγχρονη επισκόπηση των τεχνικών διακριτοποίησης σε συνδυασμό με μια ταξινόμηση των μεθόδων στην εργασία [319].

Ο Πίνακας 4.3 παρουσιάζει μια σύντομη περιγραφή και πληροφορίες σχετικά με τις αναφορές που έλαβαν οι πιο γνωστές και ευρέως χρησιμοποιούμενες μέθοδοι, που σχετίζονται με τα θέματα κανονικοποίησης και διακριτοποίησης των δεδομένων.

Τίτλος	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Discretization: An embedding technique'	922	57.625	Σε αυτή τη μελέτη παρέχονται ευρέως χρησιμοποιούμενες μέθοδοι διακριτοποίησης, καθώς και η εφαρμογή τους κυρίως σε προβλήματα ταξινόμησης. Επιπλέον, οι συγγραφείς της κάνουν μια κατηγοριοποίηση αυτών των τεχνικών και πτυχές, όπως η ακρίβεια και η ταχύτητα [240].
'KHIOPS: A statistical discretization method of continues attributes'	142	10.14	Ο συγγραφέας της εργασίας πρότεινε ένα νέο αλγόριθμο που μπορεί να χειριστεί συνεχή χαρακτηριστικά. Αυτή η μέθοδος χρησιμοποιεί το chi-square στατιστικό τεστ και βελτιστοποιεί ένα ολικό κριτήριο [43].
'CAIM discretization algorithm'	471	33.64	Η προτεινόμενη μέθοδος έχει σχεδιαστεί για επιτηρούμενα δεδομένα. Οι συγγραφείς της εργασίας δημιούργησαν έναν αλγόριθμο που διαχωρίζει το σύνολο δεδομένων χρησιμοποιώντας διαστήματα. Το σύνολο δεδομένων είναι διακριτοποιημένο χωρίς τη βοήθεια του χρήστη [225].
'A discretization algorithm based on a heterogeneity criterion'	52	4.00	Προτείνεται μια νέα ευρετική προσέγγιση για να βρεθεί το βέλτιστο σχήμα διακριτοποίησης. Επιπλέον, οι συντάκτες της εργασίας πρότειναν ένα νέο κριτήριο βάσει του ορισμού της (heterogeneity) [243].
'Evaluating the performance of cost-based discretization versus entropy-and error-based discretization'	56	4.67	Οι συγγραφείς του άρθρου συνέκριναν δύο πολύ γνωστές μεθόδους διακριτοποίησης: τη μέθοδο που βασίζεται στο κόστος και την εντροπία - και αυτή που βασίζεται στο «καθαρό» σφάλμα [200].
'Hand-geometry recognition using entropy-based discretization'	78	7.10	Σε αυτή τη μέθοδο το αρχικό σύνολο δεδομένων χωρίζεται σε μικρότερα διαστήματα με βάση την εντροπία [224].
'A discretization algorithm based on class-attribute contingency coefficient'	192	19.2	Οι συγγραφείς πρότειναν έναν αλγόριθμο, ο οποίος βασίζεται στον συντελεστή έκτακτης ανάγκης των χαρακτηριστικών κλάσης, προκειμένου να επιτευχθεί καλύτερη ακρίβεια ταξινόμησης [388].
'AMEVA: An autonomous discretization algorithm'	82	9.11	Αυτή η νέα προσέγγιση έχει συγκριθεί με το γνωστό αλγόριθμο CAIM [225] προκειμένου να εξετασθεί ποια παρέχει την καταλληλότερη διαίρεση διαστήματος [160].

Τίτλος	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Data discretization unification'	79	8.78	Οι συγγραφείς του άρθρου παρείχαν ένα δυναμικό αλγόριθμο, ο οποίος διαχειρίζεται την καλύτερη διακριτοποίηση με βάση την εντροπία. Κατέληξαν στο συμπέρασμα ότι περισσότερο διακριτοποιημένα διαστήματα συμβάλουν τις περισσότερες φορές σε λιγότερα σφάλματα ταξινόμησης [203].
'Discretization for naive-bayes learning: managing discretization bias and variance'	174	19.33	Παρέχεται ένας συνδυασμός αναλογικής διακριτοποίησης και καθορισμένης συχνότητας, προκειμένου να βελτιωθεί η ταξινόμηση των αφελών Bayes ταξινομητών [433].
'Discretization methods'	80	8.88	Παρουσιάζεται μια ερευνητική μελέτη που περιγράφει βασικές μεθόδους και τις κύριες πτυχές της διαδικασίας διακριτοποίησης [433].
'A clustering-based discretization for supervised learning'	41	5.125	Οι συγγραφείς του άρθρου, μέσω της χρήσης του αλγόριθμου συσταδοποίησης K means, παρείχαν ένα σύστημα διακριτοποίησης βασισμένο στις συστάδες, που χρησιμοποιεί το κριτήριο της εντροπίας και του μήκους περιγραφής (description length) [170].
'A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning'	275	54.6	Σε αυτή την έρευνα, παρέχονται οι πιο αντιπροσωπευτικές και σύγχρονες μέθοδοι διακριτοποίησης [149].
'ur-CAIM: improved CAIM discretization for unbalanced and balanced data'	27	13.5	Προτείνεται μια παραλλαγή του γνωστού αλγορίθμου CAIM. Τα πειράματα, που διενεργήθηκαν από τους συγγραφείς του άρθρου, έδειξαν ότι αυτή η νέα έκδοση επιτυγχάνει καλύτερα αποτελέσματα ταξινόμησης από τη βασική προσέγγιση του αλγορίθμου CIAM [53].
'Entropy-based discretization methods for ranking data'	25	12.5	Οι συγγραφείς της εργασίας πρότειναν ένα νέο εποπτευόμενο σχήμα διακριτοποίησης βασισμένο στην μ (minimum description length principle) [84].
'Data discretization: taxonomy and big data challenge'	49	24.5	Οι στόχοι αυτής της εργασίας είναι: Πρώτον, να επεξηγηθούν οι πλέον χρησιμοποιούμενες μέθοδοι διακριτοποίησης και δεύτερον, να επισημανθεί η δυνατότητα παραλληλοποίησης (parallelization) προκειμένου να εφαρμοστεί σε περιπτώσεις μεγάλων δεδομένων (Big Data) [319].

Τίτλος

ΣΠΑ ΑαΕ Συνοπτική περιγραφή
Πίνακας 4.3: Σύνομη περιγραφή και αριθμός των αναφορών που σχετίζονται με τα θέματα κανονικοποίησης και διακριτοποίησης που παρουσιάζονται στην Ενότητα 4.4. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.

4.5 Επιλογή παραδειγμάτων

Η προσέγγιση της μ (instance selection) δεν χρησιμοποιείται μόνο για την αντιμετώπιση του θορύβου [19, 20], αλλά και για την αντιμετώπιση της αδυναμίας μάθησης από τεράστια σύνολα δεδομένων [427]. Ο χρόνος εκπαίδευσης που απαιτείται από τα Τεχνητά Νευρωνικά Δίκτυα και τις Μηχανές Διανυσματικής Υποστήριξης, καθώς και ο χρόνος ταξινόμησης των μ (instance-based learners), επηρεάζεται, σαφώς, από το μέγεθος του συνόλου δεδομένων. Σε αυτή την περίπτωση, η μ (instance selection) μπορεί να θεωρηθεί ως ένα πρόβλημα βελτιστοποίησης, που επιχειρεί να διατηρήσει την ποιότητα μαζί με την ελαχιστοποίηση του μεγέθους του συνόλου δεδομένων [241]. Επιπλέον, αυξάνοντας τον αριθμό των παραδειγμάτων, η πολυπλοκότητα του παραγόμενου μοντέλου αυξάνεται με αποτέλεσμα τη μείωση της επεξηγηματικότητας των αποτελεσμάτων. Επομένως, η επιλογή παραδειγμάτων συνιστάται ιδιαίτερα στην περίπτωση μεγάλων συνόλων δεδομένων, όπως σημειώνεται στο άρθρο [152].

Η μ (sampling) είναι μια ευρέως χρησιμοποιούμενη διαδικασία, δεδομένου ότι αποτελεί μια ισχυρή υπολογιστική ενέργεια, που λειτουργεί σε ένα υπο-σύνολο (sub-sample) του συνόλου δεδομένων και είναι σε θέση να διατηρήσει ή και να αυξήσει την ακρίβεια [216]. Υπάρχει μια ποικιλία διαδικασιών για την δειγματοληψία παραδειγμάτων από ένα μεγάλο σύνολο δεδομένων (σύνολα δεδομένων που ξεπερνούν το ένα Gb). Οι πιο γνωστές είναι οι ακόλουθες [56]: (α) η μ (random sampling), που επιλέγει τυχαία ένα υποσύνολο παραδειγμάτων και (β) η μ (stratified sampling), στην περίπτωση που οι τιμές της κλάσης δεν είναι ομοιόμορφα κατανομημένες στο σύνολο δεδομένων.

Στο άρθρο [325], ο συγγραφέας του παρουσίασε ένα ενοποιημένο πλαίσιο, το οποίο καλύπτει προσεγγίσεις που σχετίζονται με την επιλογή παραδειγμάτων. Επιπλέον, στις εργασίες [214] και [147], παρέχεται επίσης μια ταξινόμηση και κατάταξη πρωτοτύπων σχημάτων μείωσης παραδειγμάτων. Γενικά, οι μέθοδοι επιλογής παραδειγμάτων έχουν κατηγοριοποιηθεί σε διαφορετικές κλάσεις, όπως αναγράφεται στην εργασία [278]:

- (α) (Type of selection): (i) Οι μ (condensation methods), οι οποίες προσπαθούν να υπολογίσουν ένα σταθερό υποσύνολο, αφαιρώντας περιπτώσεις που δεν μπορούν να αυξήσουν την ακρίβεια ταξινόμησης, (ii) οι μ (edition methods), οι οποίες προσπαθούν να αφαιρέσουν θορυβώδη στιγμιότυπα και (iii) οι μ (hybrid methods), που αναζητούν ένα υποσύνολο στο οποίο συναντώνται τόσο θορυβώδη όσο και ανωφελέ παραδείγματα.
- (β) (Direction of search): (i) Οι μ (incremental methods), που ξεκινούν με ένα κενό σύνολο και προσθέτουν παραδείγματα σύμφωνα με κάποιο κριτήριο, ενώ αντίθετα (ii) οι μ (decremental methods), που ξεκινούν με ολόκληρο το σύνολο δεδομένων και αφαιρούν παραδείγματα σύμφωνα με κάποια κριτήρια και (iii) οι μ (mixed methods), που ξεκινούν με ένα υποσύνολο και προσθέτουν ή αφαιρούν παραδείγματα, ανάλογα με την πλήρωση συγκεκριμένων κριτηρίων.
- (γ) (Evaluation of search): (i) Οι μ (wrapper methods), που θεωρούν ένα κριτήριο επιλογής με βάση την ακρίβεια που επιτυγχάνεται από έναν ταξινομητή και (ii) οι μ (filter methods), που χρησιμοποιούν μια συνάρτηση επιλογής, η οποία δε σχετίζεται με κάποιο συγκεκριμένο ταξινομητή.

Στο άρθρο [36], οι συγγραφείς του έχουν συγκρίνει έντεκα μεθόδους για την εύρεση πρωτοτύπων, στα οποία βασίζεται ο ταξινομητής των πλησιέστερων γειτόνων (nearest-neighbor

classifier). Σε αυτή τη σύγκριση, διάφορες μέθοδοι έχουν επιδείξει καλή απόδοση αλλά καμία από αυτές δεν είναι ανώτερη από όλες τις άλλες. Στην περίπτωση δυαδικών προβλημάτων (binary problems), έχει παρουσιαστεί μία σημαντική παρατήρηση [306]. Συγκεκριμένα, οι συναρτήσεις διάκρισης που βασίζονται στην ανομοιογένεια (dissimilarity-based discrimination functions), οι οποίες στηρίζονται σε σύνολα δεδομένων μειωμένων πρωτοτύπων (3-10% των παραδειγμάτων εκπαίδευσης), προσφέρουν παρόμοια ή πολύ καλύτερη ακρίβεια ταξινόμησης από τον αλγόριθμο k NN, που εφαρμόζεται σε ολόκληρο το σύνολο εκπαίδευσης. Επίσης, ο κανόνας επιλογής σχετίζεται με τον ταξινομητή k NN στους περισσότερους αλγόριθμους περιτυλίγματος (wrapper algorithms) [93].

Ένας πολύ γνωστός αλγόριθμος, για παράδειγμα, είναι ο αλγόριθμος HitMiss Network (HMN) που προτάθηκε στην εργασία [254]. Αυτός ο αλγόριθμος επικεντρώνεται στην απομάκρυνση του πλεονασμού (redundancy removal) και όχι στη μείωση του θορύβου. Συγκεκριμένα, ο προτεινόμενος αλγόριθμος χρησιμοποιήθηκε για να βελτιώσει την απόδοση του αλγορίθμου 1 NN. Ο συγγραφέας της εργασίας πρότεινε τρεις παραλλαγές του αρχικού αλγορίθμου. Καθεμία από αυτές παρείχε ξεχωριστά πλεονεκτήματα και βελτιώσεις σε σχέση με τον αλγόριθμο 1 NN. Για παράδειγμα, η παραλλαγή που ονομάζεται HMNC, διαγράφει παραδείγματα από το σύνολο δεδομένων, αλλά αυτό δεν επηρέασε την ακρίβεια του αλγορίθμου 1 NN. Ακόμα, η παραλλαγή η οποία ονομάστηκε HMN-EI και εκτελούσε επανειλημμένα την έκδοση HMN-E, ήταν η δεύτερη παραλλαγή που πρότεινε ο συγγραφέας, με την οποία κατάφερε τη μείωση του χώρου αποθήκευσης. Και οι τρεις παραλλαγές δοκιμάστηκαν εκτενώς σε πολλαπλά και διαφορετικού τύπου σύνολα δεδομένων. Τα αποτελέσματα έδειξαν ότι ο προτεινόμενος αλγόριθμος έχει καλύτερη ικανότητα γενίκευσης από άλλους αλγορίθμους επιλογής παραδειγμάτων. Επιπλέον, βελτιώνει σημαντικά την ακρίβεια του ταξινομητή 1 NN. Επιπροσθέτως, στο άρθρο [275], οι συγγραφείς πρότειναν τον μ (class boundary preserving algorithm). Αυτή η προσέγγιση χρησιμοποιεί την έννοια του μ (reachable set) και προτείνει μία επέκταση, η οποία περιλαμβάνει κάτι περισσότερο από τον πλησιέστερο γείτονα. Στη συνέχεια, με τη χρήση των πολλαπλών προσβάσιμων συνόλων και των πλησιέστερων γειτόνων δημιουργεί μία δομή γεωμετρικών μοτίβων προκειμένου να αφαιρεθούν τα περιττά παραδείγματα.

Στο άρθρο [83] έχει προταθεί μια μ μ (instance selection filtering strategy). Σε αυτή τη στρατηγική, η βασική ιδέα είναι να διαιρέσουμε το σύνολο δεδομένων σε μικρά αμοιβαίως αποκλειστικά μπλοκ (small mutually exclusive blocks) και στη συνέχεια, να εφαρμόσουμε ξεχωριστά έναν αλγόριθμο επιλογής παραδειγμάτων σε κάθε μπλοκ. Ακολουθώντας, οι περιπτώσεις που επιλέγονται από κάθε μπλοκ, συγχωνεύονται σε υποσύνολα, περίπου ίδιου μεγέθους, και ο αλγόριθμος επιλογής παραδειγμάτων εφαρμόζεται ξανά. Η διαδικασία επαναλαμβάνεται μέχρι να αρχίσει να αυξάνεται ένα σφάλμα επικύρωσης (validation error).

Στη συνέχεια αυτής της υποενοότητας, παρουσιάζουμε περισσότερες εργασίες, στις οποίες ο αναγνώστης μπορεί να βρει υλοποιήσεις μεθόδων και τεχνικών επιλογής παραδειγμάτων. Μέσω αυτών, η διεργασία επιλογής και κατ' επέκταση η αύξηση ακρίβειας της ταξινόμησης σε πληθώρα παραδειγμάτων γίνεται πιο εφικτή. Έτσι, στο άρθρο [79], παρουσιάστηκε μία προσέγγιση συσταδοποίησης για την επιλογή παραδειγμάτων. Αρχικά, αυτή η προσέγγιση δημιουργεί ομάδες παραδειγμάτων για κάθε κλάση μέσω συστάδων και στη συνέχεια, στο δεύτερο βήμα, χρησιμοποιείται μια διαδικασία περιτυλίγματος (wrapper procedure).

Επιπλέον, εξελικτικές τεχνικές έχουν χρησιμοποιηθεί για την προσέγγιση του προβλήματος επιλογής παραδειγμάτων, όπως ο αλγόριθμος που προτάθηκε στην εργασία [146]. Συγκεκριμένα, αυτή η προσέγγιση χρησιμοποιεί έναν μ μ (memetic algorithm), ο οποίος συνδυάζει εξελικτικούς αλγορίθμους και στρατηγικές τοπικής αναζήτησης. Λόγω αυτού του συνδυασμού, αποφεύγεται το έμφυτο πρόβλημα σύγκλισης που πλήττει τους εξελικτικούς αλγορίθμους. Στην πραγματικότητα, επιτυγχάνονται πολύ καλά αποτελέσματα, καθώς ο αριθμός των δεδομένων αυξάνεται. Ένας επιπλέον εξελικτικός αλγόριθμος

επιλογής παραδειγμάτων είναι ο μ (cooperative coevolutionary instance selection algorithm) που προτάθηκε στην εργασία [151]. Σε αυτή την προσέγγιση, χρησιμοποιείται η τεχνική μ και (divide and conquer). Έτσι, το αρχικό σύνολο δεδομένων υποδιαιρείται σε μικρότερα σύνολα και μία τεχνική ολικής βελτιστοποίησης, βασισμένη στους πληθυσμούς (population based), αναζητά τον καλύτερο συνδυασμό. Το πλεονέκτημα αυτής της μεθόδου, είναι ότι, ο χρήστης μπορεί να ελέγξει την αντικειμενική συνάρτηση (objective function), (όπως η ακρίβεια ή ο χώρος αποθήκευσης), μέσω της μ (fitness function). Επιπλέον, αυτή η προσέγγιση δεν απαιτεί τον υπολογισμό μετρικών απόστασης ή τη χρήση ενός συγκεκριμένου αλγορίθμου ταξινόμησης. Τα πειραματικά αποτελέσματα έχουν δείξει ότι ο προτεινόμενος αλγόριθμος δεν είναι κατώτερος από οποιονδήποτε άλλο από τους συγκρίσιμους αλγορίθμους. Ακόμα, αυτή η προσέγγιση είναι ανώτερη από άποψη απαιτήσεων και είναι πολλά υποσχόμενη στην περίπτωση μεγάλων συνόλων δεδομένων σχετικά με όρους υπολογιστικού κόστους και αποτελεσματικότητας.

Για τον χαρακτηρισμό του συνόλου δεδομένων έχουν προταθεί διάφορα μ (measures) [52]. Αυτά τα μέτρα χρησιμοποιούνται για την επιλογή, από ένα πλήθος προεπιλεγμένων μεθόδων, της μέθοδου (ή συνδυασμού μεθόδων) που αναμένεται να παράγει τα καλύτερα αποτελέσματα. Επιπλέον, μια μέθοδος ταξινόμησης παραδειγμάτων ανά κλάση, που χρησιμοποιεί μ (borders), έχει προταθεί στην εργασία [180]. Τα όρια των παραδειγμάτων είναι εκείνα που βρίσκονται κοντά στα όρια απόφασης μεταξύ των κλάσεων και είναι τα πιο χρήσιμα για τους αλγορίθμους μάθησης.

Στο άρθρο [236], οι συγγραφείς του εισήγαγαν την αντιπροσωπευτική προσέγγιση ανίχνευσης δεδομένων (representative data detection approach), η οποία βασίζεται στην ανάλυση και την πρόβλεψη των ακραίων προτύπων. Ένα μοντέλο μάθησης αρχικά εκπαιδεύεται για να μάθει τα πρότυπα των (μη) αντιπροσωπευτικών δεδομένων, που επιλέγονται από μια συγκεκριμένη μέθοδο επιλογής παραδειγμάτων, σύμφωνα με ένα μικρό ποσό δεδομένων εκπαίδευσης. Έτσι, μπορούν να χρησιμοποιηθούν τα πιο χρήσιμα παραδείγματα, ακόμα και από ένα τεράστιο σύνολο δεδομένων εκπαίδευσης.

Πρόσφατα [22] παρουσιάστηκαν δύο νέοι αλγόριθμοι γραμμικής πολυπλοκότητας για το σκοπό της καλύτερης επιλογής παραδειγμάτων. Και οι δύο αλγόριθμοι χρησιμοποιούν τον μ (locality-sensitive hashing) για να βρουν ομοιότητες μεταξύ των παραδειγμάτων. Επίσης, στο άρθρο [387], οι συγγραφείς του έχουν διερευνήσει την επίδραση της επιλογής παραδειγμάτων στο φιλτράρισμα ορισμένων θορυβωδών δεδομένων, αναφορικά με τη διεργασία καταλογισμού (imputation). Συγκεκριμένα, οι συγγραφείς αυτής της μελέτης εξέτασαν αν η διαδικασία επιλογής παραδειγμάτων πρέπει να προηγείται ή να ακολουθεί τη διαδικασία καταλογισμού. Για το σκοπό αυτό, έδωσαν τέσσερις τρόπους-συνδυασμούς για να βρουν ποιο είναι προτιμότερο. Αυτό έχει ως αποτέλεσμα την αναγνώριση-δημιουργία του καταλληλότερου συνόλου εκπαίδευσης για τον ταξινομητή τους. Επιπλέον, εκτέλεσαν εκτεταμένα υπολογιστικά πειράματα για να βρουν τον καλύτερο συνδυασμό σε μια ποικιλία συνόλων δεδομένων. Οι τέσσερις συνδυασμοί δοκιμάστηκαν για τους ταξινομητές K -NN και SVM.

Στην εργασία [237], οι συγγραφείς της έχουν μελετήσει το πρόβλημα μ (support vector recognition problem) κυρίως στο πλαίσιο των μεθόδων μείωσης (παραδειγμάτων), προκειμένου να ανασυσταθεί το σύνολο εκπαίδευσης. Οι συγγραφείς του άρθρου επικεντρώθηκαν στο γεγονός της άνισης κατανομής των παραδειγμάτων στον χώρο διανυσμάτων, για να προτείνουν έναν αποτελεσματικό αλγόριθμο αυτόματης επιλογής παραδειγμάτων, αναφορικά με μεθόδους που βασίζονται στη γεωμετρία.

Πρόσφατα, παρουσιάστηκε μια εργασία [126], όπου οι συγγραφείς της χρησιμοποίησαν έναν μ (multi-objective evolutionary algorithm), μέσω του οποίου κατάφεραν να προσεγγίσουν ένα βέλτιστο σύνολο τόσο χαρακτηριστικών όσο και παραδειγμάτων. Συγκεκριμένα, η προτεινόμενη μεθοδολογία εφαρμόστηκε σε μη-

ισοροπημένα σύνολα δεδομένων. Χρησιμοποιώντας το γνωστό ταξινομητή C4.5 σε συνδυασμό με την πολύ γνωστή εξελικτική μέθοδο NSGA-II, ήταν σε θέση να επεκτείνουν τον χώρο αναζήτησης, ο οποίος επιτρέπει την κατασκευή μ μ (ensemble of classifiers). Έτσι, μέσω της επιλογής χαρακτηριστικών ξεπερνάται το πρόβλημα επικάλυψης (overlapping), που είναι εγγενές στα πολλαπλών κλάσεων μη-ισοροπημένα σύνολα δεδομένων, καθώς και μέσω της επιλογής παραδειγμάτων, επιλύθηκαν δύο σημαντικά προβλήματα, αυτό της (elimination of noise) και του ζητήματος της (imbalance issue). Οι συγγραφείς διεξήγαγαν υπολογιστικά πειράματα, συμπεριλαμβανομένων συγκρίσεων με ευρέως χρησιμοποιούμενους και γνωστούς ταξινομητές σε δυαδικά και πολλαπλών κλάσεων μη-ισοροπημένα σύνολα δεδομένων. Τα πειραματικά αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος EFIS-MOEA παρουσιάζει καλύτερη απόδοση από τις συγκρινόμενες και είναι υποσχόμενη από την άποψη της δυνατότητάς της να εφαρμοστεί σε μεγάλα σύνολα δεδομένων.

Σε αυτό το σημείο θα θέλαμε να κάνουμε μια σύντομη αναφορά στην προσέγγιση της (prototype generation), η οποία δημιουργεί νέα τεχνητά πρωτότυπα για να αυξήσει την ακρίβεια. Στο άρθρο [386], οι συγγραφείς του διεξήγαγαν μια έρευνα για τις μεθόδους παραγωγής πρωτοτύπων, που έχουν σχεδιαστεί κυρίως για τον αλγόριθμο των πλησιέστερων γειτόνων. Επιπλέον, πραγματοποίησαν μια πειραματική μελέτη προκειμένου να μετρήσουν την απόδοση, όσον αφορά την ακρίβεια και τις δυνατότητες μείωσης των παραδειγμάτων.

Μια ενδιαφέρουσα προσέγγιση, στην οποία αξιοποιείται ένα σύνολο ταξινομητών μέσω της χρήσης γενιάς παραδειγμάτων (instance generation), παρουσιάστηκε στην εργασία [270]. Η φάση της εκπαίδευσης συνίσταται με την επανάληψη N φορές της παραγωγής πρωτοτύπων και στη συνέχεια, οι βαθμολογίες που προκύπτουν από την ταξινόμηση ενός δοκιμαστικού παραδείγματος, χρησιμοποιώντας κάθε σύνολο πρωτοτύπων, συνδυάζονται με μια τελική ψηφοφορία.

Σε γενικές γραμμές, οι ταξινομητές αναμένεται να είναι σε θέση να γενικεύσουν πάνω σε άγνωστα παραδείγματα οποιασδήποτε κλάσης με την ίδια ακρίβεια, κάτι που αποτελεί μια ιδανική κατάσταση. Επίσης, σε πολλές εφαρμογές, οι μ (learners) αντιμετωπίζουν μη-ισοροπημένα σύνολα δεδομένων, τα οποία μπορούν να προκαλέσουν μια συνθήκη μεροληψίας του μαθητή αναφορικά με μια κλάση. Αυτή η μεροληψία (bias) είναι το αποτέλεσμα μιας κλάσης που υποεκπροσωπείται σε μεγάλο βαθμό στα δεδομένα εκπαίδευσης σε σύγκριση με τις άλλες κλάσεις. Αυτό το γεγονός σχετίζεται με τον τρόπο με τον οποίο σχεδιάζονται οι εκπαιδευόμενοι (learners). Οι μ (inductive learners) σχεδιάζονται, συνήθως, για να ελαχιστοποιήσουν τα λάθη αναφορικά με τα παραδείγματα εκπαίδευσης. Οι κλάσεις που περιέχουν λίγα παραδείγματα μπορούν να αγνοηθούν ως επί το πλείστον από τους αλγορίθμους μάθησης, καθώς το κόστος της καλής απόδοσης στην υπερ-εκπροσωπούμενη κλάση υπερβαίνει το κόστος της κακής απόδοσης στη μικρότερη κλάση [176].

Επισημαίνουμε ότι τα μη-ισοροπημένα σύνολα δεδομένων έχουν πρόσφατα λάβει προσοχή σε διάφορες εργασίες της Μηχανικής Μάθησης [230]. Υπάρχουν διάφορες διαδικασίες για να αντιμετωπιστεί το πρόβλημα της ανισοροπίας στα σύνολα δεδομένων, μεταξύ άλλων, αναφέρουμε τα ακόλουθα [377]:

- (i) Διπλότυπα παραδείγματα εκπαίδευσης της υπο-εκπροσωπούμενης κλάσης (under-represented class), που στην πραγματικότητα επαναλαμβάνουν τα παραδείγματα (over-sampling),
- (ii) Αφαιρούνται παραδείγματα εκπαίδευσης της υπερ-εκπροσωπούμενης κλάσης (over-represented class), πραγματοποιώντας, έτσι, μείωση μεγέθους (under-sampling),
- (iii) Συνδυασμός των δύο παραπάνω προσεγγίσεων με αύξηση της «μικρής» κλάσης και μείωση της «μεγάλης».

(iv) Προσεγγίσεις μάθησης που στηρίζονται σε συνεργατικές μεθόδους (ensembles).

Στο άρθρο [62], οι συγγραφείς του πρότειναν μια δημοφιλή μέθοδο στον τομέα της MM και ΕΔ, την μ (Synthetic Minority Over-sampling TEchnique (SMOTE)), όπου παράγονται συνθετικά (τεχνητά) δείγματα αντί για υπερ-δειγματοληψία (over-sampling). Ο αλγόριθμος SMOTE παράγει τον ίδιο αριθμό δειγμάτων συνθετικών δεδομένων για κάθε πρωτότυπο παράδειγμα μειοψηφίας (minority instance). Αυτό συμβαίνει χωρίς να ληφθούν υπόψη γειτονικά παραδείγματα, γεγονός που αυξάνει την εμφάνιση επικάλυψης μεταξύ των κλάσεων [407].

Στην εργασία [125], προτάθηκε μια μεθοδολογία που χρησιμοποιεί τις Μηχανές Διανυσματικής Υποστήριξης ως προεπεξεργαστή. Η προσέγγιση αυτή αντικαθιστά τις πραγματικές τιμές στόχου των δεδομένων εκπαίδευσης με αυτές που προβλέπονται από τον αλγόριθμο SVM. Στη συνέχεια, τα τροποποιημένα δεδομένα εκπαίδευσης χρησιμοποιούνται για την εκπαίδευση άλλων αλγορίθμων.

Στη μελέτη [393], παρουσιάστηκε εκτεταμένη πειραματική έρευνα με χρήση θορυβωδών και μη-ισορροπημένων δεδομένων. Τα αποτελέσματα των πειραμάτων καταδεικνύουν τις επιπτώσεις του θορύβου στη μάθηση από μη-ισορροπημένα δεδομένα. Συγκεκριμένα, ο θόρυβος που προκύπτει από την αλλοίωση των παραδειγμάτων της αληθινής κλάσης μειοψηφίας φαίνεται να είναι ο σοβαρότερος. Στο άρθρο [55], οι συγγραφείς του πρότειναν το συνδυασμό των αλγορίθμων διαστρωμάτωσης (stratification algorithm) και επιλογής παραδειγμάτων για την κλιμάκωση του συνόλου δεδομένων προς τα κάτω (μείωση δεδομένων). Αναλυτικότερα, χρησιμοποιήθηκαν δύο μοντέλα διαστρωμάτωσης για να αυξηθεί η παρουσία κλάσεων μειοψηφίας, πριν λάβει χώρα η διαδικασία επιλογής παραδειγμάτων.

Όπως έχουμε ήδη αναφέρει στην Ενότητα 4.1 και όπως σημειώνεται στη βιβλιογραφία [246], το πρόβλημα της διαχείρισης μ - μ μ (imbalanced data-sets) είναι ένα από τα σημαντικότερα και πιο δύσκολα προβλήματα στην ΕΔ. Το πρόβλημα ταξινόμησης είναι ένα πρόβλημα που αντιμετωπίζεται ευρέως και έχει πολλές διαφορετικές πτυχές και προεκτάσεις. Οι πιο κοινές προσεγγίσεις περιλαμβάνουν: τη δειγματοληψία δεδομένων ή την κατάλληλη τροποποίηση του αλγορίθμου, ώστε να είναι σε θέση να χειριστούν τα μη-ισορροπημένα σύνολα δεδομένων. Σύμφωνα με τους συγγραφείς του άρθρου [246], η λύση επέρχεται με τον κατάλληλο συνδυασμό των παραπάνω προσεγγίσεων. Συγκεκριμένα, παρατήρησαν ότι οι λύσεις που είναι ευαίσθητες ως προς το κόστος, αν συνδυαστούν με τη δειγματοληψία δεδομένων και την αλγοριθμική τροποποίηση, μπορούν να επιτύχουν καλύτερο κόστος εσφαλμένης ταξινόμησης στην κλάση μειοψηφίας. Επιπλέον, έχουν ελαχιστοποιήσει τα λάθη υψηλού κόστους (high-cost errors).

Υπολογιστικά πειράματα σε δεκαεπτά σύνολα πραγματικών δεδομένων, που χρησιμοποιούν οκτώ διαφορετικούς ταξινομητές, έχουν δείξει ότι η υπερ-δειγματοληψία (over-sampling) της κλάσης μειοψηφίας υπερβαίνει την υπο-δειγματοληψία (under-sampling) της κλάσης πλειοψηφίας, όταν τα σύνολα δεδομένων είναι έντονα μη-ισορροπημένα. Από την άλλη πλευρά, δεν υπάρχουν σημαντικές διαφορές για τις βάσεις δεδομένων με χαμηλά ποσοστά ανισορροπίας [150].

Στη συνέχεια, ο αναγνώστης μπορεί να βρει στον Πίνακα 4.4 μια σύντομη περιγραφή και πληροφορίες σχετικά με τις αναφορές που έλαβαν οι πιο γνωστές και ευρέως χρησιμοποιούμενες μέθοδοι, οι οποίες σχετίζονται με ζητήματα επιλογής παραδειγμάτων.

Τίτλος	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Nearest prototype classifier designs: An experimental study'	188	11.06	Η συγκεκριμένη εργασία δίνει μια λεπτομερή σύγκριση μεταξύ διαφόρων προσεγγίσεων για την εύρεση πρωτοτύπων [36].
'SMOTE: synthetic minority over-sampling technique'	6422	401.375	Στο συγκεκριμένο άρθρο παρουσιάζεται ο γνωστός αλγόριθμος SMOTE, όπου παράγονται τεχνητά δείγματα αντί για υπερ-δειγματοληψία (over-sampling) [62].
'On issues of instance selection'	196	12.25	Στην παρούσα εργασία εξετάζονται τα κύρια θέματα και οι εξελίξεις σχετικά με τις τεχνικές επιλογής παραδειγμάτων [241].
'A unifying view on instance selection'	119	7.44	Παρέχεται ένα πλαίσιο που περιλαμβάνει προσεγγίσεις επιλογής παραδειγμάτων, καθώς επίσης τη δειγματοληψία και συγκεκριμένα παραδείγματα επισημαίνοντας τα πλεονεκτήματα και τα μειονεκτήματά τους [325].
'Learning drifting concepts: Example selection vs example weighting'	454	32.49	Προτείνεται ένα ευφύες μοντέλο που βασίζεται στον SVM ταξινόμητή, προκειμένου να φιλτράρει και να προσαρμόσει μεταβαλλόμενες καταστάσεις [216].
'Prototype selection for dissimilarity-based classifiers'	326	27.17	Διεξάγονται εκτεταμένα υπολογιστικά πειράματα με αρκετές μετρικές και μη-μετρικές αναπαραστάσεις ανομοιότητας. Επίσης, παρέχονται πρωτότυπα σχήματα επιλογής [306].
'A memetic algorithm for evolutionary prototype selection: A scaling up approach'	161	16.10	Προτείνεται ένας αποτελεσματικός συνδυασμός εξελικτικών αλγορίθμων και τοπικών τεχνικών αναζήτησης για την αντιμετώπιση του προβλήματος επιλογής πρωτοτύπων [146].
'A survey on evolutionary instance selection and generation'	135	16.875	Αυτή η εργασία παρουσίασε το χρήσιμο ρόλο των εξελικτικών αλγορίθμων, όσον αφορά τη μείωση δεδομένων. Οι συγγραφείς της παρείχαν μια σειρά εφαρμογών Εξόρυξης Δεδομένων, όπου εφαρμόστηκαν οι εξελικτικοί αλγόριθμοι [94].
'A review of instance selection methods'	235	29.375	Πρόκειται για μια εργασία επισκόπησης, όπου οι συγγραφείς της συνοψίζουν τις πιο γνωστές μεθόδους επιλογής παραδειγμάτων [278].

Τίτλος	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'Prototype selection for nearest neighbor classification: Taxonomy and empirical study'	447	74.50	Οι συγγραφείς του άρθρου παρουσίασαν μια εκτενή ανασκόπηση, εξετάζοντας τις πιο χρησιμοποιούμενες μεθόδους επιλογής πρωτυπών, για να βοηθήσουν τον αλγόριθμο πλησιέστερων γειτόνων να ξεπεράσει ένα πλήθος εγγενών μειονεκτημάτων [147].
'Analysis of preprocessing vs cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics'.	157	26.17	Σε αυτή την προσέγγιση, η δειγματοληψία δεδομένων με γνωστούς ταξινομητές φαίνεται να προσφέρει λύση στη διαχείριση μη-ισορροπημένων δεδομένων [246].
'A taxonomy and experimental study on prototype generation for nearest neighbor classification'	195	32.50	Οι συγγραφείς παρουσίασαν μια έρευνα αναφορικά με τις μεθόδους παραγωγής πρωτοτύπων για τον ταξινομητή k NN [386].
'Instance selection of linear complexity for big data'	22	11.00	Σε αυτή την εργασία προτείνονται δύο νέοι αλγόριθμοι που εξετάζουν τα χαρακτηριστικά των παραδειγμάτων προκειμένου να βρουν ομοιότητες. Και οι δύο αλγόριθμοι χρησιμοποιούν έναν τοπικό-ευαίσθητο κατακερματισμό (locality-sensitive hashing) για να επιτευχθεί ο παραπάνω σκοπός [22].
'Combining instance selection for better missing value imputation'	14	7.00	Οι συγγραφείς του άρθρου εξέτασαν αν η διαδικασία επιλογής των παραδειγμάτων πρέπει να προχωρήσει ή να ακολουθήσει τη διαδικασία καταλογισμού [387].
'An efficient instance selection algorithm to reconstruct training set for support vector machine'	15	15.00	Σε αυτή την εργασία έχει μελετηθεί το πρόβλημα αναγνώρισης διανύσματος υποστήριξης σχετικά με τις μεθόδους μείωσης, προκειμένου να ανασταθεί το σύνολο δεδομένων εκπαίδευσης [237].

Πίνακας 4.4: Σύνομη περιγραφή και αριθμός αναφορών που σχετίζονται με τα θέματα επιλογής παραδειγμάτων, όπως παρουσιάζονται στην ενότητα 4.5. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.

4.6 Επιλογή χαρακτηριστικών

Είναι γνωστό και ευρέως αποδεκτό ότι αρκετοί μαθητές (learners), όπως ο αλγόριθμος k -NN, είναι πολύ ευαίσθητοι σε σχέση με (irrelevant features). Επιπλέον, η παρουσία άσχετων χαρακτηριστικών μπορεί να κάνουν την εκπαίδευση κάποιων αλγορίθμων, όπως οι Μηχανές Διανυσματικής Υποστήριξης και τα Τεχνητά Νευρωνικά Δίκτυα άκρως αναποτελεσματική και σε πολλές περιπτώσεις μη πρακτική. Επιπλέον, οι Bayesian ταξινομητές δεν αποδίδουν καλά στην περίπτωση πλεοναζόντων μεταβλητών (redundant variables). Προκειμένου να αντιμετωπιστούν τα θέματα αυτά, προτείνεται να χρησιμοποιηθεί η προσέγγιση της (feature selection). Η διαδικασία αυτή, στοχεύει στον εντοπισμό και κατάργηση όσο το δυνατόν περισσότερων άσχετων και περιττών χαρακτηριστικών. Αυτό μειώνει τη διαστασιμότητα των δεδομένων και διευκολύνει τους αλγορίθμους μάθησης, ώστε να λειτουργούν γρηγορότερα και αποτελεσματικότερα. Γενικά, τα χαρακτηριστικά μπορούν να διακριθούν ως εξής [193]:

- (α) (Relevant): Τα χαρακτηριστικά που έχουν επιρροή στην κλάση και ο ρόλος τους δεν μπορεί να θεωρηθεί από τα υπόλοιπα χαρακτηριστικά. Συνεπώς, πρόκειται για χαρακτηριστικά με σημαντικό ρόλο στο σύνολο δεδομένων.
- (β) (Irrelevant): Άσχετα χαρακτηριστικά που δεν επηρεάζουν καθόλου την κλάση.
- (γ) (Redundant): Ο πλεονασμός συναντάται όταν ένα χαρακτηριστικό μπορεί να πάρει το ρόλο ενός άλλου (ίσως ο απλούστερος τρόπος για να μοντελοποιηθεί ο πλεονασμός). Στην πράξη, δεν είναι τόσο απλό να προσδιοριστεί ο πλεονασμός χαρακτηριστικών, όταν ένα χαρακτηριστικό συσχετίζεται (ίσως εν μέρει) με ένα σύνολο χαρακτηριστικών. Συνεπώς, είναι η δυσκολότερη κατηγορία προς διαχείριση.

Γενικά, οι αλγόριθμοι επιλογής χαρακτηριστικών αποτελούνται από δύο προσεγγίσεις [193]:

- (i) Ένας μ (selection algorithm), που παράγει τα προτεινόμενα υποσύνολα χαρακτηριστικών, προκειμένου να βρει ένα βέλτιστο υποσύνολο και
- (ii) Ένας μ (evolutionary algorithm), ο οποίος αποφασίζει για την ποιότητα του προτεινόμενου υποσυνόλου χαρακτηριστικών, επιστρέφοντας κάποια μέτρα «καταλληλότητας» στον αλγόριθμο επιλογής.

Από την άλλη πλευρά, χωρίς την εφαρμογή ενός κατάλληλου κριτηρίου διακοπής, είναι δυνατόν η διαδικασία της επιλογής χαρακτηριστικών να εκτελεστεί εξαντλητικά ή χωρίς τερματισμό, στο χώρο των υποψήφιων υποσυνόλων. Συνήθως, τα κριτήρια διακοπής είναι:

- (i) Είτε η προσθήκη (ή η διαγραφή) οποιουδήποτε χαρακτηριστικού που δεν προσφέρει ένα καλύτερο υποσύνολο, και/ή
- (ii) Είτε εάν έχει βρεθεί ένα βέλτιστο υποσύνολο σύμφωνα με κάποια συνάρτηση αξιολόγησης.

Στην εργασία [304], διάφορες τεχνικές επιλογής χαρακτηριστικών έχουν συγκριθεί, καταλήγοντας στο συμπέρασμα ότι, καμία από αυτές δεν είναι ανώτερη από όλες τις υπόλοιπες. Όσον αφορά τη στρατηγική επιλογής, οι μέθοδοι φιλτραρίσματος ταξινομούν τα χαρακτηριστικά ανεξάρτητα από τον ταξινομητή. Στις μεθόδους φιλτραρίσματος, ένα χαρακτηριστικό μπορεί να επιλεγεί λαμβάνοντας υπόψη ορισμένα προκαθορισμένα κριτήρια όπως: (α) την μ (mutual information) [69], (β) το μ (class separability measure) [253], ή (γ) την μ (variable ranking) [57].

Από την άλλη πλευρά, οι μ μ (wrapper methods) χρησιμοποιούν έναν ταξινομητή για να αξιολογήσουν τα υποσύνολα χαρακτηριστικών. Αυτές οι μέθοδοι είναι υπολογιστικά δαπανηρές στην πράξη και δεν είναι ανεξάρτητες από τον τύπο ταξινομητή που χρησιμοποιείται, όπως σημειώνεται χαρακτηριστικά στην εργασία [242]. Οι μέθοδοι περιτυλίγματος, χρησιμοποιούν την προσέγγιση της μ (cross-validation) για να προβλέψουν τα οφέλη από την προσθήκη ή την αφαίρεση ενός χαρακτηριστικού από το υποσύνολο χαρακτηριστικών που χρησιμοποιείται. Στην περίπτωση της μ μ (forward stepwise selection), δημιουργείται επαναληπτικά ένα υποσύνολο χαρακτηριστικών. Κάθε ένα από τα χρησιμοποιηθέντα χαρακτηριστικά προστίθεται στο μοντέλο με τη σειρά του και επιλέγεται το χαρακτηριστικό που βελτιώνει περισσότερο την ακρίβεια του μοντέλου. Στην περίπτωση της μ (backward stepwise selection), ο αλγόριθμος ξεκινά με την οικοδόμηση ενός μοντέλου που περιλαμβάνει όλα τα διαθέσιμα χαρακτηριστικά εισόδου. Σε κάθε επανάληψη, ο αλγόριθμος εντοπίζει τη μεταβλητή της οποίας η απουσία προκαλεί τη μεγαλύτερη βελτίωση της απόδοσης (ή προκαλεί την ελάχιστη χειρότερηση του μοντέλου). Ένα πρόβλημα με την προς τα εμπρός επιλογή είναι ότι, μπορεί να αποτύχει να συμπεριλάβει χαρακτηριστικά που είναι αλληλεξαρτώμενα, καθώς προσθέτει μια μεταβλητή τη φορά. Από την άλλη πλευρά, είναι σε θέση να εντοπίσει μικρά αποτελεσματικά υποσύνολα πολύ σύντομα, αφού οι πρώτες αξιολογήσεις που περιλαμβάνουν σχετικά λίγα χαρακτηριστικά είναι γρήγορες. Αντίθετα, στην προς τα πίσω επιλογή, οι αλληλεπιδράσεις διαχειρίζονται καλώς, αλλά στην αρχή οι αξιολογήσεις είναι υπολογιστικά δαπανηρές.

Διάφορες μέθοδοι επιλογής χαρακτηριστικών, αντιμετωπίζουν την περίπτωση πολλαπλών κλάσεων απευθείας (αντί να την «αποδομούν» σε διάφορα προβλήματα δύο κλάσεων). Η διαδοχική προς τα εμπρός αιωρούμενη επιλογή (sequential forward floating selection) και η διαδοχική προς τα πίσω αιωρούμενη επιλογή (sequential backward floating selection), χαρακτηρίζεται από το μεταβαλλόμενο αριθμό χαρακτηριστικών που περιλαμβάνονται ή εξαλείφονται σε διαφορετικά στάδια της διαδικασίας [370]. Στο άρθρο [251], οι συγγραφείς έχουν εισαγάγει έναν αλγόριθμο περιτυλίγματος για την επιλογή χαρακτηριστικών χρησιμοποιώντας Μηχανές Διανυσματικής Υποστήριξης με συναρτήσεις πυρήνα (kernel functions). Η μέθοδός τους, βασίζεται σε μια διαδοχική προς τα πίσω επιλογή, χρησιμοποιώντας τον αριθμό των λανθασμένων ταξινομήσεων σε ένα υποσύνολο επικύρωσης (validation subset), ως το μέτρο για να αποφασίσει ποιο χαρακτηριστικό πρέπει να εξαλειφθεί σε κάθε βήμα.

Επιπλέον, ο μ (genetic algorithm), είναι μια ακόμα γνωστή προσέγγιση για την αντιμετώπιση θεμάτων επιλογής χαρακτηριστικών [357]. Σε αυτή την προσέγγιση, για κάθε επανάληψη, επιλέγεται ένα χαρακτηριστικό και αποφασίζεται αν θα συμπεριληφθεί στο υποσύνολο ή θα αποκλειστεί από αυτό. Όλοι οι συνδυασμοί άγνωστων χαρακτηριστικών χρησιμοποιούνται με ίσες πιθανότητες. Λόγω της πιθανοτικής φύσης της διαδικασίας αναζήτησης, ένα χαρακτηριστικό που θα πρέπει να είναι στο υποσύνολο θα είναι ανώτερο των άλλων, ακόμα και αν εξαρτάται από ένα άλλο χαρακτηριστικό. Ένα σημαντικό χαρακτηριστικό του γενετικού αλγορίθμου είναι ότι έχει σχεδιαστεί για να αξιοποιεί την (epistasis), (δηλαδή την αλληλεξάρτηση μεταξύ των δυαδικών ψηφίων στη συμβολοσειρά), και έτσι, είναι κατάλληλος για την επιλογή χαρακτηριστικών. Από την άλλη πλευρά, οι γενετικοί αλγόριθμοι απαιτούν, συνήθως, μεγάλο αριθμό αξιολογήσεων προκειμένου να φτάσουν σε έναν ελαχιστοποιητή.

Προκειμένου να συνδυαστούν τα πλεονεκτήματα των μοντέλων φιλτραρίσματος και περιτυλίγματος, έχουν προταθεί διάφορα υβριδικά μοντέλα (hybrid models) για την αντιμετώπιση δεδομένων μεγάλης διαστάσεως. Στην εργασία [391], οι συγγραφείς της παρουσίασαν έναν υβριδικό αλγόριθμο φίλτρου-περιτυλίγματος για την επιλογή χαρακτηριστικών (hybrid filter-wrapper feature subset selection), που βασίζεται στη μ μ μ (particle swarm optimization) [211, 296] για την ταξινόμηση των SVM. Το μοντέλο φίλτρου βασίζεται στις αμοιβαίες πληροφορίες και είναι ένα συνδυασμένο μέτρο της

συνάφειας (relevance) και του πλεονασμού (redundancy) χαρακτηριστικών αναφορικά με το επιλεγμένο υποσύνολο χαρακτηριστικών. Επίσης, το μοντέλο περιτυλίγματος μπορεί να θεωρηθεί ως τροποποιημένος διακριτός αλγόριθμος βελτιστοποίησης με χρήση σμήνους ωματιδίων.

Οι συγγραφείς του άρθρου [190], πρότειναν μια έννοια (neighborhood margin) και μ (neighborhood soft margin), προκειμένου να υπολογιστεί η ελάχιστη απόσταση μεταξύ διαφορετικών κλάσεων. Συγκεκριμένα, έχουν χρησιμοποιήσει το κριτήριο του μαλακού περιθωρίου γειτονιάς, για να εκτιμήσουν την ποιότητα των υποψηφίων χαρακτηριστικών και να δημιουργήσουν έναν άπληστο αλγόριθμο επιλογής χαρακτηριστικών προς τα εμπρός.

Ένα στοιχείο στο οποίο αξίζει να σταθούμε είναι το ακόλουθο. Αντί να προσεγγίσουμε χωριστά την επιλογή παραδειγμάτων ή τα προβλήματα της επιλογής χαρακτηριστικών, σε διάφορες ερευνητικές προσπάθειες, μελετάται η ταυτόχρονη επιλογή παραδειγμάτων και χαρακτηριστικών [93]. Για να συμβεί αυτό, οι συντάκτες του άρθρου παρείχαν ένα εξελικτικό σχήμα, που εφαρμόστηκε στο πρόβλημα ταξινόμησης του αλγορίθμου k NN. Τα πειραματικά αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος υπερέρχει έναντι άλλων γνωστών εξελικτικών προσεγγίσεων, όπως οι αλγόριθμοι FS-GGA, FS-SSGA, FS-CHC και 1, όσον αφορά το πρόβλημα μείωσης των δεδομένων.

Ακόμα, στο άρθρο [350], οι συγγραφείς του, με βάση τη θεωρία των ακατέργαστων συνόλων (rough set theory), έχουν αντιμετωπίσει το πρόβλημα της επιλογής χαρακτηριστικών για ευαίσθητα σε κόστος δεδομένα με ελλίπουσες τιμές. Έτσι, πρότειναν μια πολυαντικειμενική συνάρτηση αξιολόγησης, προκειμένου να προσδιοριστεί η σημασία των υποψηφίων χαρακτηριστικών. Συγκεκριμένα, αυτό συμβαίνει λαμβάνοντας υπόψη όχι μόνο την ισχύ της θετικής περιοχής (positive region) και της συνοριακής περιοχής (boundary region), αλλά και το σχετικό τους κόστος (associated costs).

Γενικά, στις προσεγγίσεις επιλογής χαρακτηριστικών, ένα χαρακτηριστικό παίρνει ένα δυαδικό βάρος, όπου το '1' υποδεικνύει ότι το χαρακτηριστικό έχει επιλεγεί, ενώ το '0' το αντίθετο. Ωστόσο, η στάθμιση χαρακτηριστικών (feature weighting) αντιστοιχεί μια τιμή, συνήθως από το διάστημα $[0; 1]$ σε κάθε χαρακτηριστικό, και όσο μεγαλύτερη είναι αυτή η τιμή, τόσο πιο σημαντικό θα είναι το χαρακτηριστικό. Η χρήση των μοντέλων που στηρίζονται στο βάρος (weight-based models), έχει καθιερωθεί ευρέως σε διάφορα ζητήματα ταξινόμησης [411]. Για αυτό τον σκοπό, τα σταθμισμένα νευρωνικά δίκτυα (weighted neural networks), μπορούν να θεωρηθούν ως ένα από τα πιο χρησιμοποιούμενα παραδείγματα στο χώρο [290]. Από την άλλη, οι μέθοδοι των Μηχανών Διανυσματικής Υποστήριξης [349] και των πλησιέστερων γειτόνων [258] μπορούν να χρησιμοποιήσουν βάρη για να επιτύχουν καλύτερη απόδοση. Επιπρόσθετα, είναι γνωστό ότι, η σωστή απόδοση των βαρών κατά τη διάρκεια της διαδικασίας εκπαίδευσης, βελτιώνει την ακρίβεια ταξινόμησης του μοντέλου. Στο άρθρο [385], οι συγγραφείς του έχουν ενσωματώσει ένα νέο σχήμα στάθμισης χαρακτηριστικών, χρησιμοποιώντας δύο διαφορετικές μεθόδολογίες παραγωγής πρωτοτύπων. Στην ουσία, οι συγγραφείς του, με το προτεινόμενο υβριδικό εξελικτικό σχήμα βελτίωσαν την απόδοση άλλων εξελικτικών προσεγγίσεων, που έχουν χρησιμοποιηθεί για τη βελτιστοποίηση των θέσεων των πρωτοτύπων.

Είναι ευρέως γνωστό ότι τα μ χρησιμοποιούν τη διαδικασία μ και μ . Επομένως, τείνουν να αποδίδουν καλά αν υπάρχουν λίγα και πολύ σημαντικά χαρακτηριστικά, αλλά στην περίπτωση που υπάρχουν πολλές πολύπλοκες αλληλεπιδράσεις μεταξύ τους. Γενικά, το πρόβλημα της αλληλεπίδρασης χαρακτηριστικών μπορεί να αντιμετωπιστεί με την κατασκευή νέων από το βασικό σύνολο χαρακτηριστικών. Αυτή η διαδικασία ονομάζεται μ / μ μ (feature construction/transformation). Σε αυτή την περίπτωση, τα νέα χαρακτηριστικά που δημιουργούνται ενδέχεται να οδηγήσουν στη δημιουργία πιο μεστών και ακριβών ταξινομητών. Επιπλέον, η ανακάλυψη σημαντικών χαρακτηριστικών συμβάλλει στην καλύτερη κατανόηση του παρα-

γόμενου ταξινομητή και στην καλύτερη κατανόηση της μαθησιακής διαδικασίας.

Οι συγγραφείς της εργασίας [357], έχουν χρησιμοποιήσει τις προσεγγίσεις του γενετικού προγραμματισμού και του γενετικού αλγορίθμου, ως εργαλεία προεπεξεργασίας δεδομένων με εφαρμογή τα δέντρα απόφασης. Συγκεκριμένα, για αυτή την περίπτωση χρησιμοποίησαν τον αλγόριθμο C4.5, που χρησιμοποιείται για τη δημιουργία ενός δέντρου απόφασης. Αναλυτικότερα, χρησιμοποίησαν γενετικό προγραμματισμό προκειμένου να κατασκευάσουν χαρακτηριστικά και τον γενετικό αλγόριθμο ως εργαλείο επιλογής χαρακτηριστικών. Ακόμα, έχουν αποδείξει, χρησιμοποιώντας δέκα σύνολα δεδομένων, ότι το πλαίσιο που όρισαν επιτυγχάνει καλύτερα αποτελέσματα συγκριτικά με τη μεμονωμένη χρήση δέντρων απόφασης. Επιπρόσθετα, έχουν αποδείξει ότι η αποτελεσματική προεπεξεργασία των δεδομένων εισόδου στα δέντρα απόφασης, μπορεί να βελτιώσει σημαντικά την απόδοση της ταξινόμησης.

Συμπληρωματικά, η είναι ένα είδος αναλυτικής διαδικασίας, που βασίζεται σε έναν υπόχωρο και είναι σε θέση να εκτιμήσει ένα αρχικό δείγμα με τη χρήση χαρακτηριστικών διανυσμάτων χαμηλών διαστάσεων [186]. Ακόμα, ένας μεγάλος χώρος χαρακτηριστικών μπορεί να μετατραπεί σε ένα χώρο χαμηλής διαστασιμότητας, εκπαιδεύοντας ένα μ (multilayer neural network) με λίγα κρυμμένα στρώματα. Έτσι, οι γνωστές και ευρέως χρησιμοποιούμενες μέθοδοι βελτιστοποίησης μ (gradient descent) μπορούν, στη συνέχεια, να χρησιμοποιηθούν για την εξομάλυνση (fine-tuning) των βαρών σε τέτοια δίκτυα αυτόματης κωδικοποίησης (autoencoder networks) [182]. Επομένως, ενώ ο αλγόριθμος PCA περιορίζεται σε ένα γραμμικό χάρτη, οι autoencoders δεν το κάνουν.

Κλείνοντας αυτή την ενότητα, αξίζει να αναφέρουμε ορισμένες ακόμα χρήσιμες πληροφορίες. Έχει αποδειχθεί ότι η κατασκευή χαρακτηριστικών (feature construction) μειώνει την πολυπλοκότητα του χώρου, που καλύπτεται από τα δεδομένα εισόδου. Στο άρθρο [305], οι συγγραφείς του παρουσίασαν έναν επαναληπτικό αλγόριθμο για τη βελτίωση της απόδοσης οποιασδήποτε επαγωγικής διαδικασίας εκμάθησης, χρησιμοποιώντας την κατασκευή χαρακτηριστικών ως ένα βήμα προεπεξεργασίας. Η επιλογή, μεταξύ των διαδικασιών επιλογής χαρακτηριστικών και κατασκευής χαρακτηριστικών, εξαρτάται από τον τομέα εφαρμογής και το διαθέσιμο σύνολο εκπαίδευσης. Σε γενικές γραμμές, η επιλογή χαρακτηριστικών οδηγεί σε εξοικονόμηση του κόστους μετρήσεων. Αυτό συμβαίνει γιατί ορισμένα από τα χαρακτηριστικά απορρίπτονται, και τα επιλεγμένα χαρακτηριστικά διατηρούν την αρχική τους φυσική ερμηνεία. Επιπλέον, τα διατηρούμενα χαρακτηριστικά μπορεί να είναι σημαντικά για την κατανόηση της φυσικής διαδικασίας που παράγει τα μοτίβα. Αντιθέτως, τα μετασχηματισμένα χαρακτηριστικά, που παράγονται από την κατασκευή χαρακτηριστικών, είναι ικανά να παρέχουν μια καλύτερη διακριτική ικανότητα σε σχέση με το καλύτερο υποσύνολο των διαθέσιμων χαρακτηριστικών. Από την άλλη πλευρά, αυτά τα νέα χαρακτηριστικά μπορεί να μην έχουν σαφή φυσικό νόημα.

Ο Πίνακας 4.5, παρουσιάζει μια σύντομη περιγραφή, καθώς και τις σχετικές αναφορές που έχουν λάβει οι πιο γνωστές και ευρέως χρησιμοποιούμενες μέθοδοι για την επιλογή χαρακτηριστικών.

Τίτλος	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms'	844	40.19	Οι αξιολογήσεις που έγιναν από τους συγγραφείς του άρθρου, έδειξαν ότι οι μέθοδοι που χρησιμοποιούν τη στάθμιση χαρακτηριστικών έχουν καλύτερη απόδοση, χαμηλότερες απαιτήσεις προεπεξεργασίας και χρειάζονται λιγότερα δεδομένα εκπαίδευσης [411].
'Orthogonal forward selection and backward elimination algorithms for feature subset selection'	134	9.57	Προτείνονται δύο αλγόριθμοι επιλογής υποομάδων ορθογώνιων χαρακτηριστικών. Ο πρώτος βασίζεται στην ορθογώνια προς τα εμπρός επιλογή, ενώ ο δεύτερος βασίζεται στην προς τα πίσω εξάλειψη παραδειγμάτων [253].
'Evaluating feature selection methods for learning in data mining applications'	216	15.43	Μια σειρά από μεθόδους επιλογής χαρακτηριστικών συγκρίνονται, καταλήγοντας στο συμπέρασμα ότι καμία δεν ξεπερνά σε επιδόσεις όλες τις άλλες [304].
'Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information'	158	12.15	Οι συντάκτες του άρθρου παρείχαν μια αποτελεσματική προσέγγιση επιλογής χαρακτηριστικών, προκειμένου να υπολογίσουν αποτελεσματικά τις στατιστικές πληροφορίες υψηλής τάξης [69].
'Optimal number of features as a function of sample size for various classification rules'	186	14.31	Σε αυτή τη μελέτη, λαμβάνει χώρα ο παράλληλος υπολογισμός του υποσυνόλου των ωφέλιμων χαρακτηριστικών. Για να επιτευχθεί αυτό, οι συγγραφείς εφάρμοσαν αρκετούς ταξινομητές [193].
'Genetic programming with a genetic algorithm for feature construction and selection'	101	7.77	Οι συγγραφείς της εργασίας, συνδύασαν το γενετικό προγραμματισμό με γενετικούς αλγόριθμους, ως εργαλεία προεπεξεργασίας για τα δέντρα απόφασης. Συγκεκριμένα, χρησιμοποίησαν τον γενετικό προγραμματισμό, ως κατασκευαστή χαρακτηριστικών και έναν γενετικό αλγόριθμο ως επιλογή χαρακτηριστικών [357].
'Reducing the dimensionality of data with neural networks'	7835	652.92	Οι συγγραφείς της συγκεκριμένης εργασίας, «ρίχνουν» τη διαστασιότητα του συνόλου δεδομένων χρησιμοποιώντας δίκτυα MLP [182].
'Kernel PCA for novelty detection'	424	38.54	Ο συγγραφέας του άρθρου μέσω της προσέγγισης πυρήνα-PCA πέτυχε να μειώσει ένα μεγάλο χώρο χαρακτηριστικών σε ένα αντίστοιχο χαμηλότερης διάστασης [186].

Τίτλος	ΣΠΑ	ΑαΕ	Συνοπτική περιγραφή
'A wrapper method for feature selection using support vector machines'	223	24.78	Σε αυτή τη μελέτη, η διαδοχική διαδικασία προς τα πίσω επιλογής χρησιμοποιείται για να εξαλείψει τα άχρηστα χαρακτηριστικά βήμα προς βήμα [251].
'm ² PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification'	168	24.00	Σε αυτή την προσέγγιση, ο γνωστός αλγόριθμος PSO χρησιμοποιείται ως στάδιο προεπεξεργασίας για τον ταξινομητή SVM [391].
'Multi-criteria feature selection on cost-sensitive data with missing values'	8	4.00	Οι συγγραφείς της συγκεκριμένης μελέτης, έδωσαν ένα σχήμα για να προσδιορίσουν τη σημασία των υποψηφίων χαρακτηριστικών, λαμβάνοντας υπόψη το σχετικό τους κόστος [350].

Πίνακας 4.5: Σύντομη περιγραφή και αριθμός αναφορών, που σχετίζονται με τα θέματα επιλογής χαρακτηριστικών όπως παρουσιάζονται στην ενότητα 4.6. Όπου «ΣΠΑ» δηλώνει το Συνολικό Πλήθος Αναφορών, ενώ «ΑαΕ» δηλώνει τον αριθμό Αναφορών ανά Έτος.

4.7 Σύνοψη και παρατηρήσεις

Οι αλγόριθμοι Προβλεπτικής Εξόρυξης Δεδομένων (predictive data mining), μάς επιτρέπουν να ανακαλύπτουμε γνώση από μεγάλα σύνολα δεδομένων, τα οποία υπόκεινται σε κατάλληλες συνθήκες. Το πιο σημαντικό ζήτημα είναι το αρχικό σύνολο δεδομένων να είναι αξιόπιστο. Εάν τα δεδομένα που λαμβάνονται ως είσοδος από τους αλγορίθμους εξόρυξης δεδομένων είναι αναξιόπιστα ή μολύνονται από θόρυβο, τότε δεν μπορούν να παράσχουν καλά και ανταγωνιστικά αποτελέσματα. Με άλλα λόγια, τα σύνολα δεδομένων πρέπει να είναι υψηλής ποιότητας, προκειμένου να δομηθούν αξιόπιστα μοντέλα που θα παρέχουν ακριβή αποτελέσματα. Για τους λόγους αυτούς, η προεπεξεργασία δεδομένων είναι απαραίτητη για την αξιοποίηση αλγορίθμων «ευφους» εξόρυξης δεδομένων σε διαδικασίες ανίχνευσης γνώσης (knowledge discovery).

Δεδομένου ότι η ποιότητα του συνόλου δεδομένων και η απόδοση των αλγορίθμων εξόρυξης δεδομένων σχετίζονται άμεσα, έχουν διεξαχθεί σημαντικές προσπάθειες από διάφορους ερευνητές προκειμένου να καταγραφούν ορισμένα ζητήματα που αφορούν τη μεταξύ τους αλληλεπίδραση. Ως εκ τούτου, μέσω της πειραματικής έρευνας που διεξήχθη στο άρθρο [76], έχει αποδειχθεί ότι η αναπαράσταση του συνόλου δεδομένων επηρεάζει αρκετά τις μεθόδους μάθησης, και ότι η απόδοση του αλγορίθμου είναι καλύτερη, εάν έχει πραγματοποιηθεί αποτελεσματική προεπεξεργασία δεδομένων. Επιπλέον, η επίδραση της προεπεξεργασίας ποικίλλει ανάλογα με τον αλγόριθμο. Έτσι, στο παρόν κεφάλαιο παρουσιάστηκαν σύγχρονοι, γνωστοί και ευρέως χρησιμοποιούμενοι αλγόριθμοι για κάθε βήμα προεπεξεργασίας δεδομένων.

Συνοπτικά, στην περίπτωση που ένα σύνολο δεδομένων είναι τεράστιο, ενδέχεται να μην είναι εφικτό να εφαρμοστεί κάποιος αλγόριθμος εξόρυξης δεδομένων. Για το σκοπό αυτό, είναι απαραίτητο να μειωθεί η ποσότητα των διαθέσιμων δεδομένων, η οποία μπορεί να γίνει μέσω της σωστής επιλογής των πραγματικά χρήσιμων δεδομένων. Αυτή η εργασία, μπορεί να επιτευχθεί μέσω της επιλογής παραδειγμάτων. Έτσι, μέσω της κατάλληλης επιλογής παραδειγμάτων, αποφεύγονται άσχετα δεδομένα ή δεδομένα που περιέχουν θόρυβο και ως εκ τούτου, το δείγμα γίνεται υψηλότερης ποιότητας. Έτσι, θα μπορεί να χρησιμοποιηθεί από αλγόριθμο Εξόρυξης Δεδομένων ή ευρύτερα Μηχανικής Μάθησης.

Στις περισσότερες περιπτώσεις, τα δεδομένα που λείπουν πρέπει να υποστούν κάποια προεπεξεργασία, ούτως ώστε να γίνει εφικτή η επεξεργασία ολόκληρου του συνόλου δεδομένων από έναν αλγόριθμο Εξόρυξης Δεδομένων. Στην περίπτωση που τα χαρακτηριστικά του συνόλου δεδομένων είναι συνεχή, οι συμβολικοί αλγόριθμοι, όπως τα δέντρα απόφασης (decision trees) ή οι μαθητές των κανόνων (rule learners), μπορούν να ενσωματωθούν με έναν αλγόριθμο διακριτοποίησης, που μετατρέπει τα αριθμητικά χαρακτηριστικά σε διακεκριμένα. Έτσι, το σύνολο δεδομένων γίνεται διαχειρίσιμο και έτοιμο για επεξεργασία και εξαγωγή συμπερασμάτων.

Ένα πολύ σημαντικό αποτέλεσμα, που προέκυψε από τη σύγκριση μεταξύ των διαφόρων μεθόδων επιλογής χαρακτηριστικών, παρουσιάστηκε στην εργασία [192]. Αυτό δείχνει ότι καμία από τις μεθόδους που εξετάστηκαν και έλαβαν μέρος σε αυτή τη σύγκριση δεν παρουσιάζει καλύτερα αποτελέσματα έναντι όλων των πλαισίων που τέθηκαν. Παρά το γεγονός αυτό, παρατηρήθηκαν κάποιες συσχετίσεις μεταξύ συγκεκριμένων μεθόδων επιλογής χαρακτηριστικών και του μεγέθους του συνόλου δεδομένων. Αυτό το στοιχείο, επαληθεύει το συμπέρασμα ότι, η ποιότητα του δείγματος, είτε είναι πολύ μεγάλη είτε περιέχει θορυβώδεις τιμές, διαδραματίζει βασικό ρόλο για τα προβλήματα που καλούμαστε να αντιμετωπίσουμε στην Εξόρυξη Δεδομένων. Επιπλέον, η τοποθέτηση και επιλογή χαρακτηριστικών μπορεί να οδηγήσει στην κατασκευή ενός νέου μοντέλου χρησιμοποιώντας το υπάρχον σύνολο. Σε πολλές περιπτώσεις, η κατασκευή χαρακτηριστικών είναι σε θέση να προσφέρει καλύτερη ικανότητα διάκρισης συγκριτικά με μια καλή επιλογή χαρακτηριστικών.

Υπάρχουν διάφοροι αλγόριθμοι ταξινόμησης, που αν και ανταποκρίνονται καλά σε μια

συγκεκριμένη κατηγορία προβλημάτων, δε μπορούν να εφαρμοστούν για την αντιμετώπιση άλλων παρόμοιων προβλημάτων. Αυτός ο τύπος αλγορίθμων μπορεί να εφαρμοστεί και σε άλλα προβλήματα μέσω του μετασχηματισμού των δεδομένων [142].

Ένα άλλο ενδιαφέρον σημείο είναι το ζήτημα της μ (dataset shift). Η μετατόπιση του συνόλου δεδομένων εμφανίζεται όταν τα δεδομένα δοκιμών βιώνουν ένα περιστατικό το οποίο οδηγεί: (i) Σε μια αλλαγή στην κατανομή ενός μόνο χαρακτηριστικού, (ii) σε ένα συνδυασμό χαρακτηριστικών ή (iii) στα όρια της κλάσης. Ως αποτέλεσμα, η κοινή υπόθεση ότι τα δεδομένα εκπαίδευσης και δοκιμών ακολουθούν τις ίδιες κατανομές, συχνά καταχράται σε πραγματικές εφαρμογές [318]. Για περισσότερες λεπτομέρειες σχετικά με τη μετατόπιση του συνόλου δεδομένων και τις σχετικές διεργασίες σε αυτό το πεδίο, παραπέμπουμε τον αναγνώστη στην εργασία [266].

Παρά την πρόοδο που έχει σημειωθεί από την ερευνητική κοινότητα τα τελευταία χρόνια, εξακολουθούν να υπάρχουν ορισμένα θέματα προς επίλυση. Για παράδειγμα, το ζήτημα της μ (managing data) με την πάροδο του χρόνου σε μη σταθερά περιβάλλοντα [318]. Σε ένα τέτοιο πρόβλημα, είναι απαραίτητο ένα προσαρμοστικό μοντέλο [247]. Επιπλέον, οι μέθοδοι προεπεξεργασίας δεδομένων για τη διαχείριση μ (data streams), [320] θα μελετηθούν εκτενώς στο μέλλον, αφού είναι σε θέση να αντιμετωπίσουν σημαντικά ζητήματα του Διαδικτύου και των τεχνολογιών για μαζική συλλογή δεδομένων σε απευθείας σύνδεση.

Επιπλέον, το ερώτημα «

μ , μ , μ , μ ;» παραμένει αναπάντητο. Μία πιθανή ακολουθία βημάτων, θα μπορούσε να είναι τα βήματα που χρησιμοποιήθηκαν για τη διαίρεση των ενοτήτων αυτού του κεφαλαίου. Παρόλα αυτά, για να έχει νόημα αυτός ο ισχυρισμός, απαιτούνται εκτεταμένα πειράματα και περαιτέρω μελέτη. Πρόθεσή μας είναι να αντιμετωπιστούν αυτά τα κρίσιμα ζητήματα και να επεξεργαστούν οι απαραίτητες λεπτομέρειες στο άμεσο μέλλον. Τέλος, προκειμένου ένας άπειρος αναγνώστης να εξοικειωθεί με το θέμα αυτό, παρέχουμε έναν απλό κώδικα υλοποίησης των πιο κοινά χρησιμοποιούμενων αλγορίθμων για την προεπεξεργασία δεδομένων στο Παράρτημα της διδακτορικής διατριβής.

Ακόμα, ο Πίνακας 4.6, απεικονίζει την εφαρμοσιμότητα των τεχνικών προεπεξεργασίας δεδομένων που παρουσιάζονται στο Κεφάλαιο 4, σε σχέση με τους πιο γνωστούς αλγορίθμους. Ο αναγνώστης μπορεί να διακρίνει με το σήμα 'X', τους αλγορίθμους που απαιτούν ένα αντίστοιχο βήμα προεπεξεργασίας. Από την άλλη πλευρά, με το σημάδι '7', υποδηλώνεται η ανάγκη για προεπεξεργασία δεδομένων, για παράδειγμα στην περίπτωση των δέντρων απόφασης και της διαδικασίας επιλογής παραδειγμάτων. Επιπλέον, το σήμα ' ', υποδεικνύει σε ποια μέθοδο μπορεί να εφαρμοστεί ένα από τα παραπάνω βήματα προεπεξεργασίας, αλλά υπό ορισμένες περιπτώσεις.

Συμπερασματικά, λαμβάνοντας υπόψη τις πληροφορίες που έχουμε αναλύσει σε κάθε ενότητα, ο αναγνώστης είναι σε θέση να συνειδητοποιήσει τη σημασία των διαδικασιών προεπεξεργασίας των δεδομένων στην προβλεπτική εξόρυξη δεδομένων. Είναι προφανές ότι, η εφαρμογή της προεπεξεργασίας (ή όχι), μπορεί να επηρεάσει την απόδοση ενός ταξινομητή. Η σειρά με την οποία πρέπει να εκτελεστούν τα βήματα δεν είναι προφανής. Αρκετές μελέτες, όπως σημειώνεται στο παρόν κεφάλαιο, έχουν διεξαχθεί για να προσδιοριστεί η αποτελεσματικότητα τέτοιων αλγορίθμων ταξινόμησης, υπό ορισμένα βήματα προεπεξεργασίας δεδομένων. Δυστυχώς, κάτι τέτοιο δεν είναι εφικτό για όλα τα υπάρχοντα προβλήματα. Ωστόσο, πιστεύουμε ότι μέσω της ανάλυσης αναφορών, που παρέχουμε για κάθε τεχνική προεπεξεργασίας, είμαστε σε θέση να λάβουμε πληροφορίες για τον αντίκτυπο κάθε μεθόδου. Για παράδειγμα, δε μπορεί να θεωρηθεί σύμπτωση ότι ο αλγόριθμος SMOTE, σχετικά με τη διαδικασία επιλογής χαρακτηριστικών, έχει ξεχωρίσει από τους αλγορίθμους της ίδιας κατηγορίας, και έχει συγκεντρώσει τις περισσότερες αναφορές σε σύγκριση με τις υπόλοιπες τεχνικές. Συνεπώς, ο αναγνώστης, λαμβάνοντας αυτό υπόψη, είναι σε θέση να αναγνωρίσει ποια μέθοδος

		LMs	DTs	ANN	RLs	Bayesian	SVM	LLs
		7	×	7	×	×	7	7
		×	7	×	7	7	×	×
		7	7		7	7		
		7				7	7	7
	μ		7	7	7	7		7
	μ	7	7	×	7	7	×	×
	μ	×	×	×	×	7	×	×

Πίνακας 4.6: Εφαρμογή/χρησιμότητα των διαδικασιών προεπεξεργασίας σχετικά με τους γνωστούς μαθητές. Τα 'LMs' υποδηλώνουν τα γραμμικά μοντέλα, το 'DTs' υποδεικνύει τα δέντρα απόφασης, το 'ANN' υποδεικνύει τα τεχνητά νευρωνικά δίκτυα, το 'RLs' υποδηλώνει τους μαθητές κανόνων, τα 'SVM' χαρακτηρίζουν τις μηχανές διανυσματικής υποστήριξης και τα 'LLs' υποδηλώνουν τους «τεμπέληδες» μαθητές.

παρουσιάζει σημαντικό αντίκτυπο, και συνεπώς, να την επιλέξει για την προεπεξεργασία δεδομένων.

Εν κατακλείδι, το Κεφάλαιο 4, θεωρούμε πως συμβάλλει σημαντικά, ώστε ο αναγνώστης να ενημερωθεί σφαιρικά για τις πιο πρόσφατες, εκτενώς χρησιμοποιούμενες και ευρέως αποδεκτές τεχνικές προεπεξεργασίας δεδομένων. Συγκεκριμένα, ο αναγνώστης δύναται να ενημερωθεί ξεχωριστά για κάθε βήμα προεπεξεργασίας και έτσι, να είναι σε θέση να αναλογιστεί πλεονεκτήματα και μειονεκτήματα, προκειμένου να συνθέσει δικές του πρωτότυπες μεθόδους για κάθε βήμα προεπεξεργασίας. Ακόμα, το παρόν κεφάλαιο δίνει μια πλήρη εικόνα για τα ανοιχτά ζητήματα που υπάρχουν, και θεωρούμε πως θέτει τις βάσεις, ώστε μελλοντικά να προταθεί μια ιεραρχία βημάτων, αναλόγως με το υπό εξέταση πρόβλημα και το δοθέν σύνολο δεδομένων. Τέλος, αποτελεί το βασικό στάδιο, όπως φαίνεται και στο Σχήμα 4.1, το οποίο έπεται της συλλογής δεδομένων, προκειμένου αυτό να καταστεί χρησιμοποιήσιμο και αποτελεσματικό, προτού δοθεί ως είσοδος σε κάποιον αλγόριθμο μάθησης.

Εντοπισμός και αναγνώριση ακραίων τιμών

Σχεδόν ο,τι κάνεις είναι ασήμαντο, αλλά είναι σημαντικό να το κάνεις.

— μ (1869-1948)

Ένας μεγάλος αριθμός παραγόντων επηρεάζει την εφαρμογή μεθόδων () (Machine Learning (ML)) και τη χρήση μ (Statistical models) για μια δεδομένη εργασία. Η ύπαρξη μ (outliers) σε ένα σύνολο δεδομένων είναι ένα ζήτημα που συναντάται συχνά και πρέπει να αντιμετωπιστεί αποτελεσματικά. Ο προσδιορισμός αυτών των τιμών είναι ένα δύσκολο, αλλά πολύ χρήσιμο έργο. Σε πολλές περιπτώσεις τα (errors) ή οι μ μ (dissimilar values) στην πλειοψηφία των δεδομένων είναι άχρηστα. Παρόλα αυτά, πολύτιμες πληροφορίες μπορούν να κρύβονται στο σύνολο των ακραίων τιμών. Τα τελευταία χρόνια, μολονότι έχουν αναπτυχθεί αρκετά μοντέλα για την ανίχνευση των τιμών αυτών, υπάρχει πάντα χώρος για νέες, ευφυείς, αποδοτικότερες και λιγότερο χρονοβόρες τεχνικές για αυτό το ζήτημα. Στο παρόν κεφάλαιο παρέχουμε μια νέα μ (ensemble method) για ανίχνευση των ακραίων τιμών [10]. Για να ελεγχθεί η προτεινόμενη μεθοδολογία, γίνονται συγκρίσεις με ευρέως χρησιμοποιούμενες τεχνικές για την ανίχνευση των ακραίων τιμών. Τα αποτελέσματα που προέκυψαν δείχνουν ότι το προτεινόμενο μοντέλο είναι εύρωστο (robust) και αρκετά ανταγωνιστικό έναντι των υπολοίπων μεθόδων.

5.1 Μια σύντομη επισκόπηση

Ένα κοινό πρόβλημα που παρατηρείται στα σύνολα δεδομένων είναι η ύπαρξη μ (outliers) ή μ (anomalies). Αυτό το πρόβλημα είναι αρκετά συνηθισμένο και μπορεί να επηρεάσει σημαντικά την απόδοση των αλγορίθμων MM ή των Στατιστικών μοντέλων [99]. Η ραγδαία ανάπτυξη της τελευταίας δεκαετίας στον τομέα των υλικών και των τεχνολογιών, έχει καταστήσει εφικτό το χειρισμό μεγάλων συνόλων δεδομένων. Συνεπώς, μπορούν να χρησιμοποιηθούν μεγάλα ποσά δεδομένων στον τομέα της μ (Computer Science) και έτσι, μπορούν να αντληθούν χρήσιμα συμπεράσματα. Επιπλέον, νέες στρατηγικές και μοντέλα που ανήκουν στον τομέα της MM και της μ () (Data Mining (DM)), καθιστούν τα αξιόπιστα δεδομένα απαραίτητα. Αυτό το ζήτημα είναι σημαντικό, προκειμένου να δημιουργηθούν ακριβή, ευφυή και γρήγορα μ (Computing Systems).

Οι όροι μ (outlier) ή μ (anomaly) ή μ (exception) ή μ (discordant) χρησιμοποιούνται συχνά για αυτόν τον τύπο δεδομένων [59]. Στις περισσότερες περιπτώσεις, οι δύο πρώτες από τις παραπάνω έννοιες είναι ευρέως αποδεκτές και αναγνωρίσιμες. Ένα σημαντικό ερώτημα όμως, που απαιτεί μια απάντηση είναι το εξής: « μ μ μ » Αυτά τα δεδομένα είναι συχνά σφάλματα στις

καταγεγραμμένες τιμές ή εξαιρέσεις και έτσι, δεν χαρακτηρίζουν την κοινή λειτουργία των δεδομένων κατά πλειοψηφία. Κατά συνέπεια, μια πραγματικά χρήσιμη ενέργεια είναι να εκτελεστεί μια τεχνική απομάκρυνσης αυτών των ανωμαλιών ως προετοιμασία για περαιτέρω επεξεργασία των δεδομένων. Παρόλο που κάποιος μπορεί να πιστεύει ότι οι ακραίες τιμές σε ένα σύνολο δεδομένων είναι άχρηστες, αυτό δεν είναι πάντα αληθές, δεδομένου ότι αυτές οι εξαιρέσεις μπορούν να παράσχουν ουσιαστικές πληροφορίες για υπάρχουσες μ στο σύνολο δεδομένων. Ως εκ τούτου, ο προσδιορισμός των ακραίων τιμών είναι μια εξαιρετικά ωφέλιμη διεργασία, λόγω των πρόσθετων πληροφοριών που μπορούν να παράσχουν, για ένα δεδομένο σύνολο παρατηρήσεων.

Η σημασία ενός «καθαρού» συνόλου δεδομένων μπορεί εύκολα να γίνει κατανοητή αν ληφθεί υπόψη η επιθυμητή, καλή ικανότητα γενίκευσης μιας μεθόδου MM. Εάν ένα σύνολο δεδομένων είναι μ ή μ παρουσιάζει άλλα συγκεκριμένα προβλήματα, τότε η υπό κατασκευή μέθοδος είναι δυνατόν να δώσει ανεπαρκή αποτελέσματα όσον αφορά άγνωστα δεδομένα. Γενικά, στον τομέα της MM μάς ενδιαφέρει να προβλέψουμε ένα αποτέλεσμα που καθορίζεται από ορισμένα δεδομένα [413]. Έτσι, πολλές τεχνικές που χρησιμοποιούνται στη MM μπορούν να θεωρηθούν ως τυπικές πιθανοτικές μέθοδοι. Η MM μοιράζεται, επίσης, ομοιότητες με τον τομέα της Στατιστικής, καθώς οι ερευνητές ενδιαφέρονται για τις καλές προβλέψεις. Επιπλέον, η υπολογιστική αποδοτικότητα παίζει σημαντικό ρόλο στην κατασκευή ενός αλγορίθμου [413]. Συνήθως, ο βασικός σκοπός είναι, είτε να παράσχει κάποιες αρχικές ιδέες που συνδέονται με ένα πεδίο εφαρμογής, όπου υπάρχουν ήσσονος σημασίας *a priori* γνώσεις ή, να είναι σε θέση να προβλέψει γρήγορα και με μια δεδομένη ακρίβεια μελλοντικές παρατηρήσεις.

Στη βιβλιογραφία έχουν ληφθεί υπόψη οι μ μ μ και μ μ (multivariate outliers) [184]. Οι μονοπαραγοντικές ακραίες τιμές αφορούν ένα μόνο χαρακτηριστικό σε ένα χώρο διαστάσεων, ενώ οι πολυπαραγοντικές εξαιρέσεις αφορούν πολλά χαρακτηριστικά σε μια πολυδιάστατη περιοχή. Προκειμένου να αντιμετωπιστεί το πρόβλημα μ ή μ μ , αναπτύχθηκε μια πληθώρα τεχνικών. Οι τεχνικές αυτές ταξινομούνται ως εξής [184]:

- (α) μ (Supervised): Όταν το σύνολο δεδομένων χωρίζεται σε δύο επισημασμένα σύνολα, το ένα ως μ και το άλλο ως μ και περιλαμβάνει την εκπαίδευση ενός ταξινομητή,
- (β) μ (Unsupervised): Όταν ανιχνεύονται ακραίες τιμές σε ένα μη επισημασμένο σύνολο δεδομένων δοκιμών, υποθέτοντας ότι η πλειονότητα των δεδομένων στο σύνολο είναι φυσιολογικά δεδομένα και
- (γ) μ μ (Semi-supervised): Όταν κατασκευάζεται μια μέθοδος που παρουσιάζει κανονική συμπεριφορά αναφορικά με ένα δεδομένο σύνολο φυσιολογικών παραδειγμάτων εκπαίδευσης και στη συνέχεια, εξετάζει την πιθανότητα ένα παράδειγμα δοκιμής να παράγεται από τον αλγόριθμο που εκπαιδεύτηκε.

Στο Κεφάλαιο 5, προτείνεται μια νέα μεθοδολογία, προκειμένου να εξεταστεί το κατά πόσο ένα παράδειγμα είναι ακραία τιμή ή όχι. Για να δοκιμάσουμε την απόδοση της προτεινόμενης μεθόδου, τέθηκαν σε σύγκριση πολύ γνωστές και ευρέως χρησιμοποιούμενες τεχνικές ανίχνευσης ακραίων τιμών. Τα πειραματικά αποτελέσματα που προέκυψαν δείχνουν έντονα ότι το προτεινόμενο σχήμα είναι σθεναρό και παρέχει ανταγωνιστικά αποτελέσματα. Το υπόλοιπο του Κεφαλαίου 5 είναι οργανωμένο ως εξής: Στην Ενότητα 5.2 συγκαταλέγονται παρόμοιες μελέτες που σχετίζονται με το πρόβλημα της αναγνώρισης ή/και απομάκρυνσης των ακραίων τιμών, καθώς και γνωστοί αλγόριθμοι και τεχνικές που αντιμετωπίζουν το πρόβλημα της ανίχνευσης των ανωμαλιών. Στην Ενότητα 5.3, αναλύονται η προσέγγισή μας και η προτεινόμενη μεθοδολογία [10] μαζί με τα αντίστοιχα πειραματικά αποτελέσματα.

Τέλος, στην Ενότητα 5.4 καταγράφονται οι τελικές παρατηρήσεις και ορισμένες μελλοντικές ερευνητικές μελέτες.

5.2 Γνωστές τεχνικές: Μια εμπειρική κατηγοριοποίηση

Όπως έχει ήδη επισημανθεί στην Ενότητα 5.1 αλλά και Ενότητα 4.2, το ζήτημα της ανίχνευσης και/ή εξάλειψης των ακραίων τιμών είναι πολύ σημαντικό και σχετίζεται με πολλά προβλήματα της ΕΔ [264, 355, 358]. Προς το σκοπό αυτό έχουν αναπτυχθεί πρόσφατα πολλές μέθοδοι. Στην ενότητα που ακολουθεί, παρουσιάζουμε τα πιο αξιόπιστα και γνωστά μοντέλα για το πρόβλημα ανίχνευσης ακραίων τιμών. Για διευκόλυνση του αναγνώστη, αυτές οι μέθοδοι κατηγοριοποιούνται και παρουσιάζονται σε έξι κατηγορίες.

5.2.1 Μέθοδοι που βασίζονται στον πλησιέστερο γείτονα

Οι πιο γνωστοί αλγόριθμοι αυτής της κατηγορίας είναι ο k - Nearest Neighbors (kNN) Detector και ο αλγόριθμος AveKNN [16]. Συγκεκριμένα, για ένα παράδειγμα του συνόλου δεδομένων, ο πρώτος αλγόριθμος αξιολογεί την απόσταση των k πλησιέστερων γειτόνων του. Αυτή η απόσταση είναι το εξαγόμενο αποτέλεσμα-σκορ, που είναι ο μηχανισμός για την καταμέτρηση της σχετικής πυκνότητας. Στη δεύτερη μέθοδο, μπορούν να χρησιμοποιηθούν τρεις «ανιχνευτές» k NN: (α) Ο μ : Σε αυτόν τον ανιχνευτή η απόσταση από τον k γείτονα θεωρείται ως η πιο ακραία μέτρηση, (β) Θεωρούμε τη μ : Σε αυτή τη στρατηγική ο μέσος όρος όλων των k NN αλγορίθμων χρησιμοποιείται ως σκορ για τις ακραίες τιμές και (γ) Θεωρούμε τον μ ανιχνευτή: Σε αυτή την προσέγγιση, ο μέσος όρος των αποστάσεων χρησιμοποιείται ως το αποτέλεσμα των ακραίων τιμών.

Πρόσφατα, έχει αναπτυχθεί ένα μη εποπτευόμενο μοντέλο που ονομάζεται μ μ μ (Histogram-Based Outlier Score (HBOS)) [157]. Στην προσέγγιση αυτή, η κύρια ιδέα βασίζεται σε ιστογράμματα, όπως υποδεικνύεται από το όνομα της μεθόδου. Συγκεκριμένα, το μοντέλο θεωρεί τον αυτοπροσδιορισμό των χαρακτηριστικών, που παράγουν μια πολύ γρήγορη μέθοδο, σε σχέση με άλλες πολυμεταβλητές μεθόδους από πλευράς κόστους μικρότερης ακρίβειας.

Στο άρθρο [47], έχει αναπτυχθεί ένα μοντέλο που βασίζεται στον μ (Local Outlier Factor (LOF)). Αυτή η μέθοδος αναγνωρίζει ανωμαλίες σε πολυδιάστατα σύνολα δεδομένων. Ο προαναφερθείς παράγοντας επισημαίνει αν ένα στοιχείο είναι ακραία τιμή ή όχι. Έτσι, ο αλγόριθμος αξιολογεί την τοπική απόκλιση της πυκνότητας ενός συγκεκριμένου σημείου σε σχέση με τους γείτονές του. Ακριβέστερα, η τοποθεσία (locality) προσδιορίζεται από τον αλγόριθμο k NN, η απόσταση του οποίου χρησιμοποιείται για την αξιολόγηση της τοπικής πυκνότητας. Μέσω αυτής της σύγκρισης, της τοπικής πυκνότητας ενός παραδείγματος με τις πυκνότητες των γειτόνων του, μπορεί κανείς να εντοπίσει περιπτώσεις που έχουν σημαντικά χαμηλότερη πυκνότητα συγκριτικά με τους γείτονές τους. Αυτές οι τιμές αξιολογούνται τελικώς ως ανωμαλίες.

5.2.2 Πιθανοτικές μέθοδοι

Οι συγγραφείς του άρθρου [223], έχουν αναπτύξει ένα μοντέλο που μπορεί να εφαρμοστεί σε σύνολα δεδομένων μεγάλης διαστάσεως και διαφέρει από προσεγγίσεις που βασίζονται στις αποστάσεις. Συγκεκριμένα, το λεγόμενο μ μ (Angle-Based Outlier Detection (ABOD)) χρησιμοποιεί γωνίες αντί για αποστάσεις. Γενικά, η γωνία παρέχει ένα πιο ισχυρό μέτρο από αυτό της απόστασης, στην περίπτωση δεδομένων μεγάλης διαστάσεως. Οι συγγραφείς του άρθρου έχουν παρατηρήσει ότι η γωνιακή διακύμανση είναι ακριβέστερη, και για να επιταχυνθεί η συνολική διαδικασία, χρησιμοποιήθηκε ένα υποσύνολο των δεδομένων για αξιολόγηση.

Μια προσέγγιση μ με την ονομασία μ (Stochastic Outlier Selection (SOS)) παρουσιάστηκε στην εργασία [201]. Πρόκειται για μια στοχαστική μέθοδο και βασίζεται στη συγγένεια ή συνάφεια (affinity), προκειμένου να προσδιοριστεί η σχέση μεταξύ των δεδομένων με «συγγενική σχέση». Η συνάφεια αντιστοιχεί στο ποσό της ομοιότητας μεταξύ δύο τιμών. Επομένως, μια τιμή δεδομένων εκτιμάται ως ακραία τιμή, εάν η συνάφειά της μεταξύ όλων των άλλων σημείων είναι κακή.

Ένας συνδυασμός της Mahalanobis (Mahalanobis Squared Distance (MSD)) και της ορίζουσας ελάχιστης συνδιακύμανσης (Minimum Covariance Determinant (MCD)) χρησιμοποιήθηκε στο άρθρο [175], προκειμένου να παραχθεί μια μέθοδος ανίχνευσης ανωμαλιών. Συγκεκριμένα, εάν μια τιμή δεν περιλαμβάνεται σε κάποια συστάδα, τότε θεωρείται ως ακραία τιμή. Από στατιστική άποψη, ο ελάχιστος συντελεστής κατανομής (Minimum Distribution Coefficient) μπορεί να εφαρμοστεί μόνο σε δεδομένα από τη Gaussian κατανομή, αλλά θα μπορούσε επίσης, να χρησιμοποιηθεί ικανοποιητικά για δεδομένα από μία μονομερή συμμετρική κατανομή.

5.2.3 Μέθοδοι που βασίζονται στον υπόχωρο

Η γνωστή και ευρέως χρησιμοποιούμενη μέθοδος ανάλυσης κύριων συνιστωσών (Principal Component Analysis (PCA)) έχει χρησιμοποιηθεί στην εργασία [352], προκειμένου να ανιχνεύσει ανωμαλίες σε ένα σύνολο δεδομένων. Σε αυτή την προσέγγιση, τα ιδιοδιανύσματα του πίνακα συνδιακύμανσης δεδομένων με υψηλές ιδιοτιμές επικυρώνουν το μεγαλύτερο μέρος της διακύμανσης στα δεδομένα. Συνεπώς, η διαστασιμότητα μειώνεται και δημιουργείται ένα υπερεπίπεδο από τα K ιδιοδιανύσματα των κυριότερων συνιστωσών, τα οποία καταγράφουν το μεγαλύτερο μέρος της διακύμανσης στα δεδομένα. Οι ακραίες τιμές, που είναι αρκετά ανόμοιες με τις κανονικές τιμές δεδομένων, μετριούνται τόσο με την απόστασή τους από τις κύριες, όσο και τις δευτερεύουσες συνιστώσες.

Στην εργασία [222], έχει αναπτυχθεί μία μέθοδος μ (Subspace Outlier Detection (SOD)), η οποία αντιμετωπίζει το πρόβλημα ανίχνευσης ακραίων τιμών σε διάφορες υπο-κατηγορίες ενός χώρου μεγάλης διαστάσεως. Χρησιμοποιώντας αυτή τη μέθοδο δεν είναι δυνατό να βρεθεί μια ακραία τιμή στον αρχικό χώρο δεδομένων. Έτσι, για κάθε δεδομένο, η μέθοδος SOD ψάχνει τον υπόχωρο του παράλληλου-άξονα (axis-parallel subspace) για να διαπιστώσει πόσο διαφορετικό είναι το αντικείμενο από τους γείτονές του σε αυτόν τον υπόχωρο.

5.2.4 Συνδυαστικές μέθοδοι

Ένα μοντέλο που δεν εξαρτάται από τις αποστάσεις ή τα μέτρα πυκνότητας έχει αναπτυχθεί στο άρθρο [238]. Συγκεκριμένα, η παρεχόμενη μέθοδος, ονομάζεται μ (Isolation Forest (iForest)) και διακρίνει τις ακραίες τιμές με βάση την ιδέα της απομόνωσης. Ένα γεγονός που αξίζει να σημειωθεί, είναι ότι ο αλγόριθμος iForest είναι σε θέση να αξιοποιήσει την υπο-δειγματοληψία (sub-sampling). Αυτός ο αλγόριθμος, έχει απαιτήσεις μικρής γραμμικής πολυπλοκότητας χρόνου, καθώς και μικρή μνήμη αποθήκευσης. Επιπλέον, μπορεί να χειριστεί αποτελεσματικά δύο συχνά παρατηρούμενα φαινόμενα. Συγκεκριμένα: τον μ (swamping) και την μ (masking).

Μια μη-εποπτευόμενη μέθοδος συνόλου, η οποία βασίζεται σε τοπικές περιοχές κοντά σε ένα παράδειγμα δοκιμής, ονομάζεται μ (Locally Selective Combination in Parallel Outlier Ensembles (LSCP)), παρουσιάστηκε στο άρθρο [440]. Συγκεκριμένα, η τοπική περιοχή προσδιορίζεται, ώστε να είναι η συλλογή των πλησιέστερων δεδομένων εκπαίδευσης σε τυχαίες δειγματοληψίες υποκλάσης χαρακτηριστικών, καθώς λαμβάνουν χώρα συχνότερα από ένα καθορισμένο όριο σε πολλαπλές επαναλήψεις. Η χρήση της τοπικής περιοχής έχει ως αποτέλεσμα μια τοπική ψευδοεξαρτημένη αλήθεια και η συσχέτιση Pearson προορίζεται μεταξύ των αποτελεσμάτων

της ανωμαλίας της κύριας ανάλυσης του ανιχνευτή και της ψεύτικης αλήθειας. Επιπλέον, δημιουργείται ένα ιστόγραμμα από τα αποτελέσματα συσχέτισης Pearson. Τέλος, το αποτέλεσμα είναι το μέσο σκορ ανωμαλίας των κατάλληλα επιλεγμένων ανιχνευτών.

Στην εργασία [227], έχει προταθεί μια νέα προσέγγιση που ανιχνεύει ακραίες τιμές, συνδυάζοντας την απόφαση πολλαπλών αλγορίθμων για τον εντοπισμό ακραίων τιμών για υψηλής διαστασιμότητας θορυβώδη σύνολα δεδομένων. Ως αποτέλεσμα, παρατηρήθηκαν ποικίλες αποδόσεις. Ένας ανιχνευτής χαρακτηριστικών «σάκου» (Feature Bagging (FB)) χρησιμοποιήθηκε ως μ - μ (meta estimator), ο οποίος ρυθμίζει μια ομάδα βασικών ανιχνευτών σε διάφορα υποδείγματα του αρχικού συνόλου δεδομένων. Αυτός, χρησιμοποιεί το μέσο όρο ή το συνδυασμό άλλων μοντέλων για να επιτύχει καλύτερη ακρίβεια πρόβλεψης και να αποφύγει το φαινόμενο υπερταυρίσματος (over-fitting).

5.2.5 Μέθοδοι που βασίζονται στον ταξινομητή

Οι συγγραφείς της εργασίας [342], πρότειναν ένα μοντέλο, το οποίο ονομάζεται μ μ (One-Class Support Vector Machine (OCSVM)), το οποίο είναι παρόμοιο με τη γνωστή μέθοδο SVM, για να ανιχνεύσει τις ακραίες τιμές ενός συνόλου δεδομένων. Έτσι, απαιτείται η επιλογή ενός πυρήνα καθώς και μια κλιμακωτή παράμετρος. Συχνά, προτιμάται η επιλογή του πυρήνα RBF.

5.2.6 Μέθοδοι που βασίζονται στη συσταδοποίηση

Μια νέα μέθοδος που ονομάζεται μ μ (Cluster-Based Local Outlier Factor (CBLOF)) προτάθηκε στο άρθρο [177]. Το μοντέλο που χρησιμοποίησε την παραπάνω τεχνική ονομάστηκε μέθοδος FindCBLOF, και τα πειραματικά αποτελέσματα έχουν δείξει ότι η παρεχόμενη μέθοδος είναι αρκετά ανταγωνιστική σε σύγκριση με άλλες παραλλαγές συσταδοποίησης. Συγκεκριμένα, το προτεινόμενο μοντέλο λαμβάνει υπόψη το αρχικό σύνολο δεδομένων και τον αλγόριθμο συσταδοποίησης. Στη συνέχεια, το σκορ που δίνει για τις ακραίες τιμές παρέχεται βάσει του μεγέθους της συστάδας στην οποία βρίσκονται τα δεδομένα, καθώς και της απόστασης από την πλησιέστερη μεγαλύτερη συστάδα.

5.3 Προτεινόμενη μεθοδολογία και πειραματικά αποτελέσματα

Οι συνδυαστικές μέθοδοι (ensemble methods) αποτελούν μοντέλα που εμπιστεύονται όλο και περισσότερο οι ερευνητές. Αυτό οφείλεται στο γεγονός ότι αυτές οι μέθοδοι μπορούν να κατασκευαστούν εμπειρικά και συνδυάζουν τις αποφάσεις αξιόπιστων μεθόδων, προκειμένου να παραχθεί ένα ακόμα πιο αξιόπιστο και αποτελεσματικό τελικό μοντέλο. Έχουν παρουσιαστεί πολυάριθμες μέθοδοι για τη δημιουργία ενός τέτοιου συνεργατικού πλαισίου, όπως σημειώνεται στην εργασία [95]. Σε αυτή την ενότητα, αναλύουμε μια νέα συνεργατική μέθοδο για τον εντοπισμό των ακραίων τιμών. Στη συνέχεια, παρέχουμε τη μεθοδολογία που υιοθετείται στην υβριδική μέθοδό μας. Η προτεινόμενη μεθοδολογία, η οποία αποτελεί στρατηγική έξι βημάτων, παρουσιάζεται στον Αλγόριθμο 1.

Αξίζει να σημειωθεί ότι, το προτεινόμενο μοντέλο συνόλου ταξινομητών, μπορεί απλώς να παραλληλιστεί χρησιμοποιώντας έναν ανιχνευτή ανά μηχανή. Ο παράλληλος και κατανεμημένος υπολογισμός έχει μεγάλη σημασία για τους επαγγελματίες, καθώς, όταν αξιοποιείται μια παράλληλη ή κατανεμημένη υλοποίηση, μια μεθοδολογία μπορεί: (α) να αυξήσει την ταχύτητά της, και (β) να επεκτείνει το φάσμα εφαρμογών, όπου μπορεί να χρησιμοποιηθεί (δηλαδή μπορεί να επεξεργαστεί περισσότερα δεδομένα).

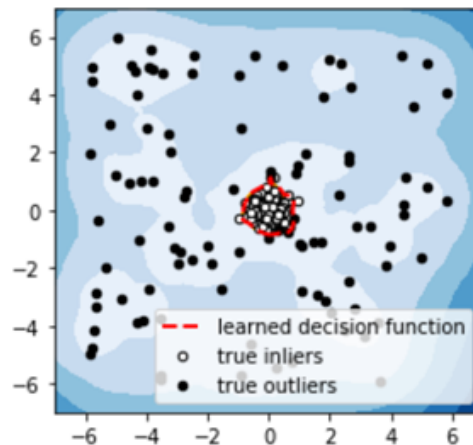
Αξίζει, επίσης, να σημειωθεί ότι η πολυπλοκότητα χρόνου, που απαιτεί η επιλογή του καλύτερου υποσυνόλου ανιχνευτών, αυξάνεται σε σχέση με τον αριθμό των βασικών ανιχνευτών που χρησιμοποιούνται. Από αυτή την άποψη, ο ευρετικός κανόνας για τον έλεγχο

των αλγορίθμων, χρησιμοποιώντας ένα μικρό υποσύνολο του συνόλου εκπαίδευσης, μειώνει την υπολογιστική πολυπλοκότητα. Οι ανιχνευτές που χρησιμοποιούνται αρχικά για την δημιουργία του συνόλου, τεστάρονται σε ένα μικρό υποσύνολο του συνόλου εκπαίδευσης και μόνο οι τρεις καλύτεροι από αυτούς συμμετέχουν στην τελική απόφαση του μοντέλου του συνόλου (ensemble model).

- 1 :** Λαμβάνεται ένα δείγμα περίπου 20% του αρχικού συνόλου δεδομένων.
- 2 :** Το εξαγόμενο σύνολο δεδομένων χωρίζεται τυχαία σε τρία ίσα ή σχεδόν ίσα μέρη.
- 3 :** Δύο από αυτά τα τρία μέρη χρησιμοποιούνται για την εκπαίδευση των ανιχνευτών και τα υπόλοιπα δεδομένα χρησιμοποιούνται ως σύνολο δοκιμών.
- 4 :** Το αποτέλεσμα τριών δοκιμών υπολογίζεται κατά μέσο όρο.
- 5 :** Οι τρεις αλγόριθμοι που έχουν επιτύχει τις καλύτερες επιδόσεις, σύμφωνα με τη ROC, επιλέγονται για να δημιουργήσουν μια ισχυρή τελική μέθοδο.
- 6 :** Οι τρεις καλύτεροι αλγόριθμοι, εκτελούνται σε ολόκληρο το αρχικό σύνολο για να παράγουν το μοντέλο πρόβλεψης λαμβάνοντας τον μέσο όρο βαθμολογίας όλων των ανιχνευτών.

Αλγόριθμος 1. Προτεινόμενη μεθοδολογία

Ένα απλοποιημένο επεξηγηματικό παράδειγμα, της προαναφερθείσας μεθοδολογίας παρουσιάζεται στο Σχήμα 5.1.



Σχήμα 5.1: Ένα απλοποιημένο επεξηγηματικό παράδειγμα

Θεωρούμε ένα σύνολο δεδομένων με $n = 400$ σημεία δεδομένων, όπου η παράμετρος ακραίων τιμών προσδιορίζεται ως $ofraction = 0.25$. Έτσι, έχουμε $n_{inliers} = 300$ σημεία και $n_{outliers} = 100$ σημεία. Στη συνέχεια, όπως υποδεικνύεται στο Βήμα 1, του Αλγορίθμου 1, 80 σημεία από το σύνολο δεδομένων λαμβάνονται τυχαία. Στη συνέχεια, το εξαγόμενο σύνολο δεδομένων, σύμφωνα με το Βήμα 2, χωρίζεται σε τρία μέρη με 27, 27 και 26 σημεία. Στο Βήμα 3, δύο από τα προαναφερθέντα σύνολα χρησιμοποιούνται ως σύνολα εκπαίδευσης, ενώ το άλλο σύνολο χρησιμοποιείται ως σύνολο δοκιμών. Συγκεκριμένα, τα μαύρα σημεία στην Εικόνα 5.1 απεικονίζουν τις αληθείς ακραίες τιμές των δεδομένων, ενώ τα λευκά υποδηλώνουν τους inliers. Τα αποτελέσματα των τριών τεστ λαμβάνονται κατά μέσο όρο, όπως επισημαίνεται στο Βήμα 4. Στη συνέχεια, οι τρεις αποδοτικότεροι και αποτελεσματικότεροι εκπαιδευόμενοι (στην περίπτωση μας ο αλγόριθμος CBLOF, ο IForest και ο HBOS ήταν οι τρεις αλγόριθμοι με την καλύτερη απόδοση ROC για το δεδομένο παράδειγμα) συνδυάζονται, προκειμένου να δημιουργηθεί ένας ισχυρότερος τελικός ταξινομητής. Η μεθοδολογία μας τερματίζεται στο Βήμα 6, όπου οι καλύτεροι αλγόριθμοι εκτελούνται σε όλα τα αρχικά

δεδομένα και τα σημεία που περιλαμβάνονται στους κόκκινους διακεκομμένους κύκλους είναι τα σημεία μάθησης, τα οποία έχουν προσδιοριστεί ως δεδομένα *inliers* από το μοντέλο μας.

Προκειμένου να ελεγχθεί η προτεινόμενη μέθοδος, οι καμπύλες (Receiver Operating Characteristics (ROC)), χρησιμοποιούνται ως μέτρο της απόδοσης των μεθόδων που εξετάζονται σε αυτή τη σύγκριση. Οι καμπύλες ROC τυπικά χαρακτηρίζουν το πραγματικά θετικό και το ψευδώς αρνητικό ποσοστό στον άξονα U και Q αντίστοιχα. Αυτό δείχνει ότι, η επάνω αριστερή γωνία της γραφικής παράστασης είναι το καλύτερο δυνατό σημείο - ένα ψευδώς θετικό ποσοστό από μηδέν, και ένα πραγματικά θετικό ποσοστό από ένα. Αυτό δεν είναι εφικτό στην πραγματικότητα, αλλά δείχνει ότι μια μεγαλύτερη μ (Area Under the Curve (AUC)) είναι συνήθως καλύτερη. Τα παρατηρούμενα αποτελέσματα παρουσιάζονται στον Πίνακα 5.1.

Επιπλέον, διεξήχθησαν στατιστικές δοκιμές προκειμένου να ελεγχθεί η σημασία των αποτελεσμάτων. Συγκεκριμένα, διεξήχθη η μη παραμετρική δοκιμή του Friedman [185]. Σύμφωνα με αυτό το στατιστικό τεστ, η μηδενική υπόθεση (όλες οι μέθοδοι παρουσιάζουν την ίδια απόδοση σε σχέση με την καμπύλη ROC) απορρίπτεται έντονα. Κατά συνέπεια, οι στατιστικές δοκιμές υποδεικνύουν ότι, υπάρχουν μοντέλα των οποίων η διαφορά απόδοσης ήταν στατιστικά σημαντική έναντι των υπολοίπων. Στον Πίνακα 5.3, παρουσιάζεται η κατάταξη των αλγορίθμων χρησιμοποιώντας το κριτήριο του Friedman. Επιπλέον, τα αποτελέσματα που ελήφθησαν από το στατιστικό τεστ Holm Bonferroni-Dunn [106], με τη χρήση της προτεινόμενης μεθόδου ως μεθόδου ελέγχου, παρουσιάζονται στον Πίνακα 5.4. Μπορεί να φανεί ότι η προτεινόμενη απόδοση της μεθόδου είναι καλύτερη από τις άλλες μεθόδους.

Ακόμα, χρησιμοποιήθηκε η μ - (Precision-Recall Curve (PRC)), που υποδεικνύει την αντιστάθμιση μεταξύ ακρίβειας και ανάκλησης για ένα διαφορετικό όριο. Δεδομένου ότι παρατηρείται η μεγαλύτερη δυνατή περιοχή κάτω από την καμπύλη, υποδηλώνεται υψηλή ανάκληση και υψηλή ακρίβεια για το προτεινόμενο μοντέλο. Εάν η ακρίβεια είναι υψηλή, τότε έχουμε ένα χαμηλό ψευδώς θετικό ποσοστό. Από την άλλη πλευρά, αν έχουμε μεγάλη ανάκληση, τότε υποδεικνύεται χαμηλός ψευδώς αρνητικός ρυθμός. Εάν και οι δύο βαθμολογίες είναι υψηλές, αυτό σημαίνει ότι ο ανιχνευτής επιστρέφει καλά αποτελέσματα (υψηλή ακρίβεια), καθώς επιστρέφει την πλειοψηφία όλων των θετικών αποτελεσμάτων (υψηλή ανάκληση). Στον Πίνακα 5.2, παρουσιάζονται τα αποτελέσματα που λαμβάνονται για την καμπύλη PRC. Τα αποτελέσματα των στατιστικών δοκιμών φαίνονται στους Πίνακες 5.5 και 5.6.

Για τη διεξαγωγή αυτής της πειραματικής μελέτης, χρησιμοποιήσαμε τη γνωστή γλώσσα προγραμματισμού Python. Συγκεκριμένα, το εργαλείο PyOD [441] έχει χρησιμοποιηθεί για τους αλγορίθμους ανίχνευσης ακραίων τιμών. Επιπλέον, η πλατφόρμα STAC [329] χρησιμοποιήθηκε για τη διεξαγωγή των συγκρίσεων και για τη λήψη των στατιστικών δοκιμών.

5.4 Συμπεράσματα - Μελλοντικές μελέτες

Έχει σημειωθεί σημαντική πρόοδος σε σχέση με τις μεθόδους ανίχνευσης ακραίων τιμών, όπως εύκολα μπορεί να συμπεράνει κανείς από τις παραπάνω αναφορές. Αξίζει να τονίσουμε ότι, έχουν αναπτυχθεί πολλές και ποικίλες μέθοδοι για την αντιμετώπιση του προβλήματος εντοπισμού ή/και αφαίρεσης τυχόν ανωμαλιών ή εξαιρέσεων. Οι εφαρμογές του προαναφερθέντος προβλήματος είναι πολλές και ποικίλες. Για το λόγο αυτό, η αξιόπιστη αντιμετώπιση αυτού του προβλήματος είναι επείγουσα και εξαιρετικά σημαντική.

Οι μέθοδοι συνόλου φαίνεται να είναι μια κατάλληλη και αξιόπιστη λύση σε πολλά προβλήματα της MM και ΕΔ. Έτσι, στο Κεφάλαιο 5, παρουσιάσαμε την κατασκευή μιας νέας μεθόδου για την ανίχνευση ανωμαλιών ή ακραίων τιμών σε ένα σύνολο δεδομένων. Η νέα μεθοδολογία αναλύεται και τα πειραματικά αποτελέσματα δείχνουν ότι η μέθοδός μας είναι εύρωστη και αρκετά ανταγωνιστική.

Πίνακας 5.1 : Receiver Operating Characteristics (ROC) curves

Data	# Samples	# Dimensions	Outlier Perc.	ABOD	CHLOF	FB	HBOS	IForest	KNN	LOF	MCD	OC SVM	PCA	SOS	SOD	AvdKNN	LSOP	Proposed
artlydimla	452	274	14.6018	0.7921	0.7892	0.7841	0.8379	0.8173	0.7917	0.7833	0.8317	0.7832	0.7861	0.6701	0.7528	0.7913	0.8062	0.8290
cardio	1831	21	9.6122	0.5694	0.8167	0.6223	0.8387	0.9186	0.7381	0.5790	0.8162	0.8328	0.7861	0.5351	0.6730	0.6739	0.9274	0.9361
glass	214	9	4.2056	0.8149	0.8857	0.9038	0.7703	0.7770	0.8828	0.9238	0.7986	0.5917	0.6396	0.7696	0.8378	0.8565	0.7914	0.9044
ionosphere	351	33	35.8974	0.9287	0.9004	0.8892	0.6128	0.8465	0.9274	0.8951	0.9544	0.8286	0.7932	0.7774	0.8858	0.9315	0.8225	0.9382
letter	1600	32	6.2500	0.8787	0.7864	0.8865	0.5750	0.6298	0.8771	0.8713	0.8010	0.6147	0.5378	0.8347	0.8941	0.9061	0.6950	0.8956
lympho	148	18	4.0541	0.9383	0.9944	0.9926	1.0000	1.0000	0.9833	0.9926	0.9377	0.9926	0.9963	0.7852	0.9435	0.9778	1.0000	1.0000
mnst	7603	100	9.2069	0.7897	0.8518	0.7137	0.5790	0.8062	0.8522	0.7097	0.8490	0.8595	0.8600	0.5410	0.6841	0.8406	0.8164	0.8547
mnst	3062	166	3.1679	0.0631	1.0000	0.4400	1.0000	0.9995	0.7252	0.4377	0.9998	0.9998	0.9998	0.4776	0.8632	0.4054	0.9564	1.0000
musik	5216	64	2.8758	0.4673	0.5345	0.5120	0.8555	0.7185	0.3895	0.5249	0.3914	0.5226	0.5215	0.5077	0.4981	0.3765	0.6080	0.7273
pendigits	6870	16	2.2707	0.6945	0.8673	0.4655	0.9298	0.9605	0.7576	0.4850	0.8372	0.8350	0.9361	0.5096	0.7082	0.7390	0.9499	0.9489
pinna	768	8	34.8958	0.6770	0.6517	0.5924	0.6827	0.6566	0.6988	0.6069	0.6686	0.5933	0.6107	0.5200	0.5983	0.6976	0.6644	0.6930
satellite	6435	36	31.6395	0.5769	0.7345	0.5666	0.7570	0.6928	0.6895	0.5646	0.8055	0.8636	0.5958	0.4783	0.6347	0.6739	0.6740	0.7723
satimage-2	5803	36	1.2235	0.8026	0.9899	0.4034	0.3916	0.9888	0.9479	0.3960	0.9959	0.9959	0.9943	0.5222	0.8003	0.9318	0.9956	0.9995
vehicle-2	240	6	12.5000	0.3891	0.4195	0.3990	0.3216	0.3878	0.3821	0.4010	0.3971	0.4534	0.4251	0.5312	0.4503	0.3739	0.3608	0.4783
vowel	1456	12	3.4341	0.9660	0.9276	0.9533	0.6862	0.7678	0.9795	0.9484	0.7566	0.7967	0.6226	0.7533	0.9030	0.9793	0.7242	0.9749
wbc	378	30	5.5556	0.8976	0.8847	0.9110	0.9575	0.9064	0.9201	0.9082	0.9017	0.9085	0.8860	0.6780	0.9254	0.9167	0.9019	0.9332

Πίνακας 5.2: Precision-Recall Curve (PRC)

Data	# Samples	# Dimensions	Outlier Perc.	ABOD	CBLOF	FB	HBOS	iForest	KNN	LOF	MCD	OCSVM	PCA	SOS	SOD	AveKNN	LSCP	Proposed
arrhythmia	452	274	14.6018	0.4410	0.4821	0.4560	0.5744	0.5617	0.4560	0.4554	0.5078	0.4821	0.4954	0.3615	0.4415	0.4693	0.5355	0.5572
cardio	1831	21	9.6122	0.2464	0.3932	0.1646	0.4576	0.4769	0.3348	0.1634	0.4134	0.4900	0.5816	0.1595	0.3293	0.3120	0.5031	0.5249
glass	214	9	4.2056	0.2778	0.2222	0.3889	0.0556	0.2222	0.2222	0.3889	0.0000	0.2222	0.0556	0.3889	0.2222	0.2222	0.0556	0.3889
ionosphere	351	33	35.8974	0.8492	0.8065	0.7205	0.4419	0.6401	0.8624	0.7203	0.8847	0.7135	0.5868	0.6663	0.7865	0.8779	0.6422	0.8750
letter	1600	32	6.2500	0.3924	0.2129	0.4515	0.1015	0.0991	0.3242	0.3833	0.1660	0.1562	0.0903	0.4504	0.4825	0.3872	0.1363	0.4615
lympho	148	18	4.0541	0.4167	0.9167	0.9167	1.0000	1.0000	0.9167	0.9167	0.3333	0.9167	0.9167	0.5833	0.7500	0.9167	1.0000	1.0000
mnist	7603	100	9.2069	0.3687	0.4062	0.3115	0.1233	0.3209	0.4139	0.3308	0.2929	0.3925	0.3825	0.1474	0.3025	0.4058	0.3384	0.4086
musk	3062	166	3.1679	0.0123	1.0000	0.1111	0.9753	0.9259	0.1605	0.1235	0.9630	1.0000	0.9630	0.0741	0.2840	0.0864	0.7408	0.9918
optdigits	5216	64	2.8758	0.0000	0.0000	0.0317	0.2173	0.0063	0.0000	0.0254	0.0000	0.0000	0.0000	0.0403	0.0000	0.0000	0.0063	0.0964
pendigits	6870	16	2.2707	0.0939	0.2289	0.0605	0.3368	0.3958	0.1240	0.0662	0.0994	0.3958	0.3647	0.0415	0.0771	0.1118	0.4003	0.3869
pima	768	8	34.8958	0.5314	0.4587	0.4285	0.5234	0.4810	0.5207	0.4284	0.5045	0.4243	0.4472	0.3374	0.4581	0.5197	0.4811	0.5252
satellite	6435	36	31.6395	0.3991	0.5872	0.4084	0.5716	0.5769	0.5054	0.4030	0.6866	0.5356	0.4781	0.2777	0.4595	0.4985	0.5521	0.6169
satimage-2	5803	36	1.2235	0.2455	0.9507	0.0604	0.7401	0.9277	0.3934	0.0675	0.6494	0.9507	0.8506	0.0553	0.2591	0.3782	0.8435	0.9430
vertebral	240	6	12.5000	0.0256	0.0513	0.0256	0.0000	0.0513	0.0256	0.0256	0.0000	0.0256	0.0000	0.1282	0.1025	0.0256	0.0000	0.0940
vowels	1456	12	3.4341	0.5938	0.2357	0.3309	0.0701	0.1909	0.5255	0.3121	0.0889	0.2626	0.1177	0.1177	0.2422	0.5441	0.1399	0.5545
wbc	378	30	5.5556	0.1786	0.3762	0.4238	0.5321	0.4238	0.4238	0.4238	0.3595	0.4238	0.3821	0.1083	0.3821	0.4238	0.3821	0.4599

Πίνακας 5.3: Κατάταξη των αλγορίθμων χρησιμοποιώντας το Friedman test (Using ROC)

Rank	Algorithm
2.43750	Proposed
6.43750	CBLOF
6.96875	IForest
7.00000	KNN
7.40625	MCD
7.43750	HBOS
7.62500	OCSVM
7.71875	LSCP
7.81250	AveKNN
8.59375	PCA
9.31250	SOD
9.56250	ABOD
9.56250	FB
9.62500	LOF
12.5000	SOS

Πίνακας 5.4: Post-hoc Bonferroni-Dunn (Using Proposed as control method - Using ROC)

Comparison	Statistic	Adjusted p-value	Result
Proposed vs SOS	6.36408	0.00000	H_0 is rejected
Proposed vs LOF	4.54577	0.00008	H_0 is rejected
Proposed vs ABOD	4.50625	0.00009	H_0 is rejected
Proposed vs FB	4.50625	0.00009	H_0 is rejected
Proposed vs SOD	4.34813	0.00019	H_0 is rejected
Proposed vs PCA	3.89355	0.00138	H_0 is rejected
Proposed vs AveKNN	3.39945	0.00945	H_0 is rejected
Proposed vs LSCP	3.34016	0.01172	H_0 is rejected
Proposed vs OCSVM	3.28086	0.01449	H_0 is rejected
Proposed vs HBOS	3.16228	0.02192	H_0 is rejected
Proposed vs MCD	3.14251	0.02345	H_0 is rejected
Proposed vs KNN	2.88558	0.05470	H_0 is accepted
Proposed vs IForest	2.86581	0.05823	H_0 is accepted
Proposed vs CBLOF	2.52982	0.15977	H_0 is accepted

Ο αναγνώστης μπορεί, έτσι, να ενημερωθεί για το πρόβλημα ανίχνευσης των ακραίων τιμών, ένα πρόβλημα μείζονος σημασίας για τις διεργασίες ταξινόμησης στη MM. Ακόμα, μπορεί να εξετάσει τη δυναμική των μ (ensemble methods), όπως αυτή παρουσιάζεται μέσα από την προτεινόμενη μεθοδολογία, σε ένα άκρως ανταγωνιστικό και δύσκολο πρόβλημα. Επιπλέον, μέσω του Κεφαλαίου 5, χαρτογραφείται η δυναμική της προτεινόμενης μεθόδου για μελλοντική εφαρμογή και περαιτέρω βελτίωση, μέσα από την ενσωμάτωση μη-επιτηρούμενης επιλογής χαρακτηριστικών.

Πίνακας 5.5: Κατάταξη των αλγορίθμων χρησιμοποιώντας το Friedman test (Using PRC)

Rank	Algorithm
2.50000	Proposed
6.75000	IForest
6.87500	CBLOF
7.12500	OCSVM
7.34375	KNN
7.56250	HBOS
7.68750	AveKNN
7.78125	LSCP
9.00000	FB
9.06250	SOD
9.21875	PCA
9.28125	MCD
9.37500	ABOD
9.40625	LOF
11.03125	SOS

Πίνακας 5.6: Post-hoc Bonferroni-Dunn (Using Proposed as control method - Using PRC)

Comparison	Statistic	Adjusted p-value	Result
Proposed vs SOS	5.39564	0.00000	H_0 is rejected
Proposed vs LOF	4.36790	0.00018	H_0 is rejected
Proposed vs ABOD	4.34813	0.00019	H_0 is rejected
Proposed vs MCD	4.28884	0.00025	H_0 is rejected
Proposed vs PCA	4.24931	0.00030	H_0 is rejected
Proposed vs SOD	4.15049	0.00046	H_0 is rejected
Proposed vs FB	4.11096	0.00055	H_0 is rejected
Proposed vs LSCP	3.34016	0.01172	H_0 is rejected
Proposed vs AveKNN	3.28086	0.01449	H_0 is rejected
Proposed vs HBOS	3.20181	0.01912	H_0 is rejected
Proposed vs KNN	3.06346	0.03063	H_0 is rejected
Proposed vs OCSVM	2.92511	0.04821	H_0 is rejected
Proposed vs CBLOF	2.76699	0.07921	H_0 is accepted
Proposed vs IForest	2.68794	0.10065	H_0 is accepted

Συνεργατικές μέθοδοι για προβλήματα ταξινόμησης

Η ισχύς εν τη ενώσει.

— (620-560 . . .)

Έχουν αναπτυχθεί διάφορες μέθοδοι για την αντιμετώπιση ενός προβλήματος ταξινόμησης στον τομέα κυρίως της Μηχανικής Μάθησης και των συστημάτων που καλούνται να συμβάλλουν σε μια απόφαση. Ένα υβριδικό σύστημα πρόβλεψης που συνδυάζει διάφορους ταξινομητές, αντί να επιλέγει μια μοναδική αλλά σιβαρή μέθοδο, είναι μια καλή εναλλακτική λύση για την αντιμετώπιση αυτού του προβλήματος. Προκειμένου να προσεγγίσει κανείς αυτό το ζήτημα, συναντά στη βιβλιογραφία ένα σύνολο συνεργαζόμενων κατηγοριοποιητών (classifiers) για τη δημιουργία ενός υβριδικού συστήματος υποστήριξης αποφάσεων (decision support system). Αυτή η μέθοδος βασίζεται στην παραλλαγή μιας γνωστής μεθόδου, αυτή της στοιβάξης (stacking), που συνδυάζει ισχυρές συνεργατικές μεθόδους προκειμένου να εκτελέσει μια αξιόπιστη πρόβλεψη. Στο παρόν κεφάλαιο παρουσιάζουμε μια υβριδική μέθοδο, την οποία συγκρίνουμε με γνωστές συνεργατικές μεθόδους αυτής της κατηγορίας. Τα υπολογιστικά πειράματα που διεξήχθησαν σε διάφορα τυποποιημένα σύνολα δοκιμαστικών δεδομένων έδειξαν ότι, το προτεινόμενο σχήμα δίνει πολλά υποσχόμενα αποτελέσματα, όσον αφορά την ακρίβεια, στις περισσότερες από τις περιπτώσεις. Στη συνέχεια, ως υπενθυμίσουμε ορισμένα βασικά χαρακτηριστικά των συνεργατικών μεθόδων, ώστε να γίνει αντιληπτός στον αναγνώστη, ο λόγος για τον οποίο αξίζει να επενδύσει τη μελέτη του σε αυτή την κατηγορία ταξινομητών.

Ένα σύνολο μ (ή κατηγοριοποιητών) (classifiers) [268] συνδυάζει την απόφαση μεμονωμένων ταξινομητών (με κάποιον τρόπο), προκειμένου να ταξινομηθούν νέες, άγνωστες περιπτώσεις. Έχουν παρουσιαστεί διάφοροι μέθοδοι για τη δημιουργία ενός τέτοιου συνόλου ταξινομητών [95]. Οι πιο συνηθισμένοι και ευρέως χρησιμοποιούμενοι τρόποι δημιουργίας ενός τέτοιου συνδυασμού ταξινομητών είναι οι ακόλουθοι τρεις:

- (α) Χρησιμοποιώντας μια μεμονωμένη μέθοδο μάθησης και διαφορετικά υποσύνολα δεδομένων εκπαίδευσης.
- (β) Χρησιμοποιώντας ένα μοναδικό αλγόριθμο εκπαίδευσης και διάφορες παραμέτρους εκπαίδευσης.
- (γ) Με τη χρήση διαφορετικών μεθόδων μάθησης.

Σε αυτό το σημείο, μπορεί εύλογα να αναρωτηθεί κανείς, γιατί ένα σύνολο ταξινομητών μπορεί να προσφέρει καλύτερα αποτελέσματα από ένα μεμονωμένο ταξινομητή. Οι λόγοι είναι , και και παρουσιάζονται αναλυτικά στο άρθρο [95]. Τα δεδομένα εκπαίδευσης δεν μπορούν να παράγουν πληροφοριακές γνώσεις σε ορισμένες περιπτώσεις, προκειμένου να επιλεγεί ένας μόνον ταξινομητής. Αυτό μπορεί

να είναι μια πιθανή αιτία του προαναφερθέντος προβλήματος δεδομένου ότι, το σύνολο των δεδομένων εκπαίδευσης είναι πολύ μικρό σχετικά με το μέγεθος του χώρου υποθέσεων.

Παρά το γεγονός ότι το ενδιαφέρον που έχει συγκεντρωθεί για τις μ (ensemble methods) τα τελευταία χρόνια είναι πολύ μεγάλο, απαιτείται περαιτέρω μελέτη, καθώς δεν είναι σαφές ποια μέθοδος είναι η καλύτερη. Στο παρόν κεφάλαιο παρουσιάζουμε ένα υβριδικό σχήμα, το οποίο συνδυάζει τις ακόλουθες συνεργατικές μεθόδους ταξινόμησης: (α) τον αλγόριθμο των (Random Forest) [46], τη μέθοδο των (ExtraTrees) [153] και τον αλγόριθμο Gradient Boosting [139], χρησιμοποιώντας μια παραλλαγή της μεθοδολογίας μ (stacking). Το προτεινόμενο υβριδικό σχήμα αποφάσεων συγκρίνεται με διάφορες συνεργατικές μεθόδους σε μια ποικιλία διαφορετικών δοκιμαστικών δεδομένων. Τα αποτελέσματα που προέκυψαν έδειξαν ότι η νέα αυτή μέθοδος έχει καλύτερη ακρίβεια στα περισσότερα από τα προβλήματα.

Η ιδέα για τη δημιουργία μιας νέας συνεργατικής μεθόδου προήλθε από τα αποτελέσματα τα οποία αναλύονται στο Κεφάλαιο 2. Δεδομένης της μη ύπαρξης μίας και μοναδικής μεθόδου που θα ανταποκρίνεται καλύτερα από όλες την υπόλοιπες της ίδιας κατηγορίας για το σύνολο όλων των προβλημάτων, π.χ. ταξινόμησης, μάς δίνεται η δυνατότητα να δημιουργήσουμε εμπειρικά μοντέλα μάθησης για την αντιμετώπιση προβλημάτων ταξινόμησης. Αυτή είναι η εύληπτη ερμηνεία της προτίμησης μιας συνεργατικής μεθόδου έναντι ενός και μόνο ισχυρού ταξινομητή. Έτσι, δημιουργώντας ένα σύνολο από ισχυρούς και ευρέως δοκιμασμένους ταξινομητές, καταφέραμε να επιτύχουμε καλύτερη ακρίβεια ταξινόμησης στις περισσότερες από τις περιπτώσεις που ανατρέξαμε.

Το υπόλοιπο του Κεφαλαίου 6 είναι οργανωμένο ως εξής. Στην επόμενη ενότητα παρουσιάζουμε γνωστές μεθόδους για τη δημιουργία συνεργατικών μεθόδων, ενώ στην Ενότητα 6.2 εξετάζουμε την προτεινόμενη υβριδική μεθοδολογία σε διαφορετικά σύνολα δεδομένων. Τα διεξαχθέντα υπολογιστικά πειράματα και οι συγκρίσεις με άλλες συνδυαστικές μεθόδους παρουσιάζονται στην Ενότητα 6.3. Η μελέτη μας ολοκληρώνεται στην Ενότητα 6.4, με μια σύνοψη και περαιτέρω ερευνητικές παρατηρήσεις.

6.1 Γνωστές τεχνικές σχεδιασμού συνεργατικών μεθόδων

Υπάρχει μια πληθώρα μεθόδων που χρησιμοποιούν ένα μόνο ταξινομητή για την επίλυση ενός προβλήματος πρόβλεψης. Ωστόσο, η ανάγκη για μοντέλα που θα κάνουν μια ακριβέστερη πρόβλεψη είναι έντονη. Αυτός ο σκοπός, όπως αναφέρεται στην εισαγωγή του κεφαλαίου, εξυπηρετείται από ένα συνδυασμό ταξινομητών. Προκειμένου να δημιουργηθεί ένα καλό σύνολο ταξινομητών, υπάρχουν διάφορες μεθοδολογίες. Καθεμία από αυτές μπορεί να παράσχει ένα διαφορετικό ποσοστό ορθώς ταξινομημένων περιπτώσεων. Έτσι, ο γενικός στόχος των μεθόδων, που ανήκουν σε συστήματα υποστήριξης αποφάσεων, είναι να επιτύχουμε αξιόπιστα και πολύ ακριβή αποτελέσματα πρόβλεψης. Αυτή η ενότητα παρουσιάζει ευρέως χρησιμοποιούμενες μεθόδους για τη δημιουργία μ (ensemble methods).

Η μ (bagging technique) ή η μ (bootstrap aggregating) [45] είναι μία από τις πιο γνωστές και ευρέως χρησιμοποιούμενες τεχνικές για τη δημιουργία συνεργατικών μεθόδων. Ένα βασικό χαρακτηριστικό αυτής της μεθόδου είναι η εκπαίδευση κάθε «μαθητή» (learner) χρησιμοποιώντας διαφορετικά σύνολα εκπαίδευσης. Συγκεκριμένα, έστω N , ένα σύνολο δεδομένων εκπαίδευσης μεγέθους n . Η μέθοδος bagging επιλέγει από αυτό το σύνολο n^l τυχαία παραδείγματα, δημιουργώντας ένα σύνολο υποσύνολων εκπαίδευσης, έστω N_l . Η επιλογή των περιπτώσεων γίνεται με τη χρήση μιας κατανομής, για παράδειγμα της ομοιόμορφης κατανομής, ενώ τα παραδείγματα επιλέγονται με αντικατάσταση. Δεδομένου ότι η κατασκευή των υποσυνόλων γίνεται με αυτόν τον τρόπο, είναι δυνατόν ορισμένα από τα παραδείγματα να επαναληφθούν στο σύνολο N_l και έτσι, μερικά παραδείγματα να είναι «διπλά» ή κάποια άλλα να παραλείπονται, σε σχέση με το αρχικό

σύνολο εκπαίδευσης N . Αυτή η διαδικασία παράγει ένα σύνολο ταξινομητών. Στη συνέχεια, διεξάγεται μια διαδικασία ψηφοφορίας σε σχέση με τις προβλέψεις κάθε κατηγοριοποιητή. Το αποτέλεσμα της ψηφοφορίας δίνει την τελική απόφαση του μοντέλου.

Ένας πολύ γνωστός μ - μ (meta-algorithm) είναι η (Boosting approach) [138]. Θα μπορούσαμε να πούμε ότι αυτή η μέθοδος μοιάζει με την προσέγγιση bagging, καθώς διάφορα υποσύνολα χρησιμοποιούνται για την εκπαίδευση ενός μόνο εκπαιδευόμενου (learner). Επιπλέον, αυτός ο αλγόριθμος βασίστηκε στο ερώτημα « μ μ μ μ ; » [210]. Η μέθοδος boosting σχετίζεται άμεσα με αυτή την ερώτηση, καθώς επικεντρώνεται στην απόδοση ενός αδύναμου ταξινομητή, προκειμένου να τον ενισχύσει, ειδικά σε περιπτώσεις όπου τα παραδείγματα δεν έχουν «δίδαχθεί» επαρκώς. Έτσι, αντί να επιλέγουμε τυχαία τις περιπτώσεις με βάση μια κατανομή, ο αλγόριθμος επικεντρώνεται στα παραδείγματα εκπαίδευσης που δεν έχουν «αφομοιωθεί» σωστά. Προκειμένου να δημιουργηθεί ένας ισχυρός ταξινομητής, μετά από μερικά βήματα, η πρόβλεψη γίνεται μέσω μιας σταθμισμένης διαδικασίας ψήφου (weighted voting). Αναλυτικά, κάθε πρόβλεψη ταξινομητή παίρνει ένα βάρος σύμφωνα με την ακρίβεια ταξινόμησης που πέτυχε.

Εκτός από τις προαναφερθείσες τεχνικές, που κυρίως εκπαιδεύουν τους ταξινομητές σε ένα υποσύνολο δεδομένων εκπαίδευσης, υπάρχουν μέθοδοι που δημιουργούν ένα μοντέλο, χρησιμοποιώντας ένα σύνολο αλγορίθμων μάθησης με βάση ολόκληρο το σύνολο δεδομένων εκπαίδευσης. Στη συνέχεια, επιλέγοντας ένα κατάλληλο μοτίβο ψηφοφορίας, συνδυάζουν τις απαντήσεις των ταξινομητών για να λάβουν την τελική πρόβλεψη. Μέσω αυτής της διαδικασίας, χρησιμοποιούνται διαφορετικοί αλγόριθμοι μάθησης για την αύξηση της ποικιλομορφίας των σφαλμάτων πρόβλεψης των μοντέλων. Αυτό οφείλεται στο γεγονός ότι διαφέρουν τόσο στη διαδικασία αναζήτησης όσο και στην αναπαράσταση. Συμπερασματικά, τα μοντέλα που δημιουργούνται μέσω της παραπάνω διαδικασίας είναι πιο πιθανό να κάνουν λάθη σε διαφορετικά σημεία, καθώς χρησιμοποιούνται διαφορετικές μεροληψίες μάθησης (learning biases). Τέλος, χρησιμοποιώντας την πλειονότητα των προβλέψεων, καταλήγουμε εύκολα στην τελική απόφαση, καθώς δεν απαιτείται προηγούμενη εκπαίδευση.

Μια πολύ γνωστή μέθοδος που χρησιμοποιεί την προαναφερθείσα διαδικασία είναι η μ - μ (stacked generalization approach) ή απλά μ (stacking) [419]. Σε αυτή τη μεθοδολογία, ο κύριος στόχος είναι η παραγωγή ενός ισχυρού, υψηλού επιπέδου εκπαιδευόμενου με υψηλή απόδοση γενίκευσης. Για να επιτευχθεί αυτό, συνδυάζεται ένα σύνολο διαφόρων κατηγοριοποιητών. Για να επιτευχθεί ο κατάλληλος συνδυασμός των μ (base predictions), χρησιμοποιείται ένας αλγόριθμος μάθησης. Προκειμένου να εκτελεστεί η τελική πρόβλεψη, το μοντέλο χρησιμοποιεί τα ακόλουθα βήματα: (α) μ (Level zero data), όπου όλοι οι μ (base learners) τρέχουν στο αρχικό σύνολο δεδομένων, (β) μ (Level one data). Μετά το επίπεδο μηδέν, οι προβλέψεις των ταξινομητών θεωρούνται ως νέα δεδομένα και (γ) μ (Final prediction). Μια άλλη διαδικασία μάθησης εκτελείται με τη χρήση των δεδομένων του πρώτου επιπέδου ως νέες εισοδοί και ως έξοδος, και λαμβάνεται η τελική πρόβλεψη.

Μια εναλλακτική ιδέα χρησιμοποιήθηκε στο άρθρο [383]. Η προτεινόμενη μέθοδος είναι μία επέκταση της γνωστής μεθόδου stacking [419] και βασίζεται στις μ (class probability distributions) και όχι στις προβλέψεις. Έτσι, κάθε μαθητής δεν παρουσιάζει μια τελική πρόβλεψη για την κλάση στην οποία ανήκει ένα παράδειγμα. Ως εκ τούτου, δίνει μια εκτιμώμενη πιθανότητα για όλες τις κατηγορίες. Οι συγγραφείς του άρθρου, προκειμένου να επιτύχουν αξιόπιστη μάθηση στο μετα-επίπεδο (meta-level learning), χρησιμοποίησαν τη μέθοδο $\mu\mu$ μ (Multi-response Linear Regression (MLR)).

Μια σειρά αρκετών προσεγγίσεων έχουν προταθεί για τη βελτίωση της απόδοσης της μεθόδου που αναπτύχθηκε στο άρθρο [383]. Συγκεκριμένα, στο άρθρο [109], υιοθετήθηκε

η ίδια στρατηγική εκτός από τη διαδικασία μ - μ (meta-learning). Οι συγγραφείς του άρθρου, χρησιμοποίησαν ως μοντέλο ένα δέντρο (model tree induction) αντί για την προσέγγιση MLR και τα αποτελέσματα που προέκυψαν, έδειξαν ότι επιτυγχάνεται καλύτερη απόδοση από το προτεινόμενο μοντέλο. Μια άλλη τροποποίηση προτείνεται στο άρθρο [344]. Σε αυτήν τη μέθοδο, αντί για τις συνολικές κατανομές πιθανότητας κλάσης (overall class probability distributions), λαμβάνονται υπόψη μόνο οι πιθανότητες κλάσης που αφορούν την αληθή κλάση. Τα πειραματικά αποτελέσματα έδειξαν ότι μέσω αυτής της παραλλαγής η ακρίβεια της μεθόδου μοιάζει να βελτιώνεται αισθητά.

Η σημαντικότητα της μεθόδου stacking φανερώνεται μέσα από το πλήθος των παραλλαγών που συναντώνται στη βιβλιογραφία. Ακόμα, αρκετοί ερευνητές χρησιμοποίησαν αυτή την μεθοδολογία για να βελτιώσουν άλλες υπάρχουσες μεθόδους ή προκειμένου να δημιουργήσουν νέες συνεργατικές μεθόδους. Στο άρθρο [259], παρουσιάζεται ένας άλλος αλγόριθμος, που ονομάζεται Τροϊκα, προκειμένου να αντιμετωπιστούν προβλήματα ταξινόμησης πολλαπλών κατηγοριών. Συγκεκριμένα, η προτεινόμενη μέθοδος, αντί ενός μετα-μαθητή χρησιμοποιεί τρία επίπεδα συνδυασμού ταξινομητών: (α) Το μηδενικό επίπεδο αποτελείται από τους βασικούς ταξινομητές, (β) Στο πρώτο επίπεδο συναντώνται «ειδικοί» ταξινομητές, (γ) Το δεύτερο επίπεδο αποτελείται από μετα-ταξινομητές και (δ) Στο τελευταίο επίπεδο λαμβάνει χώρα ο υπερ-ταξινομητής (super-classifier), που δέχεται τις προβλέψεις του προηγούμενου επιπέδου για να κάνει την τελική πρόβλεψη του μοντέλου. Η κύρια σύγκριση περιελάμβανε τη μέθοδο Τροϊκα, το βασικό σχήμα stacking και τη βελτίωση της τυπικής μεθόδου στοιβαξης, τη μέθοδο stackingC. Τα πειράματα που διεξήχθησαν έδειξαν ότι η προτεινόμενη μέθοδος υπερέχει των άλλων δύο μεθόδων, όσον αφορά την ακρίβεια, ανεξάρτητα από τη μέθοδο δυαδικής ταξινόμησης που προτιμάται στο βασικό επίπεδο (base-level classifiers).

Η μεθοδολογία της συνεργατικής μεθόδου στοιβαξης μπορεί να θεωρηθεί ως ένα συνδυαστικό πρόβλημα βελτιστοποίησης. Αυτό οφείλεται στο γεγονός ότι, τόσο οι ταξινομητές βάσης όσο και ο μετα-ταξινομητής, πρέπει να προσδιοριστούν με βέλτιστο τρόπο. Στο άρθρο [66], υιοθετήθηκε μια προσαρμοστική μεταερευνητική μέθοδος αναζήτησης, προκειμένου να δημιουργηθεί μια νέα προσέγγιση στοιβαξης, που ονομάζεται μ - μ -Stacking (Ant Colony Optimization (ACO) Stacking algorithm). Έτσι, δεδομένου ενός συνόλου μαθητών βάσης, έστω B , το οποίο αποτελείται από n ταξινομητές, ο αλγόριθμος ACO [101] καθορίζει ένα υποσύνολο του B , που αποτελείται από ένα μικρότερο αριθμό μαθητών. Αυτό το σύνολο ταξινομητών αναμένεται να έχει καλύτερη ακρίβεια ταξινόμησης από το πρωτότυπο. Οι συγγραφείς του άρθρου συνέκριναν την προσέγγιση ACO-Stacking με γνωστές συνεργατικές μεθόδους, όπως αυτές της Bagging, AdaBoost, StackingC και Random Forest. Τα πειραματικά αποτελέσματα έδειξαν ότι η απόδοση του αλγορίθμου ACO-Stacking είναι πολλά υποσχόμενη.

Το παραπάνω πρόβλημα, δεν είναι εύκολο να διαχειριστεί, καθώς η βέλτιστη επιλογή μεθόδων μπορεί να αποδειχθεί πολύ επίπονη διαδικασία. Για το σκοπό αυτό, έχουν αναπτυχθεί ερευνητικές προσπάθειες, παρόμοιες με την προηγούμενη προσέγγιση, στις οποίες το πρόβλημα της επιλογής της κατάλληλης διαμόρφωσης των βασικών μαθητών και του μετα-ταξινομητή είναι ο κύριος άξονας μελέτης [351]. Η κύρια διαφορά που σχετίζεται με το άρθρο [66], αφορά τον αλγόριθμο ευφυούς αναζήτησης σμήνους, που χρησιμοποίησε για τη διαδικασία της βελτιστοποίησης. Συγκεκριμένα, οι συγγραφείς του άρθρου χρησιμοποίησαν το γνωστό αλγόριθμο μ - μ (Artificial Bee Colony (ABC)) [207] για να κατασκευάσουν δύο διαφορετικούς τύπους στοιβαξης. Έτσι, η βέλτιστη διαμόρφωση των βασικών μαθητών λαμβάνει χώρα στην αρχή της διαδικασίας. Σε αυτή τη φάση, ο μετα-ταξινομητής που εφαρμόζεται, θεωρείται ότι είναι ένας σταθερός αλγόριθμος μάθησης. Κατά τη διάρκεια της μετα-φάσης (meta-phase), ο αλγόριθμος ABC βελτιστοποιεί ταυτόχρονα τους βασικούς μαθητές και τον μετα-ταξινομητή. Η προσέγγιση ABC-Stacking δοκιμάστηκε σε αρκετά γνωστά σύνολα δεδομένων και συγκρίθηκε με διαφορετικούς προγνωστικούς αλγορίθμους, και βασικές συνεργατικές μεθόδους, όπως τη μέθοδο GA-Stacking [279] και

ACO-Stacking. Τα υπολογιστικά πειράματα που διενεργήθηκαν από τους συγγραφείς του άρθρου έδειξαν ότι, η προτεινόμενη προσέγγιση έχει αντιμετωπίσει αποτελεσματικά από κοινού τη διαμόρφωση των ταξινομητών βάσης και την επιλογή του μετα-ταξινομητή. Επιπλέον, η απόδοση του αλγορίθμου ABC-Stacking είναι συγκρίσιμη με τα καλύτερα αποτελέσματα που αναφέρονται μέχρι τώρα.

Στην εργασία [315] υιοθετείται μια στρατηγική που είναι παρόμοια με τη μεθοδολογία stacking, προκειμένου να εφαρμοστεί ένας αλγόριθμος διαδοχικής μάθησης σε προβλήματα πολλαπλών κλάσεων. Για μια εκτενέστερη μελέτη της μεθοδολογίας stacking, των συνεργατικών μεθόδων στοιβαξης και σχετικά με τις προσεγγίσεις stacking, που αναπτύχθηκαν τα τελευταία 20 χρόνια, ο αναγνώστης παραπέμπεται στο άρθρο [347].

6.2 Προτεινόμενο υβριδικό σχήμα ταξινόμησης

Γενικά, όπως και με κάθε ταξινομητή, η προσθήκη νέων χαρακτηριστικών εισόδου μπορεί να βελτιώσει την ακρίβεια ταξινόμησης, όταν τα νέα χαρακτηριστικά περιέχουν νέες πληροφορίες σχετικά με την κλάση. Αυτή η βελτίωση της απόδοσης βέβαια, δεν είναι εγγυημένη, διότι οι ταξινομητές είναι «ατελείς» και ενδέχεται να μην είναι σε θέση να αξιοποιήσουν αυτές τις πληροφορίες. Εάν τα νέα χαρακτηριστικά μοιράζονται πληροφορίες με τα υπάρχοντα χαρακτηριστικά του συνόλου δεδομένων, οι νέες πληροφορίες ενδέχεται να μην βοηθήσουν. Σε όλες τις περιπτώσεις, πρέπει να ληφθεί υπόψη ένα σημαντικό ζήτημα, αυτό της

μ (curse of dimensionality). Η παρουσία περισσότερων χαρακτηριστικών μπορεί να βλάψει πολλούς ταξινομητές και μπορεί να αυξήσει τη δυνατότητα για το φαινόμενο του επερταιριάσματος (overfitting). Η κατάσταση είναι παρόμοια όταν προσθέτουμε νέους ταξινομητές βάσης σε μια μεθοδολογία , καθώς οι έξοδοι των ταξινομητών βάσης μπορούν να ιδωθούν ως νέα χαρακτηριστικά για τον τελικό ταξινομητή. Συνεπώς, όλα τα παραπάνω επιχειρήματα ισχύουν και σε αυτή την περίπτωση. Έτσι, αυτά τα χαρακτηριστικά του «δεύτερου επιπέδου» συσχετίζονται πιθανώς, επειδή όλοι οι ταξινομητές βάσης προσπαθούν να προβλέψουν την ίδια περίπτωση. Ωστόσο, εκτελούν αυτή τη διαδικασία με έναν τρόπο που χαρακτηρίζεται ως μη βέλτιστος. Η ελπίδα για ένα καλύτερο αποτέλεσμα στηρίζεται στο ότι συμπεριφέρονται με διαφορετικό τρόπο, και έτσι, ο τελικός ταξινομητής δύναται να συνδυάσει τις θορυβώδεις προβλέψεις σε μια καλύτερη τελική πρόβλεψη. Ακολουθώντας, η προσθήκη νέων ταξινομητών βάσης εκτιμάται ότι θα έχει καλύτερη πιθανότητα να βοηθήσει στο σχήμα, όταν κάνει καλή δουλειά και συμπεριφέρεται διαφορετικά από τους υπάρχοντες ταξινομητές βάσης, αλλά αυτό και πάλι δεν είναι εγγυημένο. Σε όλες τις περιπτώσεις, πρέπει να εξεταστεί ο χρόνος εκπαίδευσης. Στην περίπτωσή μας, επιλέγουμε να χρησιμοποιήσουμε στο υβριδικό σχήμα τρεις ισχυρές συνεργατικές μεθόδους ως μαθητές βάσης και συγκεκριμένα: τον αλγόριθμο Random Forest, τη μέθοδο ExtraTrees και Gradient Boosting.

Σύμφωνα με τον ορισμό του Brieman [46], ο αλγόριθμος random forest είναι ένας ταξινομητής, ο οποίος αποτελείται από ένα σύνολο ταξινομητών και στηρίζεται στις δενδρικές δομές. Ένα σημαντικό στοιχείο είναι ότι σε κάθε δέντρο, οι αποφάσεις γίνονται με τυχαίο τρόπο κατά το διαχωρισμό ενός κόμβου, χωρίς να λαμβάνονται υπόψη όλα τα χαρακτηριστικά του συνόλου. Σε αυτή τη διαδικασία, κάθε δέντρο εκπαιδεύεται σε ένα υπο-σύνολο με παρόμοιο τρόπο με αυτό της μεθόδου bagging και οι ψήφοι για την πιο δημοφιλή κατηγορία λαμβάνονται ως νέοι είσοδοι. Αυτό το είδος των δέντρων απόφασης που χρησιμοποιεί τη μέση τιμή, μπορεί να βελτιώσει την ακρίβεια της πρόβλεψης. Επιπλέον, μπορούν να ελέγξουν και να περιορίσουν το φαινόμενο της υπερπροσαρμογής. Όπως και στην τεχνική Bootstrap aggregating, η δημιουργία των υπο-δειγμάτων γίνεται με αντικατάσταση και το μέγεθος κάθε υπο-δείγματος είναι ίσο με το αρχικό δείγμα.

Μια παρόμοια μέθοδος με αυτή των τυχαίων δασών είναι τα (Extremely randomized Trees ή ExtraTrees) [153]. Αυτή η μέθοδος εφαρμόστηκε με επιτυχία και παράγει καλά αποτελέσματα όσον αφορά την ακρίβεια, ενώ παράλληλα ελέγχει

αποτελεσματικά το πρόβλημα της υπερπροσαρμογής. Η διαφορά με τα μ μπορεί να βρεθεί σε δύο σημεία: (α) Η εκπαίδευση κάθε δέντρου γίνεται χωρίς τη χρήση ενός μ (bootstrap sample), αλλά με ένα ολόκληρο-ακοιμμάτιστο σύνολο εκπαίδευσης και (β) το σημείο «κοπής» (cut-point) για τον διαχωρισμό ενός κόμβου επιλέγεται με έναν πλήρως τυχαίο τρόπο.

Η μ (gradient boosting) [139] είναι μια ευρέως χρησιμοποιούμενη προσέγγιση για την προσέγγιση των προβλημάτων παλινδρόμησης και ταξινόμησης, η οποία βασίζεται κυρίως σε ταξινομητές μ . Η κύρια στρατηγική αυτής της μεθόδου είναι παρόμοια με άλλες μ (boosting methods) (όπως οι μέθοδοι AdaBoost ή LogitBoost). Σε μια πρώτη φάση, χτίζεται ένα σύνολο αδύναμων εκπαιδευομένων (learner) για να πραγματοποιηθεί μια προκαταρκτική πρόβλεψη. Σε μια δεύτερη φάση, βελτιστοποιείται η ακρίβεια πρόβλεψης του μοντέλου με βάση την ελαχιστοποίηση μιας συνάρτησης κόστους. Συγκεκριμένα, η μέθοδος gradient boosting καθορίζει επαναληπτικά μια συνάρτηση που «δείχνει» στην κατεύθυνση της κλίσης με αρνητικό πρόσημο.

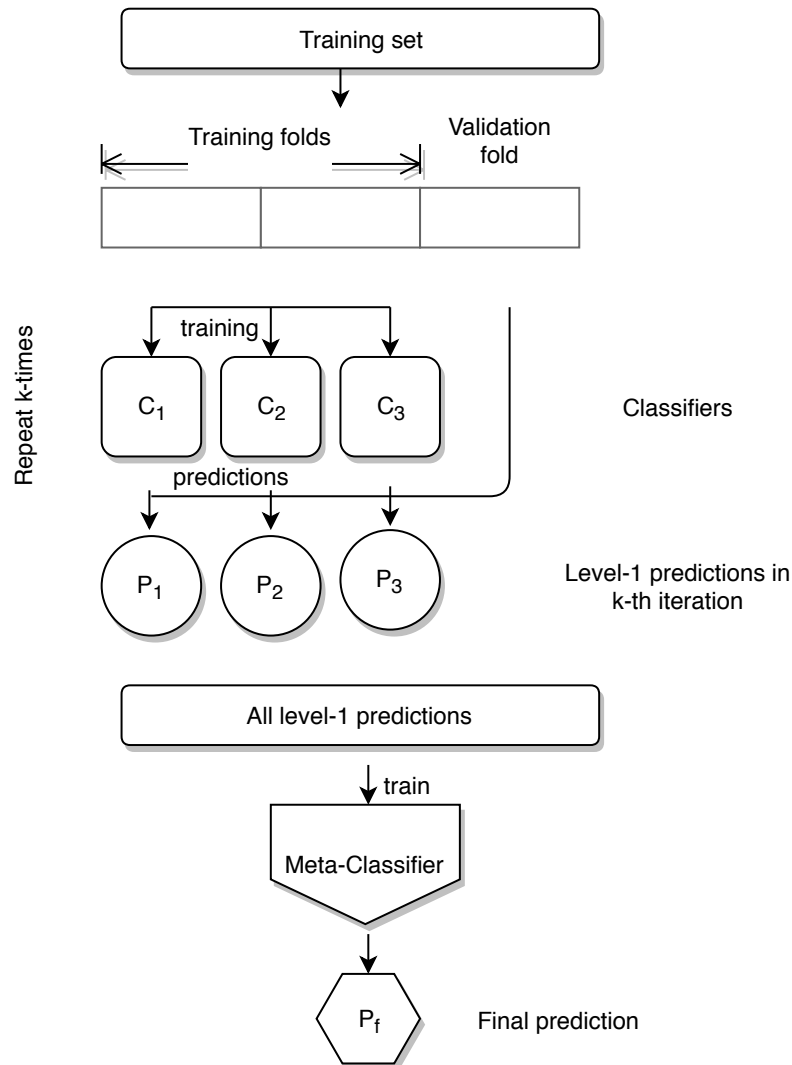
Στον προτεινόμενο αλγόριθμο μ , το σύνολο των μετα-δεδομένων θα αποτελείται από τις προβλέψεις που δίδονται από τις μ μ και την αληθή κλάση για κάθε περίπτωση εκπαίδευσης. Έτσι, ο ταξινομητής που εκπαιδεύεται σε αυτό το σύνολο δεδομένων είναι ο λεγόμενος μ - μ (meta-learner). Η προσέγγισή μας χρησιμοποιεί τις πιθανότητες κλάσης που συσχετίζονται με την αληθή κλάση και όχι με όλες τις πιθανότητες των κατηγοριών, όπως στην μ .

Ένα άλλο σημείο που αξίζει να αναφέρουμε είναι η διαδικασία μάθησης. Ο στόχος είναι να δημιουργηθεί μια διαδικασία μάθησης που θα είναι γρήγορη. Αυτό επιτυγχάνεται με τη μείωση της διασπασιμότητας των μετα-δεδομένων που καθορίζεται από έναν παράγοντα ο οποίος σχετίζεται με τον αριθμό των κλάσεων. Παρά τη μείωση αυτή, η ακρίβεια για τα προβλήματα δύο κατηγοριών παραμένει ανεπηρέαστη. Επιπλέον, ο αλγόριθμος μάθησης, που χρησιμοποιούμε στο μετα-επίπεδο, είναι ο γνωστός ταξινομητής logistic tree [376]. Αυτό οφείλεται στο γεγονός ότι αυτά τα μοντέλα επιτυγχάνονται με τη βοήθεια τετραγωνικών γραμμικών τμηματικών προσεγγίσεων (piecewise linear approximations). Το Σχήμα 6.1, απεικονίζει τη διαδικασία που υιοθετήθηκε για να δημιουργηθεί ένας ισχυρός ταξινομητής, ενώ η προτεινόμενη μέθοδος απεικονίζεται στον Αλγόριθμο 2.

6.3 Συγκρίσεις και πειραματικά αποτελέσματα

Για τα υπολογιστικά μας πειράματα, χρησιμοποιήθηκαν γνωστά σύνολα δεδομένων από το UCI Machine Learning repository [104]. Συγκεκριμένα, επελέγησαν 38 προβλήματα από διαφορετικά πεδία, τα οποία διαφέρουν τόσο στον αριθμό των χαρακτηριστικών τους όσο και στις κλάσεις τους. Στον Πίνακα 6.1, υπάρχει μια σύντομη περιγραφή αυτών των συνόλων δεδομένων, όπως ο αριθμός των χαρακτηριστικών, ο αριθμός των παραδειγμάτων και ο αριθμός των κλάσεων.

Η ακρίβεια των ταξινομητών αξιολογήθηκε σύμφωνα με την ακόλουθη διαδικασία: Το σύνολο δεδομένων εκπαίδευσης χωρίστηκε σε δέκα υποσύνολα ίσου μεγέθους, και για καθένα από αυτά ο ταξινομητής εκπαιδεύτηκε στα υπόλοιπα υποσύνολα. Στη συνέχεια, για κάθε μέθοδο, εκτελούμε την πενταπλή διαδικασία διασταυρούμενης επικύρωσης (five fold cross-validation) και μετράται η μέση τιμή πενταπλάσιας διασταυρούμενης επικύρωσης. Τα πειράματα έχουν διεξαχθεί με την γλώσσα προγραμματισμού Python χρησιμοποιώντας τις διαθέσιμες υλοποιήσεις από τη βιβλιοθήκη scikit-learn [300] και MLxtend [322]. Στον Πίνακα 6.2 παρουσιάζονται τα αποτελέσματα που ελήφθησαν για το υβριδικό μας σχήμα και για τις μεθόδους που συμμετέχουν στη σύγκριση. Το βέλτιστο σχήμα για κάθε σύνολο δεδομένων περιγράφεται με γράμματα με έντονη γραφή. Μπορούμε εύκολα να δούμε ότι η προτεινόμενη συνδυαστική μέθοδος υπερβαίνει τις υπόλοιπες τρεις συνδυαστικές μεθόδους στις περισσότερες περιπτώσεις και έτσι, μπορούμε να ισχυριστούμε ότι η προτεινόμενη μέθο-



Σχήμα 6.1: Διαδικασία δημιουργίας συνεργατικής μεθόδου με βάση τη μεθοδολογία στείβαξης

δος είναι πιο ισχυρή.

Επιπλέον, έχουν διεξαχθεί στατιστικές δοκιμές προκειμένου να ελεγχθεί η σημασία των αποτελεσμάτων. Συγκεκριμένα, λόγω του μικρού αριθμού μεθόδων σύγκρισης, διεξήχθη η μη παραμετρική δοκιμή Friedman [185]. Επομένως, στον Πίνακα 6.3 παρουσιάζεται η κατάταξη των δοκιμών Friedman. Επιπλέον, η p value σε όλες τις συγκρίσεις δείχνει ότι οι μηδενικές υποθέσεις πρέπει να απορριφθούν. Έτσι, υπάρχουν μέθοδοι των οποίων η διαφορά απόδοσης ήταν στατιστικά σημαντική για τους άλλους. Στον Πίνακα 6.4 παρουσιάζονται τα αποτελέσματα της post-hoc Holm δοκιμής [187].

Πίνακας 6.3: Κατάταξη των αλγορίθμων χρησιμοποιώντας το Friedman test

Stacking	1.48684
Gradient Boosting	2.05263
Random Forest	2.51316
ExtraTree	3.94737

Input Training data $S = \{x_i; y_i\}_{i=1}^n$, $x_i \in \mathbb{R}^n$; $y_i \in \mathcal{C}$, where \mathcal{C} denotes the classes.

Output A stacking meta-classifier ensemble P_f .

- 1 :** Adopt cross validation approach in preparing a training set for second-level Logistic Model Tree classifier.
Split the starting dataset into $k(= 3)$ equal-size subsets $S_1; S_2; S_3$ in a random way.
for $j = 1$ to k **do**
 Learn first-level classifiers
 for $t = 1$ to 3 **do**
 Train a learner l_{kt} from $S \cap S_k$
 endfor
 Construct a training set for second-level model tree classifier
 for $x_i \in S_k$ **do**
 Get a record $\{x_i^j; y_i\}$, where $x_i^j = \{p_{k1}(x_i); p_{k2}(x_i); p_{k3}(x_i)\}$
 endfor
endfor
- 2 :** Learn a second-level classifier
 Learn a new classifier p^j from the collection of $\{x_i^j; y_i\}$
- 3 :** Re-learn first-level learners
 for $t = 1$ to 3 **do**
 Learn a classifier p_t based on S
 endfor
Return $P_f(x)$

Αλγόριθμος 2. Stacking with k-fold cross validation

Πίνακας 6.4: Post-hoc Holm test using Stacking as control method

Comparison	Statistic	Adjusted p -value	Result
Stacking vs ExtraTree	8.30769	0.00000	H_0 is rejected
Stacking vs Random Forest	3.46524	0.00106	H_0 is rejected
Stacking vs Gradient Boosting	1.91033	0.04609	H_0 is rejected

6.4 Σύνοψη

Τα συστήματα υβριδικών αποφάσεων φαίνεται να είναι μια αξιόπιστη και αποτελεσματική επιλογή προκειμένου να αντιμετωπιστούν τα προβλήματα λήψης αποφάσεων σε διάφορους επιστημονικούς τομείς. Σε σύγκριση με την ακρίβεια που μπορεί να επιτύχει ένας μόνο ταξινομητής, μια ή μέθοδος φαίνεται να είναι σε θέση να επιτύχει καλύτερα αποτελέσματα ταξινόμησης. Ωστόσο, ο σχεδιασμός μιας τέτοιας μεθόδου είναι ένα δύσκολο πρόβλημα, καθώς έχουμε μια πληθώρα βασικών ταξινομητών για χρήση. Η μεθοδολογία είναι μια κατάλληλη και ακριβής διαδικασία, προκειμένου να δημιουργηθεί ένας ισχυρός συνδυασμός με τους καλύτερους ταξινομητές.

Παρά το γεγονός ότι έχουν σχεδιαστεί ποικίλες συνεργατικές μέθοδοι, η δημιουργία του καλύτερου συνδυασμού για την επίλυση συγκεκριμένων προβλημάτων ταξινόμησης παραμένει ένα φλέγον και εν εξελίξει έργο. Σε αυτό το κεφάλαιο, προτείνουμε μια ισχυρή μεθοδολογία στείβαξης με τη χρήση του αλγορίθμου Logistic Model Trees και τριών πολύ γνωστών και δοκιμασμένων συνεργατικών μεθόδων. Τα πειραματικά αποτελέσματα που διε-

Πίνακας 6.1: Συλλογή 38 προβλημάτων πολλαπλών κλάσεων από το UCI Machine Learning Repository. Στον πίνακα παρουσιάζεται ο αριθμός των παραδειγμάτων, ο αριθμός των χαρακτηριστικών, καθώς και ο αριθμός των κλάσεων.

Dataset	#Instances	#Features	#Classes
audiology	226	69	24
autos	205	25	7
badges	294	11	2
balance-scale	625	4	3
breast-cancer	286	9	2
wisconsin-breast-cancer	699	9	2
horse-colic	368	28	2
credit-rating	690	15	2
german-credit	1000	20	2
pima-diabetes	768	8	2
glass	214	9	6
grub-damage	100	11	2
haberman	306	3	2
hear-cleveland	303	13	5
heart-hungarian	294	12	2
heart-statlog	270	13	2
hepatitis	155	19	2
hypothyroid	3772	30	2
ionosphere	351	34	2
iris	150	4	3
labor	57	16	2
lymphography	148	18	4
monk1	556	7	2
monk2	601	7	2
monk3	554	7	2
mushroom	8124	22	2
primary-tumor	339	17	21
students	344	11	2
sick	3772	29	2
sonar	208	60	2
soybean	683	35	19
relation	2201	3	2
vehicle	846	18	4
vote	435	16	2
vowel	990	13	11
waveform	5000	40	3
wine	179	13	3
zoo	101	17	7

ξήχθησαν σε γνωστά σύνολα δεδομένων έδειξαν ότι το προτεινόμενο υβριδικό σχήμα δίνει πολλά υποσχόμενα αποτελέσματα στα περισσότερα από τα προβλήματα ταξινόμησης.

Παρόλα αυτά, υπάρχουν αρκετά ζητήματα που χρειάζονται περαιτέρω μελέτη, όπως οι κανόνες που δημιουργούν ένα κατάλληλο συνεργατικό σχήμα ταξινομητών με την υψηλότερη δυνατή ακρίβεια. Επιπλέον, οι περαιτέρω συγκρίσεις της μεθόδου μας με γνωστές παραλλαγές του αλγόριθμου Stacking όπως οι [66, 279] και η εφαρμογή σε προβλήματα παλινδρόμησης θα μπορούσαν να είναι ένα ενδιαφέρον έργο για μελλοντική έρευνα. Ακόμα, η δυσκολότερη, ίσως, και πιο ενδιαφέρουσα εργασία, που παραμένει ένα ανοιχτό πρόβλημα, είναι η βέλτιστη επιλογή ταξινομητών για τη συμμετοχή στο συνεργατικό σύνολο ταξινομητών. Όπως αναφέραμε και στην εισαγωγή, πρόκειται για ένα πρόβλημα που μπορεί να αντιμετωπιστεί ως συνδυαστικό πρόβλημα βελτιστοποίησης (problem of combinatorial optimization) και

Πίνακας 6.2: Ακρίβεια ταξινόμησης και τυπική απόκλιση των συγκρινόμενων μεθόδων, όπου το 'ET' δηλώνει τον αλγόριθμο ExtraTrees, το 'GB' αντιπροσωπεύει τη μέθοδο Gradient Boosting, ενώ το 'RF' δηλώνει τον αλγόριθμο Random Forest.

Dataset	Stacking		ET		GB		RF	
audiology	79.13	5.01	66.13	7.28	82.66	5.28	76.99	5.79
autos	82.15	6.78	68.49	6.22	82.54	5.30	79.12	4.17
badges	98.98	3.28	96.86	4.30	100.00	0.00	99.11	1.92
balance-scale	89.38	1.30	78.78	2.93	87.33	2.01	81.66	2.41
breast-cancer	70.55	4.20	64.07	6.71	69.58	5.87	71.82	3.81
wisconsin-breast-cancer	96.42	1.71	94.56	2.00	96.17	1.74	95.91	1.83
horse-colic	83.64	3.51	73.97	5.20	83.80	2.99	83.32	3.53
credit-rating	86.70	1.97	79.19	4.01	86.14	1.95	86.09	2.13
german-credit	75.92	2.53	66.70	2.79	75.52	2.40	74.10	2.70
pima-diabetes	75.89	3.52	67.89	2.79	75.73	3.63	74.04	2.95
glass	76.27	5.63	63.65	8.37	75.34	4.85	73.10	5.53
grub-damage	41.81	6.68	40.13	6.39	38.97	7.73	41.29	7.90
haberman	72.29	2.53	65.17	4.87	69.60	3.84	70.52	3.16
hear-cleveland	80.34	5.07	76.38	4.63	80.80	4.66	79.81	5.60
heart-hungarian	81.64	4.49	75.92	4.83	82.25	5.04	80.68	5.73
heart-statlog	80.96	4.43	74.22	5.37	78.89	4.66	79.70	4.19
hepatitis	83.10	5.98	78.19	7.41	81.94	5.59	81.94	6.45
hypothyroid	99.49	0.29	94.18	1.02	99.46	0.31	99.14	0.32
ionosphere	93.28	3.13	87.12	4.03	92.88	2.81	92.31	2.64
iris	94.27	3.79	91.73	5.54	95.33	3.43	95.33	3.47
labor	90.58	9.97	85.48	12.86	85.73	10.66	90.91	9.62
lymphography	83.79	7.67	77.31	9.13	85.15	7.13	82.42	7.07
monk1	98.71	3.03	74.69	12.95	98.87	2.75	86.15	6.52
monk2	63.46	6.91	54.80	7.42	64.74	8.42	57.50	6.43
monk3	91.65	4.01	82.69	7.32	91.17	4.20	90.66	5.19
mushroom	100.00	0.00	99.99	0.02	99.99	0.05	100.00	0.00
primary-tumor	42.48	3.29	34.28	4.69	39.82	4.68	40.81	4.22
students	83.96	3.75	80.46	3.37	84.19	3.73	82.62	3.30
sick	98.65	0.34	94.59	1.05	98.68	0.31	98.26	0.42
sonar	83.38	6.26	69.80	5.79	82.51	6.14	79.46	7.22
soybean	93.00	2.04	88.20	3.13	93.06	1.99	92.94	1.79
relation	79.02	0.94	78.94	1.02	78.94	1.02	78.95	1.05
vehicle	75.44	2.41	65.91	2.92	75.42	2.29	74.02	2.64
vote	96.00	2.13	92.92	2.96	95.31	2.32	95.49	2.87
vowel	92.93	1.93	75.82	3.31	89.78	2.09	91.37	2.03
waveform	85.61	1.21	68.37	1.52	85.52	1.20	81.36	1.22
wine	96.73	2.77	87.97	6.14	95.05	4.58	96.18	3.61
zoo	93.65	4.70	88.70	7.69	89.30	4.46	94.89	4.09

άρα, οι μέθοδοι που αναλύονται στο Κεφάλαιο 3 μπορούν να βρουν ένα επιπλέον, πεδίο εφαρμογής.

Μέρος IV

**Βελτιστοποίηση στη Μηχανική
Μάθηση**

Εξελικτικοί Αλγόριθμοι πολυ-αντικειμενικής βελτιστοποίησης στη Μηχανική Μάθηση

Μία μέλισσα μέλι ου ποιεί

— μ

Οι αλγόριθμοι (Machine Learning) αξιοποιούν ένα συγκεκριμένο σύνολο δεδομένων προκειμένου να δημιουργήσουν ένα αποτελεσματικό ή μ (predictive ή descriptive model). Η μ ή - μ (multi-objective evolutionary optimization) βοηθάει τους αλγορίθμους Μηχανικής Μάθησης να βελτιστοποιούν τις - μ (hyper-parameters) τους, συνήθως κάτω από αντικρουόμενες αντικειμενικές συνιστώσες και συμβάλλει στην επιλογή του καλύτερου μοντέλου για το δοθέν πρόβλημα. Στην παρούσα έρευνα [8], εξετάζονται πρόσφατες εξελικτικές προσεγγίσεις πολλαπλών αντικειμένων για τέσσερις βασικές διεργασίες της μ και , και συγκεκριμένα, αφορούν: (α) την μ (data preprocessing), (β) το πρόβλημα της μ (classification), (γ) την εργασία της μ (clustering) δεδομένων και (δ) τους (association rules).

Για ένα πρόβλημα βελτιστοποίησης, γενικά, η βελτιστοποίηση εξετάζει τα ακόλουθα ζητήματα, όπως ορίζονται αναλυτικά στο άρθρο [135]:

- μ (Objective function): η ποσότητα που πρέπει να βελτιστοποιηθεί (να μεγιστοποιηθεί ή να ελαχιστοποιηθεί).
- (Variables): οι εισόδοι της αντικειμενικής συνάρτησης.
- μ (Constraints): οι περιορισμοί που αποδίδονται στις εισόδους της αντικειμενικής συνάρτησης.

Επομένως, ο σκοπός ενός βελτιστοποιητή είναι να προσδιορίσει σωστά τις τιμές στις εισόδους της αντικειμενικής συνάρτησης, με τέτοιο τρόπο ώστε να επιτευχθεί η βέλτιστη λύση για τη συνάρτηση και να εκπληρωθούν όλοι οι απαιτούμενοι περιορισμοί.

Διαφορετικά προβλήματα βελτιστοποίησης που συναντώνται στον πραγματικό κόσμο συχνά υποφέρουν από τις ακόλουθες δυσκολίες [373]:

- Σε πολλές περιπτώσεις, είναι δύσκολο να διακρίνουμε τους ολικούς ελαχιστοποιητές από τους τοπικούς ελαχιστοποιητές (για το πρόβλημα της ελαχιστοποίησης μιας αντικειμενικής συνάρτησης).
- Η εκτίμηση των λύσεων μπορεί να είναι δύσκολη παρουσία θορύβου (noise).

3. Ο χώρος αναζήτησης των λύσεων μπορεί να είναι μεγάλος, οπότε και η διαστασιμότητα του προβλήματος αυξάνεται με παρόμοιο τρόπο. Αυτό δύναται να προκαλέσει ένα φαινόμενο που είναι γνωστό ως «κατάρτα της διάστασης», ένα πρόβλημα πολύ δύσκολο να διαχειριστεί.
4. Δυσκολίες που σχετίζονται με τους υποκείμενους περιορισμούς οι οποίοι αποδίδονται στις εισόδους της αντικειμενικής συνάρτησης.
5. Αναγκαιότητα για τεχνικές βελτιστοποίησης σε συσχετισμό με συγκεκριμένα προβλήματα.

Στην περίπτωση που η ποσότητα που επιζητούμε να βελτιστοποιηθεί εκφράζεται μόνο από μία αντικειμενική συνάρτηση, το πρόβλημα είναι γνωστό ως πρόβλημα βελτιστοποίησης ή μ ή μ (uni-objective ή single-objective problem). Ενώ στην περίπτωση του προβλήματος βελτιστοποίησης μ (multi-objective) εντοπίζονται περισσότεροι από έναν επιμέρους στόχοι, οι οποίοι θα πρέπει να βελτιστοποιηθούν ταυτόχρονα.

Διάφορες εφαρμογές [162, 281] της Μηχανικής Μάθησης αλλά και άλλοι τύποι προβλημάτων [194, 221, 297] αντιμετωπίστηκαν με τεχνικές [29, 341] που ανήκουν στον τομέα της Μηχανικής Μάθησης, απαιτούν την εκπλήρωση διαφόρων συνθηκών. Η ταυτόχρονη εκπλήρωση αυτών των συνθηκών, καθώς και η βελτιστοποίηση των παραμέτρων που ενσωματώνονται στις μεθόδους Μηχανικής Μάθησης, αποτελεί ένα δύσκολο πρόβλημα βελτιστοποίησης. Πράγματι, η πλειοψηφία αυτών των αλγορίθμων [85, 86, 446] απαιτεί τη βελτιστοποίηση πολλαπλών αντικειμένων, έτσι ώστε το αποτέλεσμα να είναι αξιόπιστο και ανταγωνιστικό. Για παράδειγμα, στο πρόβλημα της (feature selection), το επιθυμητό σύνολο χαρακτηριστικών πρέπει να είναι το ελάχιστο σύνολο που μεγιστοποιεί την απόδοση του ταξινομητή. Επομένως, απαιτούνται δύο προϋποθέσεις:

- (α) το ελάχιστο υποσύνολο χαρακτηριστικών από το αρχικό σύνολο δεδομένων και
- (β) αυτά τα χαρακτηριστικά που θα επιλεγούν πρέπει να μεγιστοποιούν την απόδοση του αλγορίθμου.

Ως εκ τούτου, η πλειονότητα των προβλημάτων μάθησης είναι μ και έτσι, είναι προφανές ότι θα θεωρήσουμε τα προβλήματα μάθησης ως μ - μ . Ο Freitas στο άρθρο [137] παρουσίασε τον προαναφερθέντα σκοπό, σύμφωνα με τον οποίο πρέπει να πραγματοποιείται ταυτόχρονα βελτιστοποίηση υπό ορισμένες συνθήκες, έτσι ώστε η απόδοση του μοντέλου να είναι τελικά υψηλή. Ως επί το πλείστον, απαιτείται βελτιστοποίηση ενός αριθμού παραμέτρων, προκειμένου να μεγιστοποιηθεί η ακρίβεια του μοντέλου. Για να επιτευχθεί αυτός ο στόχος, υπάρχουν τρεις διαφορετικές προσεγγίσεις:

1. Η μετατροπή του αρχικού μ - μ προβλήματος σε ένα πρόβλημα μ - μ χρησιμοποιώντας σωστά μια μ μ (weighted approach).
2. Η λεξικογραφική προσέγγιση (lexicographical approach), όπου οι αντικειμενικές ποσότητες έχουν προτεραιότητα.
3. Η γνωστή και ευρέως χρησιμοποιούμενη Pareto, η οποία δίνει ένα ολόκληρο σύνολο μη-επικρατούμενων λύσεων.

Ένα σημαντικό ζήτημα που παρέχεται από έναν αλγόριθμο βελτιστοποίησης πολλαπλών αντικειμένων [295, 395, 443] είναι ότι, αντί μιας λύσης, δύναται να επιστρέψει μια σειρά από καλές «υποψήφιας» λύσεις, τις αποκαλούμενες «μη-κυριαρχικές λύσεις», τις οποίες ο χρήστης μπορεί να συγκρίνει μεταξύ τους προκειμένου να επιλεγεί η καταλληλότερη λύση για τον

συγκεκριμένο σκοπό [87]. Το σύνολο λύσεων που επιστρέφει ο αλγόριθμος αντιπροσωπεύει το καλύτερο δυνατό «αντάλλαγμα» μεταξύ των αντικειμενικών ποσοτήτων. Ωστόσο, η Pareto [289] υποδεικνύει ότι η σχέση μεταξύ εισόδων και εξόδων του αλγορίθμου δεν είναι ισορροπημένη. Έτσι, ο Pareto (ή κανόνας 80=20) [273] «υπακούει» σε μια κατανομή, γνωστή ως μ Pareto, η οποία αποκαλύπτει ότι το 20% των εισόδων είναι υπεύθυνες για το 80% των ληφθέντων αποτελεσμάτων.

Σε πολλές περιπτώσεις, η απόφαση ενός ειδικού, του επονομαζόμενου (decision maker) [198], διαδραματίζει σημαντικό ρόλο. Έτσι, η γνώμη του υπευθύνου λήψης αποφάσεων μπορεί να ζητηθεί αρχικά, πριν από την έναρξη της διαδικασίας επίλυσης. Σύμφωνα με τις πληροφορίες που παρέχει ο εμπειρογνώμονας, απαιτείται η πλέον ενδεδειγμένη λύση για την ικανοποίηση των όρων που έχουν τεθεί. Ωστόσο, η γνώμη ενός εμπειρογνώμονα μπορεί να ζητηθεί μετά την εξεύρεση μιας σειράς κατάλληλων λύσεων. Σε αυτή την περίπτωση, ο εμπειρογνώμονας επιλέγει την καλύτερη λύση μεταξύ αυτών των επιλογών. Επιπλέον, η γνώμη του εμπειρογνώμονα μπορεί να λάβει χώρα κατά τη διάρκεια της διαδικασίας. Συγκεκριμένα, το μοντέλο ζητάει επανειλημμένα τη γνώμη του εμπειρογνώμονα να προκειμένου να βελτιώσει τις λύσεις και να επιστρέψει τελικά το απαιτούμενο βέλτιστο σύνολο λύσεων.

Παρά το γεγονός ότι τα περισσότερα από τα πραγματικά προβλήματα απαιτούν τη βελτιστοποίηση πολλαπλών αντικειμένων [72], υπάρχει πάντα προσπάθεια να περιοριστεί ο αριθμός των στόχων στο ελάχιστο [75]. Αυτό συμβαίνει λόγω του γεγονότος ότι όσο περισσότεροι στόχοι απαιτούνται για βελτιστοποίηση, τόσο περισσότερες λύσεις θα επιστραφούν από τον αλγόριθμο. Κατά συνέπεια, η διάσταση και η πολυπλοκότητα του προβλήματος αυξάνονται, και ως εκ τούτου, το πρόβλημα καθίσταται δυσκολότερο να λυθεί. Για μια ανάλυση των διαφορετικών τύπων και τεχνικών βελτιστοποίησης πολλαπλών αντικειμένων, ο αναγνώστης μπορεί να μελετήσει για περισσότερες λεπτομέρειες την εργασία [205].

Σε αυτό το κεφάλαιο, οι αναφορές μας επικεντρώνονται σε πρόσφατες ερευνητικές προσπάθειες που δημοσιεύθηκαν σε διεθνή περιοδικά, βιβλία και συνέδρια. Επιπλέον, ενσωματώνονται διάφορες αναφορές σχετικά με το πρόβλημα της πολυ-αντικειμενικής βελτιστοποίησης, οι οποίες θεωρούμε πως είναι οφέλιμες για την κατανόηση, την περαιτέρω εξήγηση της εφαρμοσιμότητας αλλά και τη σύνδεση με τα όσα αναλύσαμε στο Κεφάλαιο 2 και το Κεφάλαιο 3. Σύμφωνα με αυτό το σκοπό, οι πρώτες μείζονες προσπάθειες που έχουν ερευνηθεί διαφορετικές προσεγγίσεις βελτιστοποίησης πολλαπλών αντικειμενικών παρουσιάζονται στις εργασίες [267, 267].

Όπως αναμένεται, δεν είναι δυνατό στα πλαίσια ενός μόνο κεφαλαίου να καλυφθούν εκτενώς όλες οι πτυχές των - μ μ (multi-objective evolutionary optimization algorithms). Ωστόσο, ελπίζουμε, μέσα από αυτό το κεφάλαιο και τις πιο πρόσφατες αναφορές, ότι ο αναγνώστης θα ενημερωθεί για τα τελευταία ενδιαφέροντα της επιστημονικής κοινότητας γύρω από αυτά τα σημαντικά ζητήματα, αλλά επίσης, θα χαρτογραφήσει τις τομές που υπάρχουν με τα προηγούμενα κεφάλαια της παρούσας διατριβής.

Έτσι, η παρακάτω ενότητα παρέχει το απαραίτητο υπόβαθρο και βασικές έννοιες της βελτιστοποίησης πολλαπλών αντικειμένων. Η Ενότητα 7.2 καλύπτει μια πολύ σημαντική πτυχή της Μηχανικής Μάθησης, η οποία είναι η προεπεξεργασία των δεδομένων. Για το σκοπό αυτό, παρουσιάζονται διάφοροι εξειδικευμένοι αλγόριθμοι πολλαπλών αντικειμένων, οι οποίοι ασχολούνται με τα πιο κοινά και ευρέως χρησιμοποιούμενα βήματα καθαρισμού δεδομένων. Η εργασία ταξινόμησης και τα βασικά μοντέλα που χρησιμοποιούνται για τη διαχείριση αυτής, σύμφωνα με αλγορίθμους βελτιστοποίησης πολλαπλών αντικειμένων, περιγράφονται στην Ενότητα 7.3. Στη συνέχεια, στην Ενότητα 7.4, παρουσιάζονται προσεγγίσεις συσταδοποίησης και κανόνων συσχέτισης. Στην Ενότητα 7.5, παρουσιάζονται μερικές από τις πιο πρόσφατες εφαρμογές που αφορούν τους αλγορίθμους εξελικτικής βελτιστοποίησης πολλαπλών αντικειμένων. Το κεφάλαιο τελειώνει στην Ενότητα 7.6, με μια σύνοψη

και μια σύντομη συζήτηση για μελλοντικές ερευνητικές προσπάθειες.

7.1 Βασικές έννοιες της βελτιστοποίησης πολλαπλών αντικειμένων

Η μ (multi-objective optimization (MO)) που ονομάζεται, επίσης, μ (multiple criteria optimization) αντιμετωπίζει προβλήματα όπου οι διαφορετικοί (targets) ή μ (objectives) πρέπει να βελτιστοποιούνται ταυτόχρονα. Για τέτοιου είδους προβλήματα, ο «βέλτιστος χαρακτήρας» κατά Pareto αντικαθιστά τη βέλτιστη έννοια της βελτιστοποίησης ενός αντικειμένου (single-objective optimization) και κάθε βέλτιστη λύση κατά Pareto αντιπροσωπεύει μια ανταλλαγή μεταξύ των αντικειμενικών συναρτήσεων. Ως εκ τούτου, δύο λύσεις μπορούν να αποκτήσουν την ίδια μ (fitness value) και είναι επιθυμητό να ληφθεί η μεγαλύτερη δυνατή καταμέτρηση λύσεων με διαφορετικές εγγενείς ιδιότητες.

Ας υποθέσουμε ότι $S \subset \mathbb{R}^n$ είναι ένας n -διαστάσεων χώρος αναζήτησης και υποθέτουμε ότι

$$f_i(x) : S \rightarrow \mathbb{R}; \quad i = 1; 2; \dots; k;$$

είναι k μ (objective functions) που ορίζονται πάνω στο S . Ας υποθέσουμε ότι,

$$g_j(x) \leq 0; \quad j = 1; 2; \dots; m;$$

είναι m μ (constraints), τότε το πρόβλημα πολυ-κριτηριακής βελτιστοποίησης μπορεί να οριστεί ως εξής: Αναζητούμε ένα σημείο:

$$x = (x_1; x_2; \dots; x_n) \in S;$$

που πληροί τους περιορισμούς και βελτιστοποιεί την ακόλουθη συνάρτηση:

$$F_{nk}(x) = (f_1(x); f_2(x); \dots; f_k(x)) : \mathbb{R}^n \rightarrow \mathbb{R}^k;$$

Οι αντικειμενικές συναρτήσεις μπορεί να είναι αντικρουόμενες μεταξύ τους, επομένως είναι συνήθως αδύνατο να εντοπιστεί το ολικό ελάχιστο για όλους τους στόχους στο ίδιο σημείο. Ο στόχος της βελτιστοποίησης πολλών αντικειμένων είναι να παράσχει ένα

Pareto (Pareto optimal solutions) (σημεία) στο προαναφερθέν πρόβλημα.

Συγκεκριμένα, υποθέτουμε ότι $u = (u_1; u_2; \dots; u_k)$ και $v = (v_1; v_2; \dots; v_k)$ είναι δύο διανύσματα. Στη συνέχεια, το u θεωρούμε ότι «κυριαρχεί» (dominates) του v αν και μόνο αν $u_i \leq v_i$, για $i = 1; 2; \dots; k$, και $u_i < v_i$ για τουλάχιστον μια συνιστώσα. Αυτή η συνθήκη είναι γνωστή ως

Pareto (Pareto dominance) και χρησιμοποιείται για τον προσδιορισμό των βέλτιστων λύσεων κατά Pareto. Ως εκ τούτου, μια λύση x του προβλήματος πολυ-αντικειμενικής βελτιστοποίησης καλείται Pareto (Pareto optimal) αν και μόνο αν δεν υπάρχει άλλη λύση y , έτσι ώστε το $F_{nk}(y)$ να «κυριαρχεί» του $F_{nk}(x)$.

Το σύνολο όλων των βέλτιστων λύσεων Pareto ενός προβλήματος πολυ-κριτηριακής βελτιστοποίησης, που υποδηλώνεται με P , ονομάζεται Pareto (Pareto optimal set) ενώ το σύνολο:

$$PF = \{f_1(x); f_2(x); \dots; f_k(x) \mid x \in P\};$$

ονομάζεται μ Pareto (Pareto front). Ένα μ Pareto PF λέγεται ότι είναι (convex) αν και μόνο αν υπάρχει $w \in PF$, έτσι ώστε:

$$kuk + (1 - \mu)kvk \leq kwk; \delta u; v \in P; \delta \in [0, 1];$$

ενώ ονομάζεται (concave) αν και μόνο αν υπάρχει $W \in PF$, έτσι ώστε:

$$kuk + (1 - \mu)kvk \leq kwk; \delta u; v \in P; \delta \in [0, 1];$$

Ένα μ -Pareto μπορεί να είναι κυρτό, κοίλο, ή μερικώς κυρτό και / ή κοίλο και / ή να είναι μη συνεχές. Οι τελευταίες τρεις περιπτώσεις παρουσιάζουν τη μεγαλύτερη δυσκολία για την πλειοψηφία των τεχνικών βελτιστοποίησης πολλών αντικειμένων.

Χρησιμοποιώντας την προσέγγιση βελτιστοποίησης πολλών κριτηρίων είναι επιθυμητή η ανίχνευση όλων των μ -Pareto. Από την άλλη πλευρά, το βέλτιστο σύνολο Pareto μπορεί να είναι άπειρο και δεδομένου ότι ο υπολογισμός συνήθως περιορίζεται σε αυστηρούς περιορισμούς χρόνου και χώρου, ο κύριος στόχος της βελτιστοποίησης πολλαπλών κριτηρίων είναι η ανίχνευση του μεγαλύτερου δυνατού αριθμού βέλτιστων λύσεων Pareto, με τη μικρότερη πιθανή απόκλιση από το μ -Pareto και την κατάλληλη εξάπλωση [296].

Οι μ (evolutionary algorithms) έχουν την ικανότητα να εξελίσσουν ταυτόχρονα πολλαπλές βέλτιστες λύσεις Pareto και έτσι, είναι ιδιαίτερα αποτελεσματικές και αποδοτικές στην αντιμετώπιση προβλημάτων πολυ-αντικειμενικής βελτιστοποίησης. Οι ανιχνεύσιμες βέλτιστες λύσεις Pareto αποθηκεύονται σε δομές μνήμης, που ονομάζονται (external archives) τα οποία γενικά αυξάνουν την απόδοση των προσεγγίσεων βελτιστοποίησης πολλαπλών στόχων. Μια πληθώρα γνωστών και ευρέως εφαρμοζόμενων μ έχουν προταθεί, οι οποίες βασίζονται σε διαφορετικές τεχνικές, μεταξύ των οποίων είναι η niching fitness sharing προσέγγιση και ο μ (elitism), μεταξύ άλλων [88, 116, 296].

7.2 Προεπεξεργασία δεδομένων

Οι αλγόριθμοι Μηχανικής Μάθησης (MM) (machine learning (ML)) αυτοματοποιούν τη διαδικασία εξαγωγής γνώσης από μορφές που μπορούν εύκολα να επεξεργαστούν από συστήματα υπολογιστών. Γενικά, η «ποιότητα των δεδομένων» θα μπορούσε να μειώσει την απόδοση ενός αλγορίθμου μάθησης. Έτσι, η προεπεξεργασία δεδομένων [148] είναι ένα σημαντικό έργο στη MM που συνήθως εκτελείται με την αφαίρεση αντικειμένων και χαρακτηριστικών που περιέχουν εξωγενείς και άσχετες πληροφορίες. Στη συνέχεια, θα παραθέσουμε γνωστές μεθόδους της κατηγορίας πολυ-κριτηριακής βελτιστοποίησης για κάθε βήμα της προεπεξεργασίας δεδομένων.

7.2.1 Επιλογή χαρακτηριστικών

Το πρόβλημα της ανίχνευσης και της εξάλειψης άσχετων και περιττών (features), γνωστών και ως μ (attributes), ονομάζεται (feature selection (FS)) [350]. Αυτή η διαδικασία προσπαθεί να «συμπίεσει» την πληθυσμιακή των χαρακτηριστικών του συνόλου δεδομένων και να βοηθήσει έτσι, τους αλγορίθμους μάθησης να λειτουργήσουν ταχύτερα και αποτελεσματικότερα. Γενικά, τα χαρακτηριστικά μπορούν να διακριθούν ως εξής:

- (α) (Relevant): Είναι τα χαρακτηριστικά που συμβάλλουν με ένα σημαντικό ρόλο στην κλάση των δεδομένων και είναι χαρακτηριστικά που δεν μπορούν να ληφθούν υπόψη από τα υπόλοιπα στοιχεία του συνόλου.
- (β) (Irrelevant): πρόκειται για χαρακτηριστικά που δεν επηρεάζουν την κατηγορία στόχου.

(γ) (Redundant): Χαρακτηριστικά που μπορούν να αντικατασταθούν από άλλα χαρακτηριστικά χωρίς να επηρεάζεται το σύνολο δεδομένων.

Με την εξάλειψη των άσχετων και περιττών χαρακτηριστικών, η διαδικασία θα μπορούσε να βοηθήσει στη μείωση του χρόνου εκπαίδευσης, καθώς και στην απλοποίηση των μοντέλων μάθησης ή/και στη βελτίωση του μέτρου απόδοσης του προβλήματος. Γενικά, η διαδικασία επιλογής χαρακτηριστικών μπορεί να θεωρηθεί ως ένα πρόβλημα πολλαπλών αντικειμένων. Οι κύριοι στόχοι είναι δύο: το πρώτο αντικείμενο είναι η μεγιστοποίηση της απόδοσης του μοντέλου, ενώ το δεύτερο είναι η ελαχιστοποίηση του αριθμού των χαρακτηριστικών που θα τροφοδοτούνται στον αλγόριθμο μάθησης. Οι προαναφερόμενοι στόχοι είναι συγκρουόμενοι και η βέλτιστη επιλογή πρέπει να γίνει λαμβάνοντας υπόψη την ισορροπία μεταξύ των δύο στόχων. Η

(multi-objective feature selection) μπορεί να αποκτήσει ένα σύνολο μη-κυριαρχικών διαχωρισμών χαρακτηριστικών, ώστε να ικανοποιεί ποικίλες απαιτήσεις σε πραγματικές εφαρμογές.

Στη συνέχεια, παρουσιάζουμε εν συντομία διάφορες προσεγγίσεις για την επιλογή χαρακτηριστικών. Η αποδοτική και αποτελεσματική μέθοδος μ μ (particle swarm optimization (PSO)) [211, 296] θεωρείται ως μια μετα-ευρετική προσέγγιση που προσπαθεί να επιλύσει ένα πρόβλημα βελτιστοποίησης διατηρώντας έναν πληθυσμό υποψήφιων λύσεων (τα οποία ονομάζονται σωματίδια). Τα μέλη του σμήνους κινούνται γύρω από τον χώρο αναζήτησης σύμφωνα με ένα μαθηματικό μοντέλο που αντιμετωπίζει δύο παραμέτρους: τη θέση και την ταχύτητα των σωματιδίων. Οι συγγραφείς του άρθρου [428] διεξήγαγαν μια μελέτη σχετικά με διαφορετικούς τύπους βελτιστοποίησης πολλαπλών στόχων με χρήση σμήνους σωματιδίων για το πρόβλημα της επιλογής χαρακτηριστικών. Ο κύριος στόχος της εργασίας τους ήταν να δημιουργήσουν ένα σχήμα πολλαπλών στόχων βασισμένο στην PSO για την επιλογή χαρακτηριστικών με σκοπό την αντιμετώπιση προβλημάτων ταξινόμησης. Η προσέγγισή τους προσπαθεί να επιτύχει ένα μ Pareto από μη-κυριαρχικές λύσεις, το οποίο θα περιέχει ένα υποσύνολο του αρχικού χώρου χαρακτηριστικών, επιτυγχάνοντας ταυτόχρονα μια πιο ακριβή απόδοση ταξινόμησης χωρίς να χρησιμοποιήσει όλα τα διαθέσιμα χαρακτηριστικά.

Οι Han και Ren [173] πρότειναν μια τεχνική πολλαπλών αντικειμένων για τη βελτίωση της απόδοσης της επιλογής χαρακτηριστικών. Η μέθοδός τους μπορεί να ανταποκριθεί σε διαφορετικές απαιτήσεις, καθώς και να επιτύχει έναν καλό συνδυασμό μεταξύ διαφορετικών αντικρουόμενων στόχων.

Οι Paul και Das [298] πρότειναν ένα σχήμα επιλογής και μια μέθοδο στάθμισης που υποστηρίζεται από έναν εξελικτικό πολυ-αντικειμενικό αλγόριθμο. Τα χαρακτηριστικά των παραδειγμάτων επιλέγονται και «ζυγίζονται» ή κλιμακώνονται (scaled) και συγχρόνως εμφανίζονται τα σημεία σε ένα συγκεκριμένο υπερ-χώρο. Επιπλέον, οι αποστάσεις μεταξύ των σημείων των μη ταυτόσημων κατηγοριών αυξάνονται με τέτοιο τρόπο, ώστε να διευκολύνεται η ταξινόμησή τους.

Οι Wang et al. [410] παρουσίασαν έναν αλγόριθμο που ονομάζεται MECY-SF. Ο αλγόριθμός τους αξιοποιεί τη γενετική αναζήτηση και τη βελτιστοποίηση πολλών αντικειμένων για να ξεπεράσει τους περιορισμούς των λεγόμενων «άπληστων αλγορίθμων» επιλογής χαρακτηριστικών. Επιπλέον, υιοθετήθηκε ο γρήγορος και ελιτιστικός πολυ-αντικειμενικός γενετικός αλγόριθμος NSGA-II [88] για την επίλυση του προβλήματος πολυ-κριτηριακής επιλογής χαρακτηριστικών.

Οι συγγραφείς του άρθρου [54], μέσω μιας νέας μεθόδου πολλαπλών αντικειμένων πέτυχαν την εξαγωγή χαρακτηριστικών και την οπτικοποίηση δεδομένων. Ο προτεινόμενος αλγόριθμος βασίστηκε στο Pareto και συνδυάζεται με τον μ μ . Διάφορα μέτρα ταξινόμησης και απεικόνισης θεωρήθηκαν ως στόχοι που πρέπει να βελτιστοποιηθούν με αυτό το νέο αλγόριθμο.

Μια μέθοδος που αντιμετώπισε το πρόβλημα της επιλογής χαρακτηριστικών ως πρόβλημα βελτιστοποίησης δύο αντικειμένων προτάθηκε από τους Das A. και Das S. [80]. Κατά τη διάρκεια αυτού του αλγορίθμου διαφορετικά πραγματικά βάρη αποδίδονταν σε κάθε χαρακτηριστικό του συνόλου. Επομένως, ένα υποσύνολο σταθμισμένων χαρακτηριστικών επιλέγεται ως το καλύτερο υποσύνολο για την επερχόμενη ταξινόμηση του συνόλου δεδομένων. Τα μέτρα (relevancy) και μ (redundancy) επιλέχθηκαν για τη δημιουργία των αντικειμενικών συναρτήσεων.

Οι Hancer et al. πρότειναν [174] έναν αλγόριθμο πολλών αντικειμένων, ο οποίος στηρίχθηκε σε τεχνητές αποικίες μελισσών, προκειμένου να επιτύχει το κομμάτι της βελτιστοποίησης και κατ' επέκταση της κατάλληλης επιλογής χαρακτηριστικών. Συγκεκριμένα, οι συγγραφείς του άρθρου ανέπτυξαν μια μέθοδο επιλογής χαρακτηριστικών που θα αναζητήσει ένα βέλτιστο Pareto σύνολο χαρακτηριστικών. Πρότειναν δύο εκδοχές, δηλαδή τη μ (binary multi-objective artificial bee colony) που ονομάστηκε Bin-MOABC και την αντίστοιχη συνεχή εκδοχή που ονόμασαν ως Num-MOABC αλγόριθμο. Ο αλγόριθμός τους προσεγγίζει το πρόβλημα πολλαπλών αντικειμένων μέσω της ελάχιστης επιλογής χαρακτηριστικών που παρέχει το χαμηλότερο σφάλμα ταξινόμησης σύμφωνα με το αρχικό σύνολο χαρακτηριστικών. Έλεγξαν τον προτεινόμενο αλγόριθμο σε δώδεκα σύνολα δεδομένων αναφοράς και τα πειραματικά τους αποτελέσματα δείχνουν ότι ο αλγόριθμος Bin-MOABC παρουσιάζει καλύτερη ακρίβεια ταξινόμησης και υπερέρχει των άλλων μεθόδων που εξετάστηκαν σχετικά με τη μείωση της διαστασιμότητας του προβλήματος.

Τέλος, οι Zheng και Wang πρότειναν [442] μία μέθοδο που συνδυάζει την κοινή εντροπία της μέγιστης πληροφορίας (joint maximal information entropy (JMIE)) ως μετρική ενός υποσυνόλου χαρακτηριστικών και έναν αλγόριθμο δυαδικής βελτιστοποίησης με χρήση σμήνους σωματιδίων (binary particle swarm optimization (BPSO)) για την αναζήτηση του βέλτιστου συνόλου των χαρακτηριστικών. Οι συγγραφείς του άρθρου, διεξήγαγαν πειράματα σε πέντε σύνολα δεδομένων από τη γνωστή βάση δεδομένων UCI και τα πειραματικά τους αποτελέσματα δείχνουν ότι η προτεινόμενη τεχνική παρουσιάζει καλύτερη απόδοση σε προβλήματα επιλογής χαρακτηριστικών πολλαπλών κλάσεων. Επιπλέον, η μέθοδός τους είναι πιο συνεπής και επιτυγχάνει καλύτερη απόδοση από άποψη χρόνου συγκριτικά με τον αλγόριθμο BPSO-SVM.

7.2.2 Επιλογή παραδειγμάτων

Η μ ή (instance selection ή prototype selection) [94] μπορεί να θεωρηθεί ως πρόβλημα βελτιστοποίησης δεδομένου ότι, καταρχάς, απαιτείται η διατήρηση της ποιότητας του συνόλου δεδομένων, και δευτερευόντως η ελαττοποίηση του μεγέθους του δείγματος. Η πολυπλοκότητα του συγκεκριμένου προβλήματος τείνει να αυξηθεί ανάλογα με τον αριθμό των παραδειγμάτων εκπαίδευσης. Από την άλλη πλευρά, αυτό μπορεί να μειώσει την επεξηγησιμότητα (interpretability) των αποτελεσμάτων. Επομένως, η επιλογή παραδειγμάτων συνιστάται ιδιαίτερα στην περίπτωση μεγάλων συνόλων δεδομένων. Για το σκοπο αυτό οι Fernández et al. [126] χρησιμοποίησαν έναν εξελικτικό αλγόριθμο πολλαπλών αντικειμένων για τη διεξαγωγή της αναζήτησης, ώστε να αποκτηθεί από κοινού το καλύτερο σύνολο χαρακτηριστικών και παραδειγμάτων.

Μια προσπάθεια κατά την οποία συναντάται ένα σχήμα πολυ-κριτηριακής βελτιστοποίησης για την αντιμετώπιση του προβλήματος της κατάλληλης μ (training set selection (TSS)) συναντάται στο άρθρο [1]. Η κύρια διαφορά μεταξύ της προτεινόμενης τεχνικής και των εξελικτικών προσεγγίσεων που είχαν ήδη αναπτυχθεί είναι η a priori μ (multi-objective a priori technique). Αυτό σημαίνει ότι η μέθοδός τους διατηρεί δύο στόχους, τη μ (classification accuracy) και τη μείωση του ρυθμού (rate reduction), σε αντίθεση με όλες τις άλλες εξελικτικές μεθόδους που χρησιμοποιήθηκαν για την επιλογή του κατάλληλου υποσου-

νόλου παραδειγμάτων εκπαίδευσης έχοντας ως βάση τη μέθοδο των μηχανών διανυσματικής υποστήριξης. Οι συγγραφείς του άρθρου εξέτασαν τη μέθοδό τους χρησιμοποιώντας σύνολα δεδομένων από τη βάση UCI και τα διεξαγόμενα πειράματα δείχνουν ότι ο νέος αυτός αλγόριθμος επιδεικνύει καλύτερη επίδοση αναφορικά με τις γνωστές τεχνικές TSS και επιπλέον, ενισχύει την αποτελεσματικότητα των SVM.

7.2.3 Ελλιπή δεδομένα

Τα ή μ (incomplete ή corrupted) δεδομένα είναι ένα κοινό πρόβλημα [244] που συναντάται στις πραγματικές βάσεις δεδομένων. Υπάρχουν πολλά ζητήματα που πρέπει να ληφθούν υπόψη σύμφωνα με την επεξεργασία άγνωστων χαρακτηριστικών. Ωστόσο, ο προσδιορισμός της προέλευσης αυτής της «ασάφειας» είναι ένα από τα πλέον σημαντικά ζητήματα. Έτσι, οδηγούμαστε στους παρακάτω λόγους:

1. Κάποιο χαρακτηριστικό μπορεί να παραλείπεται επειδή κάποιος ξέχασε να το προσθέσει στη βάση δεδομένων ή για κάποιο λόγο χάθηκε.
2. Για ένα δεδομένο αντικείμενο δεν υπάρχει συγκεκριμένη τιμή χαρακτηριστικού.
3. Ο «συλλέκτης» των δεδομένων εκπαίδευσης μπορεί να μην ενδιαφέρεται για μια συγκεκριμένη τιμή χαρακτηριστικού για ένα δεδομένο παράδειγμα.

Μια ενδιαφέρουσα προσπάθεια εντοπίζεται στο άρθρο [245], όπου οι συγγραφείς του παρουσίασαν ένα γενετικό αλγόριθμο πολλαπλών αντικειμένων για τον καταλογισμό των δεδομένων. Συγκεκριμένα, η προτεινόμενη μέθοδος βασίστηκε στο γρήγορο και ελιπτικό πολυ-αντικειμενικό γενετικό αλγόριθμο NSGA-II [88], ο οποίος είναι κατάλληλος για μικτά (κατηγορικά και συνεχή) σύνολα χαρακτηριστικών και θεωρεί πληροφορίες από ελλιπή παραδείγματα για την εργασία της μοντελοποίησης. Προκειμένου να υπολογιστεί η αντικειμενική συνάρτηση επιλέχθηκαν τα ακόλουθα δύο πιο κοινά μέτρα αξιολόγησης: (α) η ρίζα του μέσου τετραγωνικού σφάλματος (root mean square error) και (β) η ακρίβεια ταξινόμησης (classification accuracy).

7.2.4 Διακριτοποίηση

Η διαδικασία της (discretization) βοηθά στη μετατροπή του χώρου των χαρακτηριστικών, από ένα χωρό πραγματικών τιμών σε ένα σταθερό αριθμό ξεχωριστών διακριτών τιμών. Ένας μεγάλος αριθμός εφικτών τιμών χαρακτηριστικών θα μπορούσε να οδηγήσει σε απαιτητικούς από άποψη χρόνου αλγορίθμους μάθησης. Αξίζει να σημειώσουμε ότι, παρά την πρόοδο που έχει σημειωθεί κατά τα τελευταία χρόνια, η επιλογή του αριθμού των διαστημάτων ή των σημείων κοπής στη διαδικασία διακριτοποίησης παραμένει ένα ανοιχτό πρόβλημα.

Οι Taha και Asadi [378] πρότειναν μια εξελικτική προσέγγιση για τη διαδικασία διακριτοποίησης χρησιμοποιώντας δύο στόχους. Η πρώτη αντικειμενική συνάρτηση ελαχιστοποιεί το σφάλμα ταξινόμησης, ενώ η δεύτερη ελαχιστοποιεί τον αριθμό των σημείων κοπής (cut points).

7.2.5 Μη ισορροπημένα σύνολα δεδομένων

Η ιδανική κατάσταση για ένα επιτηρούμενο προγνωστικό μοντέλο είναι η γενίκευση σε άγνωστα αντικείμενα οποιασδήποτε κλάσης με την ίδια ακρίβεια. Σε προβλήματα του πραγματικού κόσμου, οι μαθητές καταπιάνονται με μη ισορροπημένα σύνολα δεδομένων (imbalanced datasets) [61]. Κάτι τέτοιο δεν είναι επιθυμητό, καθώς καθιστά τον αλγόριθμο Μηχανικής Μάθησης «υποκειμενικό» ως προς μια τάξη. Αυτό μπορεί να συμβεί όταν μια τάξη υποεκπροσωπείται σε μεγάλο βαθμό στο σύνολο εκπαίδευσης σε σχέση με τις άλλες.

Οι αλγόριθμοι στην (inductive machine learning) συνήθως σχεδιάζονται για να ελαχιστοποιούν μια προκαθορισμένη μετρική σε ένα σύνολο δεδομένων εκπαίδευσης. Επιπλέον, εάν κάποια τάξη περιέχει ένα μικρό αριθμό παραδειγμάτων, στις περισσότερες περιπτώσεις, μπορεί να αγνοηθεί από τους αλγόριθμους μάθησης. Αυτό οφείλεται στο γεγονός ότι, το κόστος της καλής απόδοσης στην υπερεκπροσωπούμενη τάξη υπερβαίνει το κόστος της κακής απόδοσης στη μικρότερη κατηγορία. Πρόσφατα προτάθηκε ένας μ - $\mu\mu$ μ μ (convex-hull-based multi-objective genetic programming algorithm) [439]. Αυτός ο αλγόριθμος εφαρμόστηκε σε δυαδικές περιπτώσεις ταξινόμησης και κατάφερε να επιτύχει τη μεγιστοποίηση της κυρτής θήκης ελαχιστοποιώντας τον ψευδώς θετικό ρυθμό (false positive rate) και μεγιστοποιώντας ταυτόχρονα τον πραγματικό θετικό ρυθμό (true positive rate). Η καμπύλη receiver operating characteristic (ROC) curve χρησιμοποιήθηκε ως εκτίμηση της απόδοσης και για την καθοδήγηση της διαδικασίας αναζήτησης.

Στην προσπάθειά τους να βελτιώσουν το 2D ROC χώρο, οι συγγραφείς του άρθρου [439] ενσωμάτωσαν την πολυπλοκότητα στους στόχους. Αυτό οδήγησε στη δημιουργία ενός 3D αντικειμενικού χώρου (σε αντίθεση με τον προηγούμενο χώρο 2D ROC). Τέλος, οι Li et al. [233] εφάρμοσαν το γνωστό αλγόριθμο PSO σε δύο πτυχές για την «εξισορρόπηση» των μη ισορροπημένων συνόλων δεδομένων. Μια πτυχή είναι η αναζήτηση για την κατάλληλη ποσότητα των πλειοψηφικών παραδειγμάτων, ενώ η άλλη είναι η εκτίμηση των καλύτερων παραμέτρων ελέγχου, όπως η ένταση (intensity) και η απόσταση των γειτόνων των μειονοτικών δειγμάτων (minority samples) προκειμένου τελικώς να συντεθούν.

7.3 Επιτηρούμενη μάθηση

Στη Μηχανική Μάθηση, η μ (classification) [218] είναι το πρόβλημα όπου ένας αλγόριθμος εκπαιδεύεται χρησιμοποιώντας ένα σύνολο εκπαίδευσης σωστά αναγνωρισμένων παραδειγμάτων με τέτοιο τρόπο, ώστε να δημιουργηθεί ένα μοντέλο που θα είναι σε θέση να αναγνωρίσει σωστά τα αντικείμενα που δεν έχουν ειδωθεί.

7.3.1 Δέντρα απόφασης

Τα (decision trees) ταξινομούν παραδείγματα ξεκινώντας από τη «ρίζα» του δέντρου και στη συνέχεια, τα ταξινομούν με βάση τις τιμές χαρακτηριστικών τους. Σε ένα κάθε μ (node) αντιπροσωπεύει ένα χαρακτηριστικό ενός στιγμιότυπου το οποίο πρέπει να ταξινομηθεί, ενώ κάθε (branch) αντιπροσωπεύει μια τιμή που μπορεί να έχει ο κόμβος.

Ο Zhao [438] πρότεινε μια προσέγγιση πολλαπλών κριτηρίων που στηρίζεται στο γενετικό προγραμματισμό προκειμένου να αναπτυχθεί ένα Pareto. Αυτή η υλοποίηση επιτρέπει στο χρήστη να επιλέγει προτεραιότητες για τις αντικρουόμενες αντικειμενικές συναρτήσεις του προβλήματος, όπως τα μ έναντι μ (false negative versus false positive), μ έναντι μ (sensitivity versus specificity) και μ έναντι μ (recall versus precision.).

Ο Fieldsend [129] χρησιμοποίησε τη μέθοδο βελτιστοποίησης PSO [211, 296] προκειμένου να εκπαιδεύσει το βέλτιστο δέντρο απόφασης χρησιμοποιώντας τη φόρμουλα πολλαπλών αντικειμένων για την αντιμετώπιση των ποσοστών σφάλματος σε κάθε κατηγορία.

Οι συγγραφείς του άρθρου [30] πρότειναν έναν γενετικό αλγόριθμο για τη δημιουργία δέντρων απόφασης που ονομάζεται LEGAL-Tree. Συγκεκριμένα, πρότειναν μια λεξικογραφική προσέγγιση, όπου οι πολλαπλοί στόχοι αξιολογούνται με μια σειρά προτεραιότητας.

Μια ταξινόμηση στην οποία ο αναγνώστης συναντά μια ομαδοποίηση των έργων που εξελίσσουν τα δέντρα απόφασης χρησιμοποιώντας εξελικτικούς αλγόριθμους παρουσιάζεται στο άρθρο [29]. Οι Chikalov et al. [68] δημιούργησαν προβλήματα βελτιστοποίησης δύο-

κριτηρίων για δέντρα απόφασης. Οι συγγραφείς του άρθρου εξέτασαν διαφορετικές συναρτήσεις κόστους, όπως τον μ μ , το μ και το μ . Συγκεκριμένα, σχεδιάζουν αλγορίθμους που είναι σε θέση να προσδιορίσουν τα βέλτιστα σημεία Pareto για έναν συγκεκριμένο πίνακα αποφάσεων.

7.3.2 Κανόνες μάθησης

Οι μ μ (classification rules) [141, 413] αντιπροσωπεύουν κάθε κλάση με τη μ (disjunctive normal form). Ο στόχος είναι να βρεθεί το μικρότερο σύνολο κανόνων που είναι σύμφωνο με το σύνολο εκπαίδευσης. Πολλοί παραγόμενοι κανόνες είναι συνήθως ένα «σημάδι» (ένδειξη) ότι ο αλγόριθμος μάθησης υπερ-προσαρμόζεται στα δεδομένα εκπαίδευσης.

Οι Dehuri et al. [89] παρουσίασαν έναν ελιτιστικού χαρακτήρα γενικό αλγόριθμο πολλαπλών στόχων (elitist multi-objective genetic algorithm (EMOGA)) για την παραγωγή κανόνων ταξινόμησης. Πρότειναν έναν γενικό αλγόριθμο πολλαπλών στόχων με έναν υβριδικό τελεστή διασταύρωσης για την ταυτόχρονη βελτιστοποίηση των εξής στόχων: της μ (comprehensibility), της μ (accuracy) και του μ (interestingness of rules).

Οι Pappa και Freitas [285] παρήγαγαν, επίσης με επιτυχία, μοντέλα με ακρίβεια καθώς και «συμπαγείς» κανόνες χρησιμοποιώντας έναν αλγόριθμο γενετικού προγραμματισμού βασισμένο σε γραμματικές πολλών αντικειμένων (multi-objective grammar-based genetic programming).

Οι Srinivasan και Ramakrishnan [374] αντιμετώπισαν το πρόβλημα της εύρεσης κατάλληλων κανόνων ως ένα πρόβλημα βελτιστοποίησης πολλαπλών στόχων. Συγκεκριμένα, χρησιμοποίησαν μια προσέγγιση που ως στόχο είχε την βελτιστοποίηση τριών αντικειμένων. Αυτές οι αντικειμενικές ποσότητες ήταν η μ (accuracy), η μ (comprehensibility) και η μ (novelty).

Ο Rudzinski [333] παρουσίασε μια πολυ-αντικειμενική γενετική προσέγγιση για να παράξει μ μ των δεδομένων (interpretability-oriented fuzzy rules). Η προτεινόμενη προσέγγισή τους, επιτρέπει στο χρήστη να αποκτά συστήματα με διάφορα επίπεδα «συμβιβασμού» μεταξύ μ και μ .

7.3.3 Bayesian ταξινομητές

Ένα Bayesian (Bayesian network (BN)) [179, 426], είναι ένα γραφικό μοντέλο για πιθανοτικές σχέσεις μεταξύ των μεταβλητών. Η δομή S ενός τέτοιου δικτύου είναι ένα μ μ (directed acyclic graph). Οι κόμβοι του S είναι σε ένα-προς-ένα αντιστοιχία με τις μεταβλητές και τα τόξα του γραφήματος αντιπροσωπεύουν επιρροές μεταξύ των μεταβλητών. Η έλλειψη πιθανών τόξων στο S αντιπροσωπεύει την υπό όρους ανεξαρτησία, ενώ ένας κόμβος (μεταβλητή) είναι υπό όρους ανεξάρτητος από τους μη-απογόνους του δεδομένου των γονέων του.

Οι Rodriguez και Lozano [328], εισήγαγαν μια διαρθρωτική προσέγγιση μάθησης ενός πολυδιάστατου Bayesian μαθητή με βάση τον γρήγορο και ελιτιστικό αλγόριθμο NSGA-II [88].

Ο αλγόριθμος ENORA, ο οποίος είναι ένας πολυ-αντικειμενικός εξελικτικός αλγόριθμος για την επιλογή χαρακτηριστικών και αφορά προβλήματα ταξινόμησης πολλαπλών κατηγοριών, χρησιμοποιήθηκε στο άρθρο [283]. Συγκεκριμένα, ο συγγραφέας υπολόγισε τους μέσους εκτιμητές 1 εξάρτησης του απλού Bayes μοντέλου, μέσω του προαναφερθέντος αλγορίθμου. Το προτεινόμενο σχήμα δοκιμάστηκε σε είκοσι ένα σύνολα δεδομένων πραγματικού κόσμου και τα πειραματικά αποτελέσματα δείχνουν ότι, η εφαρμογή της μεθόδου είναι πολλά υποσχόμενη από πλευράς χρόνου και ακρίβειας.

7.3.4 Μηχανές διανυσματικής υποστήριξης

Μια μηχανή διανυσματικής υποστήριξης (support vector machine (SVM)) [178, 349] είναι ένα μοντέλο ταξινόμησης, που βασίζεται στη θεωρία της ελαχιστοποίησης του κινδύνου (structured risk minimization theory). Η επιλογή των παραμέτρων C πυρήνων και γ , σε ένα μοντέλο SVM, είναι ζωτικής σημασίας για την παραγωγή ενός αποδοτικού μοντέλου διανυσματικής υποστήριξης. Η παράμετρος C του πυρήνα της ακτινικής συνάρτησης βάσης (radial basis function (RBF)) σε ένα μοντέλο SVM, συμβιβάζει την εσφαλμένη ταξινόμηση των παραδειγμάτων εκπαίδευσης, σε αντίθεση με την απλότητα της επιφανείας απόφασης. Μια χαμηλή τιμή της παραμέτρου C προκαλεί την ομαλή απόφαση, ενώ μια υψηλή τιμή του C προσπαθεί να ταξινομήσει σωστά όλα τα παραδείγματα εκπαίδευσης, παρέχοντας στο μοντέλο την ελευθερία να επιλέξει περισσότερα δείγματα ως φορείς υποστήριξης. Οι παράμετροι C μπορούν να θεωρηθούν ως το αντίστροφο της ακτίνας επιρροής των δειγμάτων, τα οποία επιλέγονται από το μοντέλο ως φορείς υποστήριξης.

Οι συγγραφείς του άρθρου [26], χρησιμοποίησαν ένα τεχνητό μη υποκειμενο αλγόριθμο πολλαπλών στόχων (multi-objective artificial immune algorithm), προκειμένου να βελτιστοποιηθεί ο αριθμός των πυρήνων, καθώς και των υπολοίπων παραμέτρων στο μοντέλο SVM. Οι Miranda et al. [263], πρότειναν μια υβριδική αρχιτεκτονική πολλαπλών αντικειμένων, η οποία συνδυάζει τη μετα-μάθηση με αλγορίθμους βελτιστοποίησης με χρήση σωματιδίων πολλαπλών αντικειμένων, προκειμένου να αντιμετωπιστεί το πρόβλημα επιλογής παραμέτρων στις μηχανές διανυσματικής υποστήριξης.

Ένας αλγόριθμος χωρικής διαμέρισης δύο παραμέτρων (bi-parameter space partition algorithm), που είναι σε θέση να ταυριάζει όλες τις λύσεις για κάθε ζεύγος παραμέτρων, προτάθηκε για τα μοντέλα SVM. Με βάση τη διαμέριση του χώρου, πρότειναν έναν αλγόριθμο K fold διασταυρούμενης επικύρωσης (K -fold cross-validation) για τον υπολογισμό των ολικών βέλτιστων ζευγαριών των παραμέτρων.

Τέλος, στην εργασία [330], συναντάμε ένα εξελικτικό πολυ-αντικειμενικό μοντέλο για την επιλογή παραδειγμάτων και συνεπώς, για την παραγωγή μιας συνεργατικής μεθόδου με βάση το σύνολο του Pareto. Στόχος των συγγραφέων του άρθρου ήταν να ελαχιστοποιήσουν το μέγεθος των δεδομένων εκπαίδευσης και να μεγιστοποιήσουν την ακρίβεια ταξινόμησης από τα κατάλληλα παραδείγματα επιλογής.

7.3.5 Τεχνητά νευρωνικά δίκτυα

Είναι γνωστό ότι τα μοντέλα perceptrons [338, 340] είναι σε θέση να ταξινομήσουν μόνο γραμμικά διαχωρίσιμα σύνολα δεδομένων. Αν οι περιπτώσεις δεν είναι γραμμικά διαχωρίσιμες, η μάθηση δεν θα βρει ποτέ ένα υπερ-επίπεδο (hyperplane), για το οποίο όλα τα παραδείγματα να έχουν ταξινομηθεί σωστά. Για το σκοπό αυτό, έχουν προταθεί τα πολλαπλάστρα perceptrons (multilayered perceptrons) (τεχνητά νευρωνικά δίκτυα), προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα.

Οι Tan et al. [379] χρησιμοποίησαν έναν τροποποιημένο μικρο-γενετικό αλγόριθμο βελτιστοποίησης (modified micro-genetic algorithm optimizer) για δύο πτυχές. Η μία πτυχή σχετιζόταν με την επιλογή ενός μικρού αριθμού χαρακτηριστικών για ταξινόμηση, και η άλλη είχε να κάνει με τη βελτίωση της ακρίβειας του τεχνητού νευρωνικού δικτύου.

Οι Ojha et al. [277] πρότειναν ένα πολλαπλό-αντικειμενικό γενετικό πρόγραμμα (multi-objective genetic program (MOGP)), προκειμένου να δημιουργηθεί ένα ετερογενές «εύκαμπτο» νευρωνικό δέντρο (heterogeneous flexible neural tree), το οποίο είναι ένα μοντέλο νευρωνικών δικτύων εμπρόσθιας τροφοδότησης.

7.3.6 Lazy Learners

Ο αλγόριθμος των k πλησιέστερων γειτόνων (k nearest neighbor (k NN)) [77,252] βασίζεται στην αρχή ότι τα παραδείγματα σε ένα σύνολο δεδομένων θα έχουν γενικά παρόμοιες ιδιότητες. Έτσι, ο αλγόριθμος k NN βρίσκει τα k πλησιέστερα παραδείγματα στο παράδειγμα δοκιμής και προβλέπει την τάξη του, αναγνωρίζοντας την πιο συχνή κατηγορία. Η παραγωγή πρωτοτύπων (prototype generation) είναι η δημιουργία ενός μικρού συνόλου παραδειγμάτων, που αντικαθιστούν τα αρχικά δεδομένα, προκειμένου να χρησιμοποιηθούν από τον αλγόριθμο k NN για ταξινόμηση. Οι κύριες πτυχές που πρέπει να εξεταστούν κατά την εφαρμογή μιας μεθόδου παραγωγής πρωτοτύπων είναι οι ακόλουθες:

- (α) Πρέπει να ληφθεί υπόψη η ακρίβεια ενός ταξινομητή k NN χρησιμοποιώντας τα πρωτότυπα και
- (β) Το ποσοστό της μείωσης του συνόλου δεδομένων.

Και οι δύο παράγοντες βρίσκονται σε σύγκρουση και έτσι, αυτό το πρόβλημα μπορεί να αντιμετωπιστεί φυσικά, με τεχνικές βελτιστοποίησης πολλών αντικειμένων.

Οι Escalante et al. [118] πρότειναν έναν πολυ-αντικειμενικό εξελικτικό αλγόριθμο για την παραγωγή πρωτοτύπων, που ονομάζεται MOPG. Επιπλέον, οι Hu και Tan [191] παρουσίασαν μια γενιά πρωτοτύπων χρησιμοποιώντας μια τεχνική βελτιστοποίησης πολλαπλών στόχων με χρήση της γνωστής μεθόδου PSO για τον k NN ταξινομητή.

7.3.7 Συνεργατικές μέθοδοι

Η επιλογή ενός και μόνο αλγορίθμου για την παραγωγή ενός αξιόπιστου μοντέλου ταξινόμησης δεν είναι εύκολη υπόθεση. Μια απλή προσέγγιση θα μπορούσε να είναι η εκτίμηση της ακρίβειας των υποψηφίων αλγορίθμων για ένα πρόβλημα, και στη συνέχεια, η επιλογή του καλύτερου ταξινομητή. Η ιδέα του συνδυασμού ταξινομητών, ή αλλιώς των συνεργατικών μεθόδων, [95] προτείνεται ως μια κατεύθυνση για την αύξηση της ακρίβειας ταξινόμησης σε προβλήματα πραγματικού κόσμου. Στην περίπτωση αυτή, ο στόχος είναι να χρησιμοποιηθούν τα πλεονεκτήματα ενός μοντέλου για να συμπληρωθούν οι αδυναμίες του άλλου. Σε γενικές γραμμές, οι πολυ-αντικειμενικοί εξελικτικοί αλγόριθμοι για την κατασκευή συνεργατικών ταξινομητών είναι μια αρκετά ενδιαφέρουσα περιοχή για μελέτη και έρευνα.

Οι Chandra και Yao [60] παρουσίασαν έναν μ μ (ensemble learning algorithm), ο οποίος ονομάζεται DIVACE (ποικιλόμορφο και ακριβές συνεργατικό σχήμα ταξινομητών DIVERse and ACCurate Ensemble learning algorithm). Αυτή η μέθοδος προσπαθεί να βρει μια ισορροπία μεταξύ διαφορετικότητας και παρουσιάζεται αναλυτικά στο άρθρο [204].

Οι συγγραφείς του άρθρου [40] πρότειναν μια μέθοδο πολλών αντικειμένων με βάση το γενετικό προγραμματισμό, προκειμένου να «εξελιχθούν» ακριβείς και διαφορετικούς ταξινομητές με αποδεκτή ακρίβεια, τόσο για τη μειονοτική όσο και για την πλειοψηφική τάξη του συνόλου δεδομένων. Επιπλέον, οι Bhowan et al. [41] παρουσίασαν μια παρόμοια προσέγγιση, προκειμένου να παρουσιαστούν συνεργατικές μέθοδοι, χρησιμοποιώντας τον γενετικό προγραμματισμό για μη ισορροπημένα σύνολα δεδομένα.

Οι Nguyen et al. [274] χρησιμοποίησαν μια προσέγγιση γενετικού αλγορίθμου που επικεντρώνεται στους ακόλουθους τρεις στόχους, οι οποίοι σχετίζονται με τον αριθμό:

1. των σωστά ταξινομημένων παραδειγμάτων,
2. των επιλεγμένων χαρακτηριστικών και
3. των επιλεγμένων ταξινομητών.

Οι Gu et al. [169] παρουσίασαν μια έρευνα σχετικά με τις πολυ-κριτηριακές συνεργατικές μεθόδους, συμπεριλαμβανομένων των μέτρων ποικιλομορφίας, των μελών της γενιάς, καθώς και αυτών των τεχνικών επιλογής και ενσωμάτωσης.

Οι Nag και Pal [269] παρουσίασαν έναν ολοκληρωμένο αλγόριθμο για την ταυτόχρονη επιλογή χαρακτηριστικών και την συμπερίληψη διαφορετικών μαθητών, χρησιμοποιώντας έναν αλγόριθμο γενικού προγραμματισμού πολλαπλών στόχων σταθερής κατάστασης (steady state multi-objective genetic programming), ο οποίος ελαχιστοποιεί τους ακόλουθους τρεις στόχους:

- (α) τις προβλέψεις ψευδών θετικών,
- (β) τις προβλέψεις ψευδών αρνητικών και
- (γ) τον αριθμό των κόμβων στα φύλλα του δέντρου απόφασης.

Οι Albukhanajer et al. [6] προτείνουν συνεργατικούς ταξινομητές που χρησιμοποιούν πολλαπλά χαρακτηριστικά σύμφωνα με το σύνολο Pareto για την αναγνώριση εικόνων χωρίς αλλοίωση.

Τέλος, εξίσου σημαντικό, είναι το ότι πολύ πρόσφατα, οι Pourtaheri et al. [312], ανέπτυξαν δύο ευρετικές συνεργατικές μεθόδους πολλαπλών αντικειμένων συνδυάζοντας τον αλγόριθμο βελτιστοποίησης κεκλιμένων επιπέδων πολλαπλών στόχων και τον αλγόριθμο βελτιστοποίησης πολλών αντικειμένων με βάση τη μέθοδο PSO.

7.4 Μη επιτηρούμενη μάθηση

7.4.1 Συσταδοποίηση

Η μ ή (cluster analysis) [286] είναι μια διαδικασία που είναι πολύ χρήσιμη για την «εξερεύνηση» μιας συλλογής δεδομένων. Όπως υποδηλώνεται από τον όρο «συστάδα», μέσω αυτής της διαδικασίας ανιχνεύονται στοιχεία ή χαρακτηριστικά με αλληλεξάρτηση, τα οποία μπορούν να οδηγήσουν στην ομοιογενή ομαδοποίηση των δεδομένων. Μπορεί να είναι, είτε (supervised) είτε χωρίς επίβλεψη (unsupervised), και η μεγάλη διαφορά μεταξύ αυτής της διεργασίας και της διαδικασίας ταξινόμησης είναι ότι, η πρώτη δεν χρησιμοποιεί ετικέτες για να βοηθήσει στην κατηγοριοποίηση των δεδομένων προκειμένου να δημιουργήσει μια δομή συστάδων. Επιπλέον, αν ένα αντικείμενο ανήκει σε μια συγκεκριμένη ομάδα προσδιορίζεται μέσω μ - μ (intra-connectivity). Αν αυτό το μέτρο είναι υψηλό, αυτό σημαίνει ότι οι συστάδες είναι «συμπαγείς» και τα δεδομένα της ίδιας ομάδας εξαρτώνται σε μεγάλο βαθμό το ένα από το άλλο. Από την άλλη πλευρά, η μ (inter-connectivity) είναι ένα κριτήριο που δηλώνει την ανεξαρτησία μεταξύ των συστάδων. Έτσι, αν η διασύνδεση είναι χαμηλή, αυτό σημαίνει ότι οι μεμονωμένες συστάδες είναι σε μεγάλο βαθμό ανεξάρτητες μεταξύ τους. Για περισσότερες λεπτομέρειες για αλγόριθμους εξελικτικής συσταδοποίησης πολλαπλών αντικειμένων, παραπέμπουμε τον αναγνώστη στις ακόλουθες εργασίες [107,267].

Ο μ μ μ μ (mathematical programming) [120] έχει σημαντική συμβολή στο ζήτημα της ανάλυσης συσταδοποίησης. Η άμεση σύνδεση των δύο περιοχών μπορεί εύκολα να γίνει κατανοητή, καθώς απαιτείται ο ελάχιστος αριθμός συστάδων, στα οποία μπορεί να ομαδοποιηθεί το αρχικό σύνολο δεδομένων. Έτσι, αυτή η προσέγγιση μπορεί να θεωρηθεί ως πρόβλημα βελτιστοποίησης, με συγκεκριμένα χαρακτηριστικά και περιορισμούς. Ένα σημαντικό ζήτημα είναι επίσης, η κατάλληλη επιλογή λύσεων, αφού από μια σειρά εφικτών, «καλών» λύσεων, οι καλύτερες λύσεις είναι αυτές που μας ενδιαφέρουν [288].

Οι Luo et al. [249] πρότειναν μια μέθοδο για τη μοντελοποίηση της φασματικής ομαδοποίησης (spectral clustering) και μέσω συγκεκριμένων τελεστών, επέλεξαν ένα σύνολο

καλών ατόμων κατά τη διαδικασία της βελτιστοποίησης. Επιπλέον, οι συγγραφείς του άρθρου, μέσω του (ratio cut criterion), επέλεξαν μια λύση «αντιστάθμισης» από το σύνολο Pareto. Τέλος, τα διάφορα προβλήματα που αναλύθηκαν για επιτηρούμενες και μη επιτηρούμενες διεργασίες ταξινόμησης συνέβαλαν στη δημιουργία τεχνικών ομαδοποίησης με ημι-επιτήρηση (semi-supervised). Με μια μικρή ποσότητα ετικετοποιημένων δεδομένων και την κατανομή των δεδομένων, οι Alok et al. [13] πρότειναν μια μέθοδο ημι-επιτηρούμενης ομαδοποίησης χρησιμοποιώντας το πλαίσιο βελτιστοποίησης πολλαπλών κριτηρίων.

Οι Wang et al. [408], πρόσφατα, μέσω ενός εξελικτικού αλγορίθμου πολλαπλών στόχων (evolutionary multi-objective (EMO)), αντιμετώπισαν μια πολύ δύσκολη και διαχρονική πρόκληση για ένα πρόβλημα συσταδοποίησης, δηλαδή τον « μ μ

». Προς το σκοπό αυτό, οι συγγραφείς του άρθρου πρότειναν ένα αλγοριθμικό σχήμα χρησιμοποιώντας έναν αλγόριθμο EMO, και συγκεκριμένα τον ταχύτατο και ελιτιστικό πολυ-αντικειμενικό γενετικό αλγόριθμο NSGA-II, προκειμένου αυτός να επιλέξει τις μη κυριαρχούμενες λύσεις. Η διαδικασία περιλαμβάνει ένα για την επιλογή του βέλτιστου αποτελέσματος ομαδοποίησης. Οι συγγραφείς εξέτασαν το μοντέλο τους σε τρία σύνολα δεδομένων και τα πειραματικά τους αποτελέσματα δείχνουν ότι η μέθοδος EMO- k -clustering είναι αποτελεσματική και με μόνο μία εκτέλεση είναι σε θέση να αποκτήσει όλα τα αποτελέσματα ομαδοποίησης για διαφορετικές τιμές της παραμέτρου k .

Τελευταία, αλλά εξίσου σημαντική, η πολύ πρόσφατη εργασία που παρουσιάστηκε από τους Nayak et al. [272]. Συγκεκριμένα, πρότειναν ένα διαφοροεξελικτικό αλγόριθμο πολλαπλών αντικειμένων με βάση τον ελιτισμό (elitism-based multi-objective differential evolution) για την αυτόματη συσταδοποίηση. Το έργο τους χειρίζεται πολύπλοκα σύνολα δεδομένων χρησιμοποιώντας τρεις στόχους. Οι συγγραφείς του άρθρου διεξήγαγαν πειράματα σε δέκα σύνολα δεδομένων και τα αποτελέσματα δείχνουν ότι η προσέγγισή τους παρέχει μια εναλλακτική λύση για την ομαδοποίηση δεδομένων σε πολλές διαφορετικές περιοχές.

7.4.2 Κανόνες συσχέτισης

Η εξόρυξη κανόνων συσχέτισης (association rule mining (ARM)) [372] έχει ως πρωταρχικό στόχο την ανακάλυψη κανόνων συσχέτισης μεταξύ δεδομένων μιας βάσης δεδομένων. Ο πρώτος στόχος αυτής της διαδικασίας είναι να βρεθούν τα δεδομένα, που έχουν τη μεγαλύτερη εμφάνιση στη βάση δεδομένων. Στη συνέχεια, δημιουργούνται οι κατάλληλοι κανόνες συσχέτισης για ολόκληρο το σύνολο δεδομένων, χρησιμοποιώντας τις τιμές χαρακτηριστικών, που η εμφάνισή τους υπερβαίνει ένα συγκεκριμένο προκαθορισμένο όριο.

Οι Minaei-Bidgoli et al. [262] πρότειναν ένα γενετικό αλγόριθμο πολλαπλών στόχων για την εξόρυξη κανόνων συσχέτισης από αριθμητικές μεταβλητές. Είναι γνωστό ότι καλά και ευρέως χρησιμοποιούμενα μοντέλα, που χειρίζονται τη διεργασία εξόρυξης κανόνα συσχέτισης, δεν μπορούν να εφαρμοστούν σε σύνολα δεδομένων που αποτελούνται από αριθμητικά δεδομένα. Για το λόγο αυτό, είναι απαραίτητη η προεπεξεργασία των δεδομένων και ειδικότερα η διαδικασία διακριτοποίησης. Οι Minaei-Bidgoli et al. [262], χρησιμοποιώντας τρία μέτρα, συγκεκριμένα: την μ (confidence), το (interestingness) και την (comprehensibility), ορίζοντας τρεις διαφορετικές αντικειμενικές συναρτήσεις για την προσέγγισή τους, εξήγαγαν τους καλύτερους κανόνες συσχέτισης μέσω της βέλτιστης εκτίμησης του συνόλου Pareto.

Οι Beiranvand et al. [33] πρότειναν ένα πολυ-αντικειμενικό μοντέλο βελτιστοποίησης με χρήση σμήνους σωματιδίων, με όνομα MOPAR, για την εξόρυξη αριθμητικών κανόνων συσχέτισης, σε ένα μόνο βήμα, χωρίς εκ των προτέρων διακριτοποίηση. Οι συγγραφείς του άρθρου διεξήγαγαν πειράματα και τα αποτελέσματα δείχνουν ότι η προσέγγισή τους εξάγει αξιόπιστους, κατανοητούς και ενδιαφέροντες αριθμητικούς κανόνες συσχέτισης.

Οι Martin et al. [256] πρότειναν ένα πολυ-αντικειμενικό εξελικτικό μοντέλο με την ονομασία QAR-CIP-NSGA-II, το οποίο επεκτείνει τον γνωστό ελιτιστικό πολυ-αντικειμενικό

γενετικό αλγόριθμο NSGA-II [88]. Η μέθοδός τους εκτελεί μια εξελικτική διαδικασία μάθησης και μία συνθήκη επιλογής για κάθε κανόνα συσχέτισης. Επιπλέον, η προσέγγισή τους μεγιστοποιεί δύο από τις τρεις αντικειμενικές συναρτήσεις που θεώρησαν οι Minaei-Bidgoli et al.. Επιπλέον, η προσέγγισή τους μεγιστοποιεί την απόδοση των αντικειμενικών συναρτήσεων για την εξόρυξη ενός συνόλου ποσοτικών κανόνων συσχέτισης με επαρκή ερμηνεία, καθώς και ανακριβή αποτελέσματα.

7.5 Πρόσφατες εφαρμογές

Οι διάφορες εφαρμογές, που παρέχονται τα τελευταία χρόνια, δείχνουν τη σημασία των μ μ (multi-objective evolutionary optimization algorithms (MOEEOA)). Μερικές από τις πιο πρόσφατες και πολύ ενδιαφέρουσες εφαρμογές των αλγορίθμων εξελικτικής βελτιστοποίησης πολλών στόχων είναι αυτές που ακολουθούν.

Οι Mason et al. [257] ανέπτυξαν ένα τεχνητό νευρωνικό δίκτυο, το οποίο εκπαιδεύτηκε μέσω ενός μ (differential evolution (DE)). Το προτεινόμενο νευρωνικό δίκτυο έχει τη δυνατότητα να χειρίζεται προβλήματα βελτιστοποίησης πολλαπλών αντικειμένων χρησιμοποιώντας κατάλληλα μια συνάρτηση προσέγγισης. Συγκεκριμένα, η προτεινόμενη προσέγγιση χρησιμοποιεί έναν ολικό βελτιστοποιητή μίας αντικειμενικής (τον αλγόριθμος DE), προκειμένου να «εξελιχθεί» το νευρωνικό δίκτυο. Με άλλα λόγια, ο λεγόμενος αλγόριθμος MONNDE, είναι ικανός να παρέχει περαιτέρω μ (Pareto fronts) χωρίς παραπάνω προσπάθεια για βελτιστοποίηση. Οι συγγραφείς του άρθρου, εφάρμοσαν τον αλγόριθμο MONNDE στο γνωστό πρόβλημα μ μ (dynamic economic emission dispatch) και μέσω των πειραμάτων που πραγματοποίησαν, δείχνουν ότι η απόδοση του αλγορίθμου τους είναι εξίσου βέλτιστη σε σύγκριση με άλλους γνωστούς και ευρέως χρησιμοποιούμενους αλγορίθμους αυτής της κατηγορίας. Επιπλέον, δείχνουν ότι είναι αποτελεσματικότερο να βελτιστοποιείται δυναμικά η τοπολογία του νευρωνικού δικτύου με έναν άμεσο τρόπο, αντί να βελτιστοποιούνται τα βάρη του νευρωνικού δικτύου.

Οι Rao et al. [321] πρότειναν έναν εναλλακτικό μ (classifier for disease diagnosis). Συγκεκριμένα, το προτεινόμενο σχήμα περιλαμβάνει μια ακολουθιακή διαδικασία ελαχιστοποίησης, ως ταξινομητή τις μηχανές διανυσματικής υποστήριξης και τρεις εξελικτικούς αλγόριθμους για την εξέλιξη των παραμέτρων. Επιπλέον, οι συγγραφείς του άρθρου παρουσίασαν μια νέα τεχνική, η οποία ονομάζεται μ (cuboids elephant herding optimization (CEHO)). Η προσέγγισή τους εφαρμόζεται σε δεκαεπτά ιατρικά σύνολα δεδομένων και τα πειραματικά αποτελέσματα δείχνουν ότι, η προτεινόμενη τεχνική παρουσιάζει πολύ καλή απόδοση για όλα τα δοκιμασμένα σύνολα δεδομένων.

Κλείνοντας την ενότητα των εφαρμογών, οι Sabar et al. [334] εξέτασαν τη διαμόρφωση ενός μοντέλου SVM ως πρόβλημα βελτιστοποίησης δύο αντικειμένων. Η ακρίβεια του μοντέλου ήταν ο πρώτος στόχος, ενώ ο άλλος ήταν η πολυπλοκότητα του μοντέλου. Οι συντάκτες του άρθρου πρότειναν ένα νέο υπερ-ευρετικό πλαίσιο για τη βελτιστοποίηση των δύο προαναφερόμενων αντικρουόμενων στόχων. Η εξελιγμένη προσέγγιση εξετάστηκε σε δύο προβλήματα ασφαλείας στον κυβερνοχώρο και τα πειραματικά αποτελέσματα δείχνουν ότι, το προτεινόμενο σχήμα είναι πολύ αποδοτικό και αποτελεσματικό.

7.6 Σύνοψη

Γενικά, το αντικείμενο των μ μ (multi-objective evolutionary optimization algorithms (MOEEOA)) σχετίζεται με μια ενδιαφέρουσα ιδέα, με πολλές διαφορετικές πτυχές και έναν κρίσιμο ρόλο, όχι μόνο στη Μηχα-

νική Μάθηση αλλά και σε πολλά άλλα επιστημονικά πεδία. Αυτό είναι προφανές, καθώς σήμερα υπάρχει η ανάγκη αντιμετώπισης αντιφατικών στόχων απόδοσης σε πολλά επιστημονικά πεδία. Η πληθώρα των άρθρων που έχουν δημοσιευθεί σχετικά με αυτού του είδους τους αλγορίθμους δείχνουν εμφατικά ότι, η επιστημονική κοινότητα έχει στρέψει δυναμικά την προσοχή της στο συγκεκριμένο θέμα.

Οι πρώτες εξελικτικές προσεγγίσεις για την επίλυση προβλημάτων βελτιστοποίησης πολλών κριτηρίων και ιδιαίτερα οι αλγόριθμοι βελτιστοποίησης με χρήση σμήνους σωματιδίων και οι αλγόριθμοι διαφοροεξέλιξης εμφανίστηκαν πολύ ελπιδοφόρες [291, 293, 295-297]. Αξίζει να σημειωθεί ότι, οι παράγοντες που εκτιμώνται με τη βοήθεια των δύο παραπάνω αλγορίθμων [291, 297], παραμένουν η βάση της σύγχρονης έρευνας γύρω από τη μ , τη μ (δύο, τριών ή και παραπάνω αντικειμενικών ποσοτήτων) και τη μ . Επιπλέον, η πολυ-κριτηριακή βελτιστοποίηση έχει οδηγήσει σε μοντέλα Μηχανικής Μάθησης με καλύτερη απόδοση, σε αντίθεση με τα παραδοσιακά μοντέλα βελτιστοποίησης ενός μόνο στόχου.

Η σημασία των πολυ-αντικειμενικών εξελικτικών αλγορίθμων είναι εμφανής, όχι μόνο από την πληθώρα των άρθρων που έχουν παρουσιαστεί από την επιστημονική κοινότητα, αλλά και από ένα τεράστιο αριθμό διαφόρων εφαρμογών που παρουσιάστηκαν τις τελευταίες δεκαετίες, όπως στη μηχανική [159], στη βιομηχανία [229], στην οικονομία [73, 234] και σε πολλούς άλλους τομείς της επιστήμης, της τεχνολογίας και της βιομηχανίας [4, 98, 143, 155]. Ο αναγνώστης θα μπορούσε, επίσης, να αντλήσει περισσότερες λεπτομέρειες σχετικά με την ποικιλία των προβλημάτων και το μέγεθος των εφαρμογών από τα άρθρα [212] και [403].

Αυτό το κεφάλαιο, περιγράφει τον τρόπο με τον οποίο χρησιμοποιούνται οι πολυ-αντικειμενικοί εξελικτικοί αλγόριθμοι βελτιστοποίησης στον τομέα της Μηχανικής Μάθησης, με τις σχετικές λεπτομέρειες. Πρέπει να σημειωθεί ότι, ο κατάλογος των αναφορών στο τρέχον κεφάλαιο δεν παρέχει έναν πλήρη κατάλογο των εργασιών που αντιστοιχούν σε αυτό το θέμα. Ο κύριος στόχος ήταν να δοθεί μια επισκόπηση των βασικών ιδεών, παρά ενός απλού καταλόγου, όλων των ερευνητικών εργασιών που έχουν συζητηθεί ή έχουν χρησιμοποιήσει αυτές τις ιδέες. Παρ' όλα αυτά, ελπίζουμε ότι οι εργασίες που παρουσιάζουμε, θα καλύψουν τα σημαντικότερα θεωρητικά ζητήματα και θα δώσουν κατευθυντήριες γραμμές στους κύριους κλάδους της βιβλιογραφίας, σχετικά με τέτοιες τεχνικές και αλγοριθμικά σχήματα, καθοδηγώντας τον αναγνώστη σε σύγχρονες ερευνητικές κατευθύνσεις.

Τέλος, χαρτογραφούνται οι περιοχές στις οποίες δύνανται να εφαρμοστούν οι μέθοδοι που αναπτύχθηκαν στο Κεφάλαιο 3. Ακόμα, εκτίθενται σημαντικοί αλγόριθμοι του χώρου με τους οποίους δύνανται να συνδυαστούν, ώστε να δημιουργήσουμε νέες μεθόδους ή να βελτιώσουμε την απόδοση υπάρχουσών τεχνικών. Επιπλέον, ο αναγνώστης μπορεί να λάβει μια βασική ιδέα για τις εφαρμογές πολυ-αντικειμενικής βελτιστοποίησης, στις οποίες μπορούν να εφαρμοστούν τόσο το υβριδικό σχήμα που παρουσιάστηκε στο Κεφάλαιο 5 όσο και στο Κεφάλαιο 6.

Μέρος V

Συμπεράσματα–Βιβλιογραφία– Δημοσιεύσεις

Σύνοψη

Συνάψεις όλα και ουχ όλα, συμφερόμενον
διαφερόμενον, συνάδον διάδον, και εκ
πάντων εν και εξ ενός πάντα.

— (544–484 . . .)

Σο κεφάλαιο αυτό, παρουσιάζουμε τα σημαντικότερα συμπεράσματα της παρούσας διδακτορικής διατριβής. Αντικείμενο της διδακτορικής διατριβής αποτέλεσε η ανάπτυξη και θεμελίωση νέων μεθόδων Υπολογιστικών Μαθηματικών στην Υπολογιστική Νοημοσύνη. Η τομή αυτών των δύο επιστημονικών πεδίων στην περίπτωση μας, βρίσκεται στα προβλήματα Βελτιστοποίησης που συναντώνται σε πληθώρα εφαρμογών, είτε αυτών καθαυτών των πεδίων είτε άλλων, που κατέληξαν σε προβλήματα Βελτιστοποίησης. Το τεράστιο εύρος εφαρμογών της Ολικής Βελτιστοποίησης και των επιστημονικών περιοχών που συναντώνται άμεσα ή έμμεσα στη βελτιστοποίηση, σηματοδοτεί τη μεγάλη σημασία και χρησιμότητά της στις σύγχρονες επιστήμες. Παρόλο που έχουν προταθεί πολλοί και διαφορετικοί αλγόριθμοι βελτιστοποίησης, ο συνεχώς αυξανόμενος όγκος εξαιρετικά πολύπλοκων εφαρμογών που σχετίζονται με διεργασίες της πραγματικής ζωής, καθιστά αναγκαία την ανάπτυξη νέων, καινοτόμων μεθοδολογιών μέσα από μαθηματική θεμελίωση και τεκμηρίωση. Οι τεχνικές και οι αλγόριθμοι αυτοί είναι επιθυμητό να αξιοποιούν τα ξεχωριστά χαρακτηριστικά των προβλημάτων με ένα βέλτιστο τρόπο, καθώς και να είναι ικανοί να προσαρμόζονται σε τυχόν παραλλαγές, ενσωματώνοντας ή συνδυάζοντας γνώση από πολλούς τομείς.

8.1 Σύνοψη και συμπεράσματα

Η δομή της διατριβής οργανώθηκε σε πέντε μέρη. Στο πρώτο μέρος παρουσιάζονται εισαγωγικές έννοιες, οι οποίες είναι χρήσιμες στον αναγνώστη για την κατανόηση της διατριβής, ενώ στα επόμενα μέρη συναντάται το ερευνητικό έργο και η συνεισφορά μας. Έτσι, στο δεύτερο μέρος παρουσιάζονται πολύ χρήσιμες έννοιες για τη μαθηματική Βελτιστοποίηση και σημαντικά συμπεράσματα του γνωστού θεωρήματος ‘No free lunch theorems for optimization’, ενώ χαρτογραφούνται οι περιοχές στις οποίες δύναται να δημιουργηθούν ‘free lunches’ για την κατασκευή βέλτιστων αλγορίθμων. Στη συνέχεια, παρουσιάστηκαν οι βασικότερες έννοιες που σχετίζονται με τις μεθόδους δυναμικών τροχιών ανίχνευσης και την εφαρμογή τους για την επίλυση διαφορετικών προβλημάτων Βελτιστοποίησης. Η κύρια συμβολή της διατριβής παρουσιάζεται μέσω της ανάπτυξης και θεμελίωσης μιας νέας οικογένειας μεθόδων δυναμικών τροχιών, όπως αυτή εμπνεύστηκε από τη μέθοδο των Snyman-Fatti και τις μεθόδους αριθμητικής επίλυσης συνήθων διαφορικών εξισώσεων Runge-Kutta. Στη σχετική ενότητα παρουσιάζεται αναλυτικά το θεώρημα και η απόδειξη σύγκλισης. Στο τρίτο μέρος, παρουσιάζεται ένα σημαντικό ζήτημα για τη Μηχανική Μάθηση, και συγκεκριμένα η προεπεξεργασία δεδομένων, για κάθε βήμα της προετοιμασίας του συνόλου δεδομένων που καλούμαστε να

διαχειριστούμε σε προβλήματα της Εξόρυξης Δεδομένων. Στο ίδιο μέρος παρουσιάζονται δύο νέες εμπειρικές μέθοδοι, όπου η πρώτη εφαρμόζεται σε ένα κοινό πρόβλημα που συναντάται στους αλγόριθμους της Μηχανικής Μάθησης, αυτό της αναγνώρισης ακραίων τιμών σε ένα σύνολο δεδομένων. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος είναι αξιόπιστη και επιτυγχάνει σθεναρή συμπεριφορά. Η δεύτερη μέθοδος που αναπτύχθηκε, ήταν μια συνεργατική μέθοδος που εφαρμόστηκε σε ένα επίσης, συχνά παρατηρούμενο πρόβλημα της Μηχανικής Μάθησης, αυτό της ταξινόμησης. Τα συμπεράσματα που λάβαμε ήταν ενθαρυντικά, δείχνοντας ότι το προτεινόμενο σχήμα επιτυγχάνει καλύτερα αποτελέσματα ταξινόμησης κατά μέσο όρο, σε σύγκριση με τις υπόλοιπες μεθόδους. Στο τέταρτο μέρος της διδακτορικής διατριβής, παρουσιάζεται ο «σύνδεσμος» των μεθόδων που χρησιμοποιούμε στη Μηχανική Μάθηση με τη μαθηματική Βελτιστοποίηση. Συγκεκριμένα, παρέχουμε μια μελέτη αναφορικά με το πρόβλημα της πολυ-αντικειμενικής Βελτιστοποίησης και τους εξελικτικούς αλγόριθμους που έχουν εφαρμοστεί για την προσέγγιση αυτού του προβλήματος. Με αυτό τον τρόπο χαρτογραφούνται οι περιοχές εφαρμογής των μεθόδων που αναπτύχθηκαν στο δεύτερο μέρος της διατριβής.

Παρακάτω παρουσιάζονται τα κύρια συμπεράσματα των ζητημάτων που πραγματεύονται σε κάθε κεφάλαιο της διατριβής, όπως και οι μελλοντικές προεκτάσεις.

Συγκεκριμένα, αναπτύσσουμε μια νέα οικογένεια μεθόδων για ολική Βελτιστοποίηση [12], βασισμένη σε αριθμητικές μεθόδους επίλυσης συνήθων διαφορικών εξισώσεων και συγκεκριμένα, τις ευρέως γνωστές μεθόδους της οικογένειας Runge-Kutta. Από την φύση της οικογένειας Runge-Kutta, οι προτεινόμενες μέθοδοι που μπορούν να παραχθούν είναι άπειρες στο πλήθος. Τα προτεινόμενα σχήματα τεκμηριώνεται ότι συγκλίνουν σε ένα τοπικό βέλτιστο μιας αντικειμενικής συνάρτησης, ξεκινώντας από σχεδόν οποιοδήποτε αρχικό σημείο μέσω του θεωρήματος και της απόδειξης που παρουσιάζουμε στο Κεφάλαιο 3 (και συγκεκριμένα στην Ενότητα 3.3 της διδακτορικής διατριβής). Ακόμα, αξίζει να σημειωθεί ότι δεν απαιτούνται πρόσθετοι υπολογισμοί της αντικειμενικής συνάρτησης ή της κλίσης της συνάρτησης, αφού η προτεινόμενη στρατηγική χρησιμοποιεί στοιχεία που έχουν ήδη υπολογιστεί. Παρά το γεγονός ότι οι μονοτονικές στρατηγικές σύγκλισης παρέχουν έναν αποδοτικό και αποτελεσματικό τρόπο για να εξασφαλιστεί ότι η συνάρτηση σφάλματος μειώνεται επαρκώς, έχουν το μειονέκτημα ότι καμία πληροφορία, που ενδέχεται να επιταχύνει τη σύγκλιση, δεν αποθηκεύεται, και ως εκ τούτου, δεν χρησιμοποιείται [127]. Για να μετριάσουμε αυτή την κατάσταση, μπορούμε να χρησιμοποιήσουμε μ μ (nonmonotone convergence strategy) που αξιοποιούν τις συσσωρευμένες πληροφορίες σε σχέση με τις πιο πρόσφατες τιμές της αντικειμενικής συνάρτησης. Τέτοιου είδους τεχνικές έχουν εφαρμοστεί αποδοτικά στην περίπτωση της εκπαίδευσης μ μ (multilayer feedforward neural networks), μια διαδικασία που μπορεί να θεωρηθεί ως ένα ιδιαίτερο μη γραμμικό πρόβλημα ελαχιστοποίησης.

Οι παραπάνω μέθοδοι εφαρμόστηκαν με επιτυχία σε διάφορες δοκιμαστικές συναρτήσεις ποικίλης διαστασιμότητας. Η εμπειρία μας είναι ότι οι μέθοδοι συμπεριφέρονται ικανοποιητικά και είναι αξιόπιστες. Τα αποτελέσματα που ελήφθησαν μπορούν να συγκριθούν με εκείνα της μεθόδου του Cauchy (μέθοδος απότομης καθόδου) και της μεθόδου συζυγών κλίσεων Fletcher-Reeves. Ας υπενθυμίσουμε σε αυτό το σημείο ότι, γενικά, οι συγκρίσεις διαφορετικών μεθόδων βελτιστοποίησης είναι ένα δύσκολο ζήτημα. Επίσης, λόγω του ότι «δεν υπάρχει δωρεάν γεύμα στη βελτιστοποίηση», όπως εκτενώς παρουσιάστηκε στο Κεφάλαιο 2, δεν υπάρχει ένας μόνον αλγόριθμος που να λειτουργεί καλά σε όλα τα προβλήματα και αν βελτιωθεί ένας αλγόριθμος για ένα συγκεκριμένο πρόβλημα, δεν θα αποδώσει καλά σε άλλα προβλήματα. Έτσι, κατά μέσο όρο για όλα τα προβλήματα βελτιστοποίησης, χωρίς επαναδειγματοληψία, όλοι οι αλγόριθμοι βελτιστοποίησης αποδίδουν εξίσου καλά [424]. Επιπλέον, είναι γνωστό ότι, για σχεδόν οποιοδήποτε ζεύγος αλγόριθμων και διάφορα μέτρα απόδοσης αλγόριθμου, όπως ο χρόνος εκτέλεσης ή η ποιότητα των λύσεων, κάθε αλγόριθμος

θα έχει καλύτερη απόδοση από άλλους για συγκεκριμένες εισόδους [332].

Στηριζόμενοι στα πολύ χρήσιμα συμπεράσματα τα οποία αναφέρονται και στην πρόσφατη εργασία [2], μπορεί κανείς να καταλάβει ότι σκοπός των ερευνητών είναι να αποκρυπτογραφήσουν την εγγενή πολυπλοκότητα των προβλημάτων του πραγματικού κόσμου, μέσω μεθόδων που θα ανταποκρίνονται καλά σε συγκεκριμένες οικογένειες προβλημάτων. Άρα, στοχεύουμε στην «απεγκλώβιση» της σχέσης αλγορίθμου-προβλήματος, ψάχνοντας για συγκεκριμένους αλγορίθμους που θα ξεπερνούν σε απόδοση άλλους για συγκεκριμένες οικογένειες προβλημάτων (π.χ. ελαχιστοποίηση συναρτήσεων δύο μεταβλητών) με παρεμφερή χαρακτηριστικά. Αυτό θα έχει ως αποτέλεσμα τη δημιουργία των λεγόμενων 'free lunches' για ειδικές συναρτήσεις και συνεπώς, προβλήματα. Ακόμα, τα συμπεράσματα που καταγράφονται, μάς επέτρεψαν να αναζητήσουμε καινούριες μεθόδους για την επίλυση αυτού του μείζονος προβλήματος. Απάντηση έρχεται να δώσει η ανάπτυξη αλγορίθμων με μαθηματική θεμελίωση και μαθηματική απόδειξη, καθώς ένας αλγόριθμος με ισχυρό μαθηματικό υπόβαθρο μπορεί αναμφίβολα να χρησιμοποιηθεί σε περισσότερες εφαρμογές με μεγαλύτερη επιτυχία. Προς αυτή την κατεύθυνση στηρίχθηκε ο πυρήνας της διατριβής, δηλαδή στη μαθηματική θεμελίωση αλγοριθμικών τεχνικών, οι οποίες θα απευθύνονται σε πλειάδα προβλημάτων, καλύπτοντας έτσι ένα σημαντικό κενό στην επιστημονική βιβλιογραφία ως προς την Υπολογιστική Νοημοσύνη και τα Υπολογιστικά Μαθηματικά. Από την εμπειρία μας και μέσα από την έρευνα που παρουσιάζεται στο δεύτερο μέρος της διατριβής για συγκεκριμένους τύπους προβλημάτων, υπάρχουν βέλτιστοι αλγόριθμοι. Επομένως, η προσοχή της ερευνητικής κοινότητας εστιάζεται στην εύρεση τέτοιων αλγορίθμων για συγκεκριμένες κατηγορίες προβλημάτων.

Όπως προαναφέραμε, η σύγχρονη βιβλιογραφία κατακλύζεται από ένα αναρίθμητο πλήθος αλγορίθμων και τεχνικών ως προς τις προαναφερθείσες κατευθύνσεις, πάντα όμως υπάρχει περιθώριο για την ανάπτυξη νέων, αξιόπιστων μεθόδων ή τη βελτίωση υπαρχόντων τεχνικών μέσω υδρικών προσεγγίσεων. Σκοπός μας είναι, η προτεινόμενη μεθοδολογία που αναπτύχθηκε στο Κεφάλαιο 3 να δοκιμαστεί εκτενέστερα στο μέλλον και να εφαρμοστεί σε προβλήματα που εντοπίζονται στην τομή δύο σημαντικών επιστημονικών περιοχών, της Υπολογιστικής Νοημοσύνης και της Μηχανικής Μάθησης. Για τον σκοπό αυτό μελετήθηκαν τα προαπαιτούμενα που επιζητούν οι αλγόριθμοι που συναντώνται σε αυτούς τομείς και ιδιαίτερη προσοχή δόθηκε στην προεπεξεργασία των δεδομένων [11]. Όσο αξιόπιστες κι αν είναι οι μέθοδοι αυτών των ερευνητικών πεδίων, αν τα δεδομένα τα οποία λαμβάνουν στις εισόδους τους δεν είναι «ποιοτικά», τότε οι μέθοδοι είναι «καταδικασμένες» να αποτύχουν. Συνεπώς, για να προσδοκούμε αξιόπιστη, ακριβή, προσαρμοστική και ανταγωνιστική συμπεριφορά από αυτούς τους αλγορίθμους, σαν «πρώτο» βήμα θα πρέπει να επεξεργαστούμε τα δεδομένα που τους δίνουμε. Για τον λόγο αυτό, στο Κεφάλαιο 4 ο αναγνώστης συναντά μια εκτενή μελέτη που πραγματοποιήσαμε γύρω από κάθε βήμα της προεπεξεργασίας για τους αλγορίθμους που συναντάμε στη Μηχανική Μάθηση, και συγκεκριμένα σε αρκετά προβλήματα που άπτονται της Εξόρυξης Δεδομένων.

Σαν επόμενο βήμα, εντοπίσαμε προβλήματα που χαρακτηρίζουν το σύνολο των μεθόδων Μηχανικής Μάθησης και ένα από τα σημαντικότερα είναι αυτό της αναγνώρισης των ακραίων τιμών [10]. Πολλές φορές τα σύνολα δεδομένων περιέχουν τιμές που δεν θα έπρεπε να συμπεριλαμβάνονται, και χαρακτηρίζονται ως «ανωμαλίες». Βέβαια, η διαχείριση αυτών των τιμών πρέπει να είναι προσεκτική και οι τιμές αυτές δεν θα πρέπει να εξαιρούνται δίχως περαιτέρω μελέτη. Αυτό συμβαίνει διότι, μια ακραία τιμή μπορεί σε κάποιες περιπτώσεις να εμπεριέχει τρομερά χρήσιμη πληροφορία για το μοτίβο το οποίο καλούμαστε να «αποκρυπτογραφίσουμε» και γι' αυτό δεν πρέπει να «πειταχεί» αβλεπτεί (π.χ. μια υψηλή τιμή στις εξετάσεις ενός ασθενούς μπορεί να δίνει πληροφορία για κάποια νόσο). Έτσι, στο Κεφάλαιο 5, ο αναγνώστης συναντά μια νέα υβριδική μέθοδο για τον εντοπισμό των ακραίων τιμών σε ένα σύνολο δεδομένων. Αυτή η μέθοδος συγκρίθηκε σε ένα μεγάλο πλήθος συνόλου δεδομένων με διαφορετικά χαρακτηριστικά, όπως αριθμό παραδειγμάτων, αριθμό χαρακτηριστικών, κ.ά.. Στα

υπολογιστικά πειράματα που διεξήγαμε, συμπεριλήφθησαν πολλές γνωστές μέθοδοι αυτής της κατηγορίας και τα αποτελέσματα που ελήφθησαν δείχνουν ότι η προτεινόμενη μέθοδος είναι ανταγωνιστική, αξιόπιστη και κατά μέσο όρο ξεπερνά σε επιδόσεις τις υπόλοιπες [10].

Τα συμπεράσματα τα οποία αναφέρονται στο Κεφάλαιο 2, μάς ενέπνευσαν ώστε να αναπτύξουμε τη συνεργατική μέθοδο [9] που παρουσιάζεται στο Κεφάλαιο 6. Ένα από τα σημαντικότερα προβλήματα που συναντώνται στον τομέα της Μηχανικής Μάθησης είναι αυτό της ταξινόμησης. Μπορεί τα μοντέλα που έχουν αναπτυχθεί να είναι πολλά και να εφαρμόζονται με επιτυχία σε δύσκολες εφαρμογές, παρόλα αυτά δεν υπάρχει μία μέθοδος που θα ανταποκρίνεται καλά σε όλα τα προβλήματα. Αυτός ήταν ο λόγος που εστίασαμε την προσοχή μας σε ένα συνεργατικό σχήμα ταξινομητών, και όχι μόνο σε έναν καλό ταξινομητή. Με αυτό τον τρόπο το προτεινόμενο σχήμα αξιοποιεί τα πλεονεκτήματα από διαφορετικές τεχνικές και όχι μόνο από μία. Τα υπολογιστικά πειράματα που διεξήγαμε εμπεριέχουν σύνολα δεδομένων μεγάλου εύρους από άποψη παραδειγμάτων, χαρακτηριστικών και τάξεων. Σύμφωνα με τα αποτελέσματα η συνεργατική μέθοδος που προτείναμε [9] ήταν άκρως ανταγωνιστική και κατά μέσο όρο επιτύχανε καλύτερα αποτελέσματα ταξινόμησης σε σχέση με πολύ γνωστές μεθόδους ταξινόμησης.

Στο Κεφάλαιο 7, παρουσιάζονται πρόσφατες εξελικτικές προσεγγίσεις πολλαπλών αντικειμένων για τέσσερις βασικές διεργασίες της Εξόρυξης Δεδομένων και Μηχανικής Μάθησης [8], και συγκεκριμένα, αφορούν: (α) την προεπεξεργασία των δεδομένων, (β) το πρόβλημα της ταξινόμησης, (γ) το πρόβλημα της ομαδοποίησης δεδομένων και (δ) τους κανόνες συσχέτισης. Μέσα από αυτό το κεφάλαιο σκιαγραφούνται βασικές τομές μεταξύ των επιστημονικών πεδίων της Βελτιστοποίησης, της Υπολογιστικής Νοημοσύνης και της Μηχανικής Μάθησης. Το πρόβλημα της πολυ-κριτηριακής Βελτιστοποίησης είναι πολύ σημαντικό και συναντάται σε πολλές και διαφορετικές πτυχές της επιστήμης και της τεχνολογίας. Συνεπώς, έχουμε μια μεγάλη επιλογή εφαρμογών στις οποίες δύναται να εφαρμοστούν οι μέθοδοι που αναπτύχθηκαν στο Κεφάλαιο 3. Ο αναγνώστης, επιπλέον, ενημερώνεται για τη συμβολή των μεθόδων Εξελικτικού Υπολογισμού στα προβλήματα Βελτιστοποίησης.

Μια ερευνητική προσπάθεια δε θα είχε μεγάλη επίδραση, εάν είχε, ως υποθέσουμε, ένα προκαθορισμένο όριο ζωής. Αυτό σημαίνει ότι οι ερευνητικές προσπάθειες που αναπτύσσονται από την ερευνητική κοινότητα, είναι σημαντικό να διακρίνονται από «ενεργά χαρακτηριστικά», με την έννοια που αυτά συναντώνται σε έναν ζωντανό οργανισμό και συνεπώς, η έρευνα να έχει συνέπειες και εξέλιξη. Για τον λόγο αυτό στην επόμενη ενότητα παρουσιάζουμε τις μελλοντικές προεκτάσεις που έχουν τα αποτελέσματα αλλά και οι μέθοδοι που αναπτύσσονται στη συγκεκριμένη διδακτορική διατριβή.

8.2 Μελλοντικές προεκτάσεις

Το πρόβλημα της επιλογής της «καταλληλότερης» μεθόδου που πρέπει να εφαρμοστεί για για την προσέγγιση ενός συγκεκριμένου προβλήματος απασχολεί την επιστημονική κοινότητα εδώ και πολλά χρόνια. Ένα από τα σημαντικότερα θεωρητικά αποτελέσματα της μαθηματικής Βελτιστοποίησης προς την κατεύθυνση αυτή είναι το γνωστό 'No free lunch theorem', το οποίο αποδεικνύει ότι η μέση απόδοση οποιουδήποτε ζεύγους αλγορίθμων Βελτιστοποίησης, σε όλα τα πιθανά προβλήματα Βελτιστοποίησης είναι ίση/ταυτόσημη. Έτσι, ένας αλγόριθμος Βελτιστοποίησης που αποδίδει καλά σε μια κατηγορία προβλημάτων, είναι βέβαιο ότι θα αποδώσει χειρότερα σε ένα άλλο σύνολο προβλημάτων από αυτά που απομένουν. Όμως, σε αντίθεση με το 'No free lunch theorem', πρόσφατα έχει αναπτυχθεί μια θεωρία για την ύπαρξη 'Free lunches' (όπως χαρτογραφείται στο Κεφάλαιο 2) για κατηγορίες εξελικτικών αλγορίθμων (αλγόριθμοι συν-εξέλιξης), όπως και Τεχνητών Νευρωνικών Δικτύων. Έτσι, μπορεί να υπάρξει ένα βέλτιστο δίκτυο το οποίο προσεγγίζει συγκεκριμένες οικογένειες προβλημάτων με συγκεκριμένη αρχιτεκτονική. Έτσι, οι προτεινόμενες μέθοδοι του άρθρου [12] δύναται να εφαρμοστούν μελλοντικά για την εκπαίδευση τεχνητών νευρωνικών δικτύων για

την αντιμετώπιση πολύπλοκων προβλημάτων του πραγματικού κόσμου. Συμπερασματικά, περαιτέρω μελέτη δύναται να επιτρέψει τον προσδιορισμό περιοχών μάθησης τεχνητών νευρωνικών δικτύων για την ανάλυση της πολυπλοκότητας της βέλτιστης τοπολογίας (αρχιτεκτονικής).

Χρησιμοποιώντας τη μεθοδολογία που παρουσιάζουμε στο Κεφάλαιο 3, μπορούμε να δημιουργήσουμε πληθώρα από προσεγγιστικές διαδικασίες των τροχιών ανίχνευσης με βάση τις μεθόδους αριθμητικής επίλυσης συνήθων διαφορικών εξισώσεων των Adams-Bashforth και Adams-Moulton, οι οποίες μπορούν να εφαρμοστούν για την αποδοτικότερη και αποτελεσματικότερη αντιμετώπιση του προβλήματος της εκπαίδευσης τεχνητών νευρωνικών δικτύων. Επιπλέον, μπορούν να συνεισφέρουν στον καθορισμό νέων τελεστών διαφορο-εξελικτικών αλγόριθμων ολικής βελτιστοποίησης για την αντιμετώπιση του προβλήματος της εκπαίδευσης. Επιπλέον, μπορούν να καθοριστούν κλάσεις μεθόδων και μέτρα αποδοτικότητας αυτών, για την αυτόματη επιλογή της καταλληλότερης μεθόδου εκπαίδευσης τεχνητών νευρωνικών δικτύων με σκοπό την αποτελεσματικότερη αντιμετώπιση συγκεκριμένου προβλήματος ή συγκεκριμένης οικογένειας προβλημάτων. Η επιλογή της μεθόδου θα ήταν πολύ χρήσιμο αν πραγματοποιείται αυτόματα κατά τη διάρκεια της εκπαίδευσης λαμβάνοντας υπόψη την αποδοτικότητα όλων των χρησιμοποιούμενων μεθόδων στο συγκεκριμένο πρόβλημα.

Ένα δύσκολο, απαιτητικό αλλά τρομερά οφέλιμο πρόβλημα θέτει τις βάσεις του στο Κεφάλαιο 4, ώστε μελλοντικά να προταθεί μια ιεραρχία βημάτων, αναλόγως με το υπό εξέταση πρόβλημα και το δοθέν σύνολο δεδομένων. Επιπλέον, μέσω του Κεφαλαίου 5, χαρτογραφείται η δυναμική της προτεινόμενης μεθόδου για μελλοντική εφαρμογή και περαιτέρω βελτίωση, μέσα από την ενσωμάτωση μη-επιτηρούμενης επιλογής χαρακτηριστικών.

Το σύνολο των κανόνων που χρησιμοποιούνται για τη δημιουργία ενός κατάλληλου συνόλου ταξινομητών με την υψηλότερη δυνατή ακρίβεια, είναι ένα ανοιχτό ζήτημα για τις συνεργατικές μεθόδους. Έτσι, οι περαιτέρω συγκρίσεις της μεθόδου μας με γνωστές παραλλαγές του αλγόριθμου Stacking και η εφαρμογή σε προβλήματα παλινδρόμησης θα μπορούσε να είναι ένα ενδιαφέρον έργο για μελλοντική έρευνα. Ακόμα, η δυσκολότερη, ίσως, και πιο ενδιαφέρουσα εργασία, που παραμένει ένα ανοιχτό πρόβλημα, είναι η βέλτιστη επιλογή ταξινομητών για τη συμμετοχή στο συνεργατικό σύνολο ταξινομητών. Πρόκειται για ένα ζήτημα που μπορεί να αντιμετωπιστεί ως συνδυαστικό πρόβλημα βελτιστοποίησης και άρα, οι μέθοδοι που αναλύονται στο Κεφάλαιο 3 μπορούν να βρουν ένα επιπλέον πεδίο εφαρμογής.

Βιβλιογραφία

- [1] Giovanni Acampora, Francisco Herrera, Genoveffa Tortora, and Autilia Vitiello, “A multi-objective evolutionary approach to training set selection for support vector machine”, *Knowledge-Based Systems*, vol. 147, 94–108, 2018.
- [2] Stavros P Adam, Stamatios Aggelos N Alexandropoulos, Panos M Pardalos, and Michael N Vrahatis, “No free lunch theorem: A review”, In *Approximation and Optimization*, Springer, 57–82, 2019.
- [3] Charu C Aggarwal, “An introduction to outlier analysis”, In *Outlier Analysis*, Springer, 1–40, 2013.
- [4] Faez Ahmed, Kalyanmoy Deb, and Abhilash Jindal, “Multi-objective optimization and decision making approaches to cricket team selection”, *Applied Soft Computing*, vol. 13, No. 1, 402–414, 2013.
- [5] Mohammad Majidal Rifaie and John Mark Bishop, “Swarmic paintings and colour attention”, In the proceedings of the International Conference on Evolutionary and Biologically Inspired Music and Art, Springer, 97–108, 2013.
- [6] Wissam A Albukhanajer, Yaochu Jin, and Johann A Briffa, “Classifier ensembles for image identification using multi-objective pareto features”, *Neurocomputing*, vol. 238, 316–327, 2017.
- [7] Stamatios Aggelos N Alexandropoulos, Christos K Aridas, Sotiris B Kotsiantis, and Michael N Vrahatis, “A deep dense neural network for bankruptcy prediction”, In the proceedings of the International Conference on Engineering Applications of Neural Networks, Springer, 435–444, 2019.
- [8] Stamatios Aggelos N Alexandropoulos, Christos K Aridas, Sotiris B Kotsiantis, and Michael N Vrahatis, “Multi-objective evolutionary optimization algorithms for machine learning: A recent survey”, In *Approximation and Optimization*, Springer, 35–55, 2019.
- [9] Stamatios Aggelos N Alexandropoulos, Christos K Aridas, Sotiris B Kotsiantis, and Michael N Vrahatis, “Stacking strong ensembles of classifiers”, In the proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, 545–556, 2019.
- [10] Stamatios Aggelos N Alexandropoulos, Sotiris B Kotsiantis, Violetta E Piperigou, and Michael N Vrahatis, “A new ensemble method for outlier identification”, In the proceedings of the IEEE Tenth International Conference on Cloud Computing, Data Science Engineering (Confluence 2020), January 29–31, 2020, Noida, Uttar Pradesh, India, IEEE, 786–791, 2020.
- [11] Stamatios Aggelos N Alexandropoulos, Sotiris B Kotsiantis, and Michael N Vrahatis, “Data preprocessing in predictive data mining”, *The Knowledge Engineering Review*, vol. 34, 2019.
- [12] Stamatios Aggelos N Alexandropoulos, Panos M Pardalos, and Michael N Vrahatis, “Dynamic search trajectory methods for global optimization”, *Annals of Mathematics and Artificial Intelligence*, 1–35, 2019.

- [13] Abhay Kumar Alok, Sriparna Saha, and Asif Ekbal, "A new semi-supervised clustering technique using multi-objective optimization", *Applied Intelligence*, vol. 43, No. 3, 633–661, 2015.
- [14] Shun ichi Amari, Noboru Murata, K R Muller, Michael Finke, and Howard Hua Yang, "Asymptotic statistical theory of overtraining and cross-validation", *IEEE Transactions on Neural Networks*, vol. 8, No. 5, 985–996, 1997.
- [15] Fabrizio Angiulli and Fabio Fassetti, "Exploiting domain knowledge to detect outliers", *Data Mining and Knowledge Discovery*, vol. 28, No. 2, 519–568, 2014.
- [16] Fabrizio Angiulli and Clara Pizzuti, "Fast outlier detection in high dimensional spaces", In the proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 15–27, 2002.
- [17] Fabrizio Angiulli and Clara Pizzuti, "Outlier mining in large high-dimensional data sets", *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, No. 2, 203–215, 2005.
- [18] Christos K Aridas, Stamatios Aggelos N Alexandropoulos, Sotiris B Kotsiantis, and Michael N Vrahatis, "Random resampling in the one-versus-all strategy for handling multi-class problems", In the proceedings of the International Conference on Engineering Applications of Neural Networks, Springer, 111–121, 2017.
- [19] Christos K. Aridas, Sotiris B. Kotsiantis, and Michael N. Vrahatis, *Combining Prototype Selection with Local Boosting*, Springer International Publishing, Cham, 2016, 94–105.
- [20] Christos K. Aridas, Sotiris B. Kotsiantis, and Michael N. Vrahatis, "Hybrid local boosting utilizing unlabeled data in classification tasks", *Evolving Systems*, 2017.
- [21] Larry Armijo, "Minimization of functions having lipschitz continuous first partial derivatives", *Pacific Journal of mathematics*, vol. 16, No. 1, 1–3, 1966.
- [22] Álgvar Arnaiz-González, José Francisco Díez-Pastor, Juan J Rodríguez, and César García-Osorio, "Instance selection of linear complexity for big data", *Knowledge-Based Systems*, vol. 107, 83–95, 2016.
- [23] M Gethsiyal Augasta and T Kathirvalavakumar, "A new discretization algorithm based on range coefficient of dispersion and skewness for neural networks classifier", *Applied Soft Computing*, vol. 12, No. 2, 619–625, 2012.
- [24] Anne Auger, Marc Schoenauer, and Olivier Teytaud, "Local and global order 3/2 convergence of a surrogate evolutionary algorithm", In the proceedings of the Proceedings of the 7th annual conference on Genetic and evolutionary computation, 857–864, 2005.
- [25] Anne Auger and Olivier Teytaud, "Continuous lunches are free plus the design of optimal optimization algorithms", *Algorithmica*, vol. 57, No. 1, 121–146, 2010.
- [26] Ilhan Aydin, Mehmet Karakose, and Erhan Akin, "A multi-objective artificial immune algorithm for parameter optimization in support vector machine", *Applied soft computing*, vol. 11, No. 1, 120–129, 2011.
- [27] Thomas Back, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*, Oxford university press, 1996.
- [28] Carlo Barbieri, Simona Cocco, and Rémi Monasson, "On the trajectories and performance of infotaxis, an information-based greedy search algorithm", *EPL (Europhysics Letters)*, vol. 94, No. 2, 20005, 2011.

- [29] Rodrigo C Barros, André CPLFde Carvalho, and Alex A Freitas, “Decision-tree induction”, In Automatic Design of Decision-Tree Induction Algorithms, Springer, 7–45, 2015.
- [30] Marcio P Basgalupp, Andre CPLF De Carvalho, Rodrigo C Barros, Duncan D Ruiz, and Alex A Freitas, “Lexicographic multi-objective evolutionary induction of decision trees”, International Journal of Bio-Inspired Computation, vol. 1, No. 1-2, 105–117, 2009.
- [31] Gustavo EAPA Batista, Ronaldo C Prati, and Maria Carolina Monard, “A study of the behavior of several methods for balancing machine learning training data”, ACM Sigkdd Explorations Newsletter, vol. 6, No. 1, 20–29, 2004.
- [32] Roberto Battiti, “First-and second-order methods for learning: between steepest descent and newton’s method”, Neural computation, vol. 4, No. 2, 141–166, 1992.
- [33] Vahid Beiranvand, Mohamad Mobasher-Kashani, and Azuraliza Abu Bakar, “Multi-objective pso algorithm for mining numerical association rules without a priori discretization”, Expert systems with applications, vol. 41, No. 9, 4259–4273, 2014.
- [34] R Bellman, “Dynamic programming: Princeton univ. press”, N[U+30FB]J, vol. 95, 1957.
- [35] James C Bezdek, “What is computational intelligence?” Technical report, USDOE Pittsburgh Energy Technology Center, PA (United States); Oregon State . . . , 1994.
- [36] James C Bezdek and Ludmila I Kuncheva, “Nearest prototype classifier designs: An experimental study”, International Journal of Intelligent Systems, vol. 16, No. 12, 1445–1473, 2001.
- [37] Amit Bhaya and Eugenius Kaszkurewicz, “Steepest descent with momentum for quadratic functions is a version of the conjugate gradient method”, Neural Networks, vol. 17, No. 1, 65–71, 2004.
- [38] Amit Bhaya and Eugenius Kaszkurewicz, Control perspectives on numerical algorithms and matrix problems, ολ. 10, SIAM, 2006.
- [39] Amit Bhaya, Fernando Pazos, and Eugenius Kaszkurewicz, “The controlled conjugate gradient type trajectory-following neural net for minimization of nonconvex functions.” In the proceedings of the The 2010 International Joint Conference on Neural Networks (CNN), IEEE, 1–8, 2010.
- [40] Urvesh Bhowan, Mark Johnston, Mengjie Zhang, and Xin Yao, “Evolving diverse ensembles using genetic programming for classification with unbalanced data”, IEEE Transactions on Evolutionary Computation, vol. 17, No. 3, 368–386, 2012.
- [41] Urvesh Bhowan, Mark Johnston, Mengjie Zhang, and Xin Yao, “Reusing genetic programming for ensemble selection in classification of unbalanced data”, IEEE Transactions on Evolutionary Computation, vol. 18, No. 6, 893–908, 2013.
- [42] Paul T Boggs, “An algorithm, based on singular perturbation theory, for ill-conditioned minimization problems”, SIAM Journal on Numerical Analysis, vol. 14, No. 5, 830–843, 1977.
- [43] Marc Boulle, “KHIOPS: A statistical discretization method of continuous attributes”, Machine Learning, vol. 55, No. 1, 53–69, 2004.
- [44] Franklin H Branin, “Widely convergent method for finding multiple solutions of simultaneous nonlinear equations”, IBM Journal of Research and Development, vol. 16, No. 5, 504–522, 1972.
- [45] Leo Breiman, “Bagging predictors”, Machine learning, vol. 24, No. 2, 123–140, 1996.

- [46] Leo Breiman, "Random forests", *Machine learning*, vol. 45, No. 1, 5-32, 2001.
- [47] Markus M Breunig, Hans Peter Kriegel, Raymond T Ng, and Jörg Sander, "Lof: identifying density-based local outliers", In the proceedings of the Proceedings of the 2000 ACM SIGMOD international conference on Management of data, 93-104, 2000.
- [48] Carla E. Brodley and Mark A. Friedl, "Identifying mislabeled training data", *Journal of Artificial Intelligence Research*, vol. 11, 131-167, 1999.
- [49] JC Butcher, "Differential and difference equations", *Numerical Methods for Ordinary Differential Equations: Early Days the Birth of Numerical Analysis*, 2nd ed., Hoboken, New Jersey, USA: John Wiley & Sons, 1-49, 2008.
- [50] John Charles Butcher, *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*, Wiley-Interscience, 1987.
- [51] Guido Buzzi-Ferraris and Flavio Manenti, "Outlier detection in large data sets", *Computers & Chemical Engineering*, vol. 35, No. 2, 388-390, 2011.
- [52] Yoel Caises, Antonio González, Enrique Leyva, and Raúl Pérez, "Combining instance selection methods based on data characterization: An approach to increase their effectiveness", *Information Sciences*, vol. 181, No. 20, 4780-4798, 2011.
- [53] Alberto Cano, Dat T Nguyen, Sebastián Ventura, and Krzysztof J Cios, "ur-CAIM: improved CAIM discretization for unbalanced and balanced data", *Soft Computing*, vol. 20, No. 1, 173-188, 2016.
- [54] Alberto Cano, Sebastián Ventura, and Krzysztof J Cios, "Multi-objective genetic programming for feature extraction and data visualization", *Soft Computing*, vol. 21, No. 8, 2069-2089, 2017.
- [55] José Ramón Cano, Salvador García, and Francisco Herrera, "Subgroup discover in large size data sets preprocessed using stratified instance selection for increasing the presence of minority classes", *Pattern Recognition Letters*, vol. 29, No. 16, 2156-2164, 2008.
- [56] Jose Ramon Cano, Francisco Herrera, and Manuel Lozano, "Strategies for scaling up evolutionary instance reduction algorithms for data mining", In *Evolutionary Computation in Data Mining*, Springer, 21-39, 2005.
- [57] Rich Caruana and Virginia R de Sa, "Benefitting from the variables that variable selection discards", *Journal of Machine Learning Research*, vol. 3, No. Mar, 1245-1264, 2003.
- [58] Zehra Cataltepe, Yaser S Abu-Mostafa, and Malik Magdon-Ismael, "No free lunch for early stopping", *Neural computation*, vol. 11, No. 4, 995-1009, 1999.
- [59] Varun Chandola, Arindam Banerjee, and Vipin Kumar, "Anomaly detection: A survey", *ACM computing surveys (CSUR)*, vol. 41, No. 3, 1-58, 2009.
- [60] Arjun Chandra and Xin Yao, "Ensemble learning using multi-objective evolutionary algorithms", *Journal of Mathematical Modelling and Algorithms*, vol. 5, No. 4, 417-445, 2006.
- [61] Nitesh V Chawla, "Data mining for imbalanced datasets: An overview", In *Data mining and knowledge discovery handbook*, Springer, 875-886, 2009.
- [62] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer, "SMOTE: synthetic minority over-sampling technique", *Journal of Artificial Intelligence Research*, vol. 16, 321-357, 2002.

- [63] Kumar Chellapilla and David B Fogel, "Evolution, neural networks, games, and intelligence", *Proceedings of the IEEE*, vol. 87, No. 9, 1471–1496, 1999.
- [64] Shuyan Chen, Wei Wang, and Henkvan Zuylen, "A comparison of outlier detection algorithms for ITS data", *Expert Systems with Applications*, vol. 37, No. 2, 1169–1178, 2010.
- [65] Yumin Chen, Duoqian Miao, and Hongyun Zhang, "Neighborhood outlier detection", *Expert Systems with Applications*, vol. 37, No. 12, 8745–8749, 2010.
- [66] Yijun Chen and Man Leung Wong, "An ant colony optimization approach for stacking ensemble", In the proceedings of the 2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC), IEEE, 146–151, 2010.
- [67] Zhengxin Chen, *Computational intelligence for decision support*, CRC Press, 1999.
- [68] Igor Chikalov, Shahid Hussain, and Mikhail Moshkov, "Bi-criteria optimization of decision trees with applications to data analysis", *European Journal of Operational Research*, vol. 266, No. 2, 689–701, 2018.
- [69] Tommy WS Chow and D Huang, "Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information", *IEEE Transactions on Neural Networks*, vol. 16, No. 1, 213–224, 2005.
- [70] Federico Cismondi, André S Fialho, Susana M Vieira, Shane R Reti, João MC Sousa, and Stan N Finkelstein, "Missing data in medical databases: Impute, delete or classify?" *Artificial Intelligence in Medicine*, vol. 58, No. 1, 63–72, 2013.
- [71] Florin Ciucu and Jens Schmitt, "Perspectives on network calculus: no free lunch, but still good value", In the proceedings of the Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication, 311–322, 2012.
- [72] Carlos A Coello Coello, "A comprehensive survey of evolutionary-based multiobjective optimization techniques", *Knowledge and Information systems*, vol. 1, No. 3, 269–308, 1999.
- [73] Carlos A Coello Coello, "Evolutionary multi-objective optimization and its use in finance", *Handbook of Research on Nature Inspired Computing for Economy and Management*. Idea Group Publishing, 2006.
- [74] David Corne and Joshua Knowles, "Some multiobjective optimizers are better than others", In the proceedings of the The 2003 Congress on Evolutionary Computation, 2003. CEC'03., vol. 4, IEEE, 2506–2512, 2003.
- [75] Lino Costa and Pedro Oliveira, "Dimension reduction in multiobjective optimization", In the proceedings of the PAMM: Proceedings in Applied Mathematics and Mechanics, vol. 7, Wiley Online Library, 2060047–2060048, 2007.
- [76] Sven F Crone, Stefan Lessmann, and Robert Stahlbock, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing", *European Journal of Operational Research*, vol. 173, No. 3, 781–800, 2006.
- [77] Pdraig Cunningham and Delany SJ, "k-nearest neighbour classifiers. mult classif syst", 2007.
- [78] Ireneusz Czarnowski, "Prototype selection algorithms for distributed learning", *Pattern Recognition*, vol. 43, No. 6, 2292–2300, 2010.
- [79] Ireneusz Czarnowski, "Cluster-based instance selection for machine classification", *Knowledge and Information Systems*, vol. 30, No. 1, 113–133, 2012.

- [80] Ayan Das and Swagatam Das, "Feature weighting and selection with a pareto-optimal trade-off between relevancy and redundancy", *Pattern Recognition Letters*, vol. 88, 12-19, 2017.
- [81] Swagatam Das, Ajith Abraham, and Amit Konar, "Automatic clustering using an improved differential evolution algorithm", *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, vol. 38, No. 1, 218-237, 2007.
- [82] Leandro Nunes De Castro, *Fundamentals of natural computing: basic concepts, algorithms, and applications*, CRC Press, 2006.
- [83] Aidade Haro-García and Nicolás García-Pedrajas, "A divide-and-conquer recursive approach for scaling up instance selection algorithms", *Data Mining and Knowledge Discovery*, vol. 18, No. 3, 392-418, 2009.
- [84] Cláudio Rebelode Sá, Carlos Soares, and Arno Knobbe, "Entropy-based discretization methods for ranking data", *Information Sciences*, vol. 329, 921-936, 2016.
- [85] Kalyanmoy Deb, *Multi-objective optimization using evolutionary algorithms*, ολ. 16, John Wiley & Sons, 2001.
- [86] Kalyanmoy Deb, "Multi-objective optimisation using evolutionary algorithms: an introduction", In *Multi-objective evolutionary optimisation for product design and manufacturing*, Springer, 3-34, 2011.
- [87] Kalyanmoy Deb, "Multi-objective optimization", In *Search methodologies*, Springer, 403-449, 2014.
- [88] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii", *IEEE transactions on evolutionary computation*, vol. 6, No. 2, 182-197, 2002.
- [89] Satchidananda Dehuri, Srikanta Patnaik, Ashish Ghosh, and Rajib Mall, "Application of elitist multi-objective genetic algorithm for classification rule generation", *Applied Soft Computing*, vol. 8, No. 1, 477-487, 2008.
- [90] Sarah Jane Delany, Nicola Segata, and Brian Mac Namee, "Profiling instances in noise reduction", *Knowledge-Based Systems*, vol. 31, 28-40, 2012.
- [91] William A Dembski, *No free lunch: Why specified complexity cannot be purchased without intelligence*, Rowman & Littlefield, 2006.
- [92] John E Dennis Jr and Robert B Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, ολ. 16, Siam, 1996.
- [93] Joaquín Derrac, Salvador García, and Francisco Herrera, "IFS-CoCo: Instance and feature selection based on cooperative coevolution with nearest neighbor rule", *Pattern Recognition*, vol. 43, No. 6, 2082-2105, 2010.
- [94] Joaquín Derrac, Salvador García, and Francisco Herrera, "A survey on evolutionary instance selection and generation", *International Journal of Applied Metaheuristic Computing*, vol. 1, 60-92, 2010.
- [95] Thomas G Dietterich, "Ensemble methods in machine learning", In *the proceedings of the International workshop on multiple classifier systems*, Springer, 1-15, 2000.
- [96] Laurence CW Dixon, "Neural networks and unconstrained optimization", In *Algorithms for Continuous Optimization*, Springer, 513-530, 1994.
- [97] Laurence Charles Ward Dixon, "The global optimization problem. an introduction", *Toward global optimization*, vol. 2, 1-15, 1978.
- [98] Francisco Domingo-Perez, Jose Luis Lazaro-Galilea, Andreas Wieser, Ernesto Martin-Gorostiza, David Salido-Monzu, and Alvarode la Llana, "Sensor placement

- determination for range-difference positioning using evolutionary multi-objective optimization”, *Expert Systems with Applications*, vol. 47, 95–105, 2016.
- [99] Rémi Domingues, Maurizio Filippone, Pietro Michiardi, and Jihane Zouaoui, “A comparative evaluation of outlier detection algorithms: Experiments and analyses”, *Pattern Recognition*, vol. 74, 406–421, 2018.
- [100] Marco Dorigo and Mauro Birattari, “Ant colony optimization. encyclopedia of machine learning”, 2010.
- [101] Marco Dorigo and Gianni Di Caro, “Ant colony optimization: a new meta-heuristic”, In the proceedings of the Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406), vol. 2, IEEE, 1470–1477, 1999.
- [102] James Dougherty, Ron Kohavi, and Mehran Sahami, “Supervised and unsupervised discretization of continuous features”, In *Machine Learning Proceedings 1995*, Elsevier, 194–202, 1995.
- [103] Stefan Droste, Thomas Jansen, and Ingo Wegener, “Optimization with randomized search heuristics—the (a) nfl theorem, realistic scenarios, and difficult functions”, *Theoretical Computer Science*, vol. 287, No. 1, 131–144, 2002.
- [104] Dheeru Dua and Casey Graff, “Uci machine learning repository”, 2017.
- [105] Sumeet Dua and Xian Du, *Data mining and machine learning in cybersecurity*, CRC press, 2016.
- [106] Olive Jean Dunn, “Multiple comparisons among means”, *Journal of the American statistical association*, vol. 56, No. 293, 52–64, 1961.
- [107] Dipankar Dutta, Paramartha Dutta, and Jaya Sil, “Simultaneous feature selection and clustering with mixed features by multi objective genetic algorithm”, *International Journal of Hybrid Intelligent Systems*, vol. 11, No. 1, 41–54, 2014.
- [108] Eugene G D’yakonov, *Optimization in solving elliptic problems*, CRC Press, 2018.
- [109] Saso Džeroski and Bernard Ženko, “Is combining classifiers with stacking better than selecting the best one?” *Machine learning*, vol. 54, No. 3, 255–273, 2004.
- [110] Russell Eberhart and James Kennedy, “A new optimizer using particle swarm theory”, In the proceedings of the MHS’95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Ieee, 39–43, 1995.
- [111] Russ Eberhart, Pat Simpson, and Roy Dobbins, *Computational intelligence PC tools*, Academic Press Professional, Inc., 1996.
- [112] Rajmadhan Ekambaram, Sergiy Fefilatyev, Matthew Shreve, Kurt Kramer, Lawrence O Hall, Dmitry B Goldgof, and Rangachar Kasturi, “Active cleaning of label noise”, *Pattern Recognition*, vol. 51, 463–480, 2016.
- [113] Tapio Elomaa and Juho Rousu, “Efficient multisplitting revisited: Optima-preserving elimination of partition candidates”, *Data Mining and Knowledge Discovery*, vol. 8, No. 2, 97–126, 2004.
- [114] Michael G Epitropakis, Vassilis P Plagianakos, and Michael N Vrahatis, “Evolving cognitive and social experience in particle swarm optimization through differential evolution: a hybrid approach”, *Information Sciences*, vol. 216, 50–92, 2012.
- [115] Michael G Epitropakis, Dimitris K Tasoulis, Nicos G Pavlidis, Vassilis P Plagianakos, and Michael N Vrahatis, “Enhancing differential evolution utilizing proximity-based mutation operators”, *IEEE Transactions on Evolutionary Computation*, vol. 15, No. 1, 99–119, 2011.

- [116] Mark Erickson, Alex Mayer, and Jeffrey Horn, "The niched pareto genetic algorithm 2 applied to the design of groundwater remediation systems", In the proceedings of the International Conference on Evolutionary Multi-Criterion Optimization, Springer, 681-695, 2001.
- [117] H Jair Escalante, "A comparison of outlier detection algorithms for machine learning", In the proceedings of the Proceedings of the International Conference on Communications in Computing, 228-237, 2005.
- [118] Hugo Jair Escalante, Maribel Marin-Castro, Alicia Morales-Reyes, Mario Graff, Alejandro Rosales-Pérez, Manuel Montes y Gómez, Carlos A Reyes, and Jesus A Gonzalez, "Mopg: a multi-objective evolutionary algorithm for prototype generation", Pattern Analysis and Applications, vol. 20, No. 1, 33-47, 2017.
- [119] Andrew Estabrooks, Taeho Jo, and Nathalie Japkowicz, "A multiple resampling method for learning from imbalanced data sets", Computational intelligence, vol. 20, No. 1, 18-36, 2004.
- [120] Ya Ju Fan, Cem Iyigun, and W Art Chaovalitwongse, "Recent advances in mathematical programming for classification and cluster analysis", In the proceedings of the CRM Proceedings and Lecture Notes, vol. 45, Citeseer, 67-93, 2008.
- [121] Alireza Farhangfar, Lukasz Kurgan, and Jennifer Dy, "Impact of imputation of missing values on classification error for discrete data", Pattern Recognition, vol. 41, No. 12, 3692-3705, 2008.
- [122] Alireza Farhangfar, Lukasz A Kurgan, and Witold Pedrycz, "A novel framework for imputation of missing values in databases", IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 37, No. 5, 692-709, 2007.
- [123] József Farkas and Károly Jármay, Design and optimization of metal structures, Elsevier, 2008.
- [124] József Farkas, Károly Jármay, and Jan A Snyman, "Global minimum cost design of a welded square stiffened plate supported at four corners", Structural and Multidisciplinary Optimization, vol. 40, No. 1-6, 477, 2010.
- [125] MAH Farquad and Indranil Bose, "Preprocessing unbalanced data using support vector machine", Decision Support Systems, vol. 53, No. 1, 226-233, 2012.
- [126] A. Fernández, C. J. Carmona, M.J.del Jesus, and F. Herrera, "A pareto based ensemble with feature and instance selection for learning from multi-class imbalanced datasets", International Journal of Neural Systems, vol. 27, 1-17, 2017, TIN2014-57251-P, TIN2015-68454-R, P11-TIC-7765, UJA2014/06/15.
- [127] Anthony V Fiacco and Garth P McCormick, Nonlinear programming: sequential unconstrained minimization techniques, ολ. 4, Siam, 1990.
- [128] Sevan Gregory Ficici and Jordan B Pollack, Solution concepts in coevolutionary algorithms, Ph.D. thesis, Brandeis University Waltham, 2004.
- [129] Jonathan E Fieldsend, "Optimizing decision trees using multi-objective particle swarm optimization", In Swarm intelligence for multi-objective problems in data mining, Springer, 93-114, 2009.
- [130] Peter Filzmoser, Ricardo Maronna, and Mark Werner, "Outlier identification in high dimensions", Computational Statistics & Data Analysis, vol. 52, No. 3, 1694-1711, 2008.
- [131] R Fletcher, "Fortran subroutines for minimization by quasi-newton methods", Technical report, Atomic Energy Research Establishment Harwell, England, 1972.

- [132] Roger Fletcher, *Practical methods of optimization*, John Wiley & Sons, 2013.
- [133] M Julia Flores, José A Gámez, Ana M Martínez, and José M Puerta, “Handling numeric attributes when comparing bayesian network classifiers: does the discretization method matter?” *Applied Intelligence*, vol. 34, No. 3, 372–385, 2011.
- [134] Christodoulos A Floudas and Panos M Pardalos, *A collection of test problems for constrained global optimization algorithms*, ολ. 455, Springer Science & Business Media, 1990.
- [135] Christodoulos A Floudas and Panos M Pardalos, *Encyclopedia of optimization*, ολ. 1, Springer Science & Business Media, 2001.
- [136] David B Fogel, “Review of computational intelligence: imitating life [book reviews]”, *Proceedings of the IEEE*, vol. 83, No. 11, 1588, 1995.
- [137] Alex A Freitas, “A critical review of multi-objective optimization in data mining: a position paper”, *ACM SIGKDD Explorations Newsletter*, vol. 6, No. 2, 77–86, 2004.
- [138] Yoav Freund, Robert E Schapire, and others, “Experiments with a new boosting algorithm”, In the proceedings of the icml, vol. 96, Citeseer, 148–156, 1996.
- [139] Jerome H Friedman, “Greedy function approximation: a gradient boosting machine”, *Annals of statistics*, 1189–1232, 2001.
- [140] John Fulcher and Lakhmi C Jain, *Computational intelligence: a compendium*, ολ. 21, Springer, 2008.
- [141] Johannes Fürnkranz, “Separate-and-conquer rule learning”, *Artificial Intelligence Review*, vol. 13, No. 1, 3–54, 1999.
- [142] Mikel Galar, Alberto Fernández, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera, “An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes”, *Pattern Recognition*, vol. 44, No. 8, 1761–1776, 2011.
- [143] Timothy Ganesan, I Elamvazuthi, and Pandian Vasant, “Multiobjective design optimization of a nano-cmos voltage-controlled oscillator using game theoretic-differential evolution”, *Applied Soft Computing*, vol. 32, 293–299, 2015.
- [144] Luís PF Garcia, André CPLFde Carvalho, and Ana C Lorena, “Noise detection in the meta-learning level”, *Neurocomputing*, vol. 176, 14–25, 2016.
- [145] Luís PF Garcia, Ana C Lorena, Stan Matwin, and André CPLFde Carvalho, “Ensembles of label noise filters: a ranking approach”, *Data Mining and Knowledge Discovery*, vol. 30, No. 5, 1192–1216, 2016.
- [146] Salvador García, José Ramón Cano, and Francisco Herrera, “A memetic algorithm for evolutionary prototype selection: A scaling up approach”, *Pattern Recognition*, vol. 41, No. 8, 2693–2709, 2008.
- [147] Salvador Garcia, Joaquin Derrac, Jose Cano, and Francisco Herrera, “Prototype selection for nearest neighbor classification: Taxonomy and empirical study”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, No. 3, 417–435, 2012.
- [148] Salvador García, Julián Luengo, and Francisco Herrera, *Data preprocessing in data mining*, Springer, 2015.
- [149] Salvador Garcia, Julian Luengo, José Antonio Sáez, Victoria Lopez, and Francisco Herrera, “A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, No. 4, 734–750, 2013.

- [150] Vicente García, Javier Salvador Sánchez, and Ramón Alberto Mollineda, “On the effectiveness of preprocessing methods when dealing with different levels of class imbalance”, *Knowledge-Based Systems*, vol. 25, No. 1, 13–21, 2012.
- [151] Nicolás García-Pedrajas, Juan Antonio Romero Del Castillo, and Domingo Ortiz-Boyer, “A cooperative coevolutionary algorithm for instance selection for instance-based learning”, *Machine Learning*, vol. 78, No. 3, 381–420, 2010.
- [152] Nicolás García-Pedrajas and Javier Pérez-Rodríguez, “Multi-selection of instances: A straightforward way to improve evolutionary instance selection”, *Applied Soft Computing*, vol. 12, No. 11, 3590–3602, 2012.
- [153] Pierre Geurts, Damien Ernst, and Louis Wehenkel, “Extremely randomized trees”, *Machine learning*, vol. 63, No. 1, 3–42, 2006.
- [154] Amol Ghoting, Srinivasan Parthasarathy, and Matthew Eric Otey, “Fast mining of distance-based outliers in high-dimensional datasets”, *Data Mining and Knowledge Discovery*, vol. 16, No. 3, 349–364, 2008.
- [155] Nikos Giannopoulos and Andreas C Nearchou, “Bi-criteria scheduling against restrictive common due dates using a multi-objective differential evolution algorithm”, *IMA Journal of Management Mathematics*, vol. 29, No. 1, 119–136, 2018.
- [156] Philip E Gill and Walter Murray, “Quasi-newton methods for unconstrained optimization”, *IMA Journal of Applied Mathematics*, vol. 9, No. 1, 91–108, 1972.
- [157] Markus Goldstein and Andreas Dengel, “Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm”, *KI-2012: Poster and Demo Track*, 59–63, 2012.
- [158] David Gómez and Alfonso Rojas, “An empirical overview of the no free lunch theorem and its effect on real-world machine learning classification”, *Neural computation*, vol. 28, No. 1, 216–228, 2016.
- [159] Wenyin Gong, Zhihua Cai, and Li Zhu, “An efficient multiobjective differential evolution algorithm for engineering design”, *Structural and Multidisciplinary Optimization*, vol. 38, No. 2, 137–157, 2009.
- [160] L Gonzalez-Abril, Francisco Javier Cuberos, Francisco Velasco, and Juan Antonio Ortega, “AMEVA: An autonomous discretization algorithm”, *Expert Systems with Applications*, vol. 36, No. 3, 5327–5332, 2009.
- [161] Cyril Goutte, “Note on free lunches and cross-validation”, *Neural Computation*, vol. 9, No. 6, 1245–1249, 1997.
- [162] Alex Graves, Abdel rahman Mohamed, and Geoffrey Hinton, “Speech recognition with deep recurrent neural networks”, In the proceedings of the 2013 IEEE international conference on acoustics, speech and signal processing, *IEEE*, 6645–6649, 2013.
- [163] Andreas O Griewank, “Generalized descent for global optimization”, *Journal of optimization theory and applications*, vol. 34, No. 1, 11–39, 1981.
- [164] Andreas Otto Griewank and others, “A generalized descent method for global optimization”, , 1977.
- [165] Evan J Griffiths and Pekka Orponen, “Optimization, block designs and no free lunch theorems”, *Information Processing Letters*, vol. 94, No. 2, 55–61, 2005.
- [166] Luigi Grippo, Francesco Lampariello, and Stephano Lucidi, “A nonmonotone line search technique for newton’s method”, *SIAM Journal on Numerical Analysis*, vol. 23, No. 4, 707–716, 1986.

- [167] AA Groenwold, JA Snyman, and N Stander, “Modified trajectory method for practical global optimization problems”, *AIAA journal*, vol. 34, No. 10, 2126–2131, 1996.
- [168] Albert A Groenwold and JA Snyman, “Global optimization using dynamic search trajectories”, *Journal of Global Optimization*, vol. 24, No. 1, 51–60, 2002.
- [169] Shenkai Gu, Ran Cheng, and Yaochu Jin, “Multi-objective ensemble generation”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, No. 5, 234–245, 2015.
- [170] Ankit Gupta, Kishan G Mehrotra, and Chilukuri Mohan, “A clustering-based discretization for supervised learning”, *Statistics & Probability Letters*, vol. 80, No. 9, 816–824, 2010.
- [171] Isabelle Guyon and André Elisseeff, “An introduction to variable and feature selection”, *Journal of Machine Learning Research*, vol. 3, No. Mar, 1157–1182, 2003.
- [172] Jiawei Han, Jian Pei, and Micheline Kamber, *Data mining: concepts and techniques*, Elsevier, 2011.
- [173] Min Han and Weijie Ren, “Global mutual information-based feature selection approach using single-objective and multi-objective optimization”, *Neurocomputing*, vol. 168, 47–54, 2015.
- [174] Emrah Hancer, Bing Xue, Mengjie Zhang, Dervis Karaboga, and Bahriye Akay, “Pareto front feature selection based on artificial bee colony optimization”, *Information Sciences*, vol. 422, 462–479, 2018.
- [175] Johanna Hardin and David M Rocke, “Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator”, *Computational Statistics & Data Analysis*, vol. 44, No. 4, 625–638, 2004.
- [176] Haibo He and Edwardo A Garcia, “Learning from imbalanced data”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, No. 9, 1263–1284, 2009.
- [177] Zengyou He, Xiaofei Xu, and Shengchun Deng, “Discovering cluster-based local outliers”, *Pattern Recognition Letters*, vol. 24, No. 9-10, 1641–1650, 2003.
- [178] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf, “Support vector machines”, *IEEE Intelligent Systems and their applications*, vol. 13, No. 4, 18–28, 1998.
- [179] David Heckerman, Dan Geiger, and David M Chickering, “Learning bayesian networks: The combination of knowledge and statistical data”, *Machine learning*, vol. 20, No. 3, 197–243, 1995.
- [180] Pablo Hernandez-Leal, J Ariel Carrasco-Ochoa, J Fco Martínez-Trinidad, and J Arturo Olvera-Lopez, “Instancerank based on borders for instance selection”, *Pattern Recognition*, vol. 46, No. 1, 365–375, 2013.
- [181] DM Himmelblau, “Applied nonlinear programming mcgraw-hill book co”, , 1972.
- [182] Geoffrey E Hinton and Ruslan R Salakhutdinov, “Reducing the dimensionality of data with neural networks”, *science*, vol. 313, No. 5786, 504–507, 2006.
- [183] Yu Chi Ho, “The no free lunch theorem and the human-machine interface”, *IEEE Control Systems Magazine*, vol. 19, No. 3, 8–10, 1999.
- [184] Victoria Hodge and Jim Austin, “A survey of outlier detection methodologies”, *Artificial intelligence review*, vol. 22, No. 2, 85–126, 2004.

- [185] JL Hodges and Erich L Lehmann, “Rank methods for combination of independent experiments in analysis of variance”, In *Selected Works of EL Lehmann*, Springer, 403–418, 2012.
- [186] Heiko Hoffmann, “Kernel PCA for novelty detection”, *Pattern Recognition*, vol. 40, No. 3, 863–874, 2007.
- [187] Sture Holm, “A simple sequentially rejective multiple test procedure”, *Scandinavian journal of statistics*, 65–70, 1979.
- [188] Feng Honghai, Chen Guoshun, Yin Cheng, Yang Bingru, and Chen Yumei, “A SVM regression based approach to filling in missing values”, In the proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Springer, 581–587, 2005.
- [189] Reiner Horst, Panos M Pardalos, and Nguyen Van Thoai, *Introduction to global optimization*, Springer Science & Business Media, 2000.
- [190] Qinghua Hu, Xunjian Che, Lei Zhang, and Daren Yu, “Feature evaluation and selection based on neighborhood soft margin”, *Neurocomputing*, vol. 73, No. 10, 2114–2124, 2010.
- [191] Weiwei Hu and Ying Tan, “Prototype generation using multiobjective particle swarm optimization for nearest neighbor classification”, *IEEE transactions on cybernetics*, vol. 46, No. 12, 2719–2731, 2015.
- [192] Jianping Hua, Waibhav D Tembe, and Edward R Dougherty, “Performance of feature-selection methods in the classification of high-dimension data”, *Pattern Recognition*, vol. 42, No. 3, 409–424, 2009.
- [193] Jianping Hua, Zixiang Xiong, James Lowey, Edward Suh, and Edward R Dougherty, “Optimal number of features as a function of sample size for various classification rules”, *Bioinformatics*, vol. 21, No. 8, 1509–1515, 2005.
- [194] Guang Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang, “Extreme learning machine for regression and multiclass classification”, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, No. 2, 513–529, 2011.
- [195] HY Huang, JX Lin, CC Chen, and MH Fan, “Review of outlier detection”, *Application Research of Computers*, vol. 8, 002, 2006.
- [196] David Hume, *A treatise of human nature*, Courier Corporation, 2003.
- [197] D Hume and EC Mossner, “A treatise of human nature. classics series”, 1986.
- [198] C L Hwang and Abu Syed Md Masud, *Multiple objective decision making—methods and applications: a state-of-the-art survey*, oλ. 164, Springer Science & Business Media, 2012.
- [199] Sandro Incerti, Valerio Parisi, and Francesco Zirilli, “A new method for solving nonlinear simultaneous equations”, *SIAM Journal on Numerical Analysis*, vol. 16, No. 5, 779–789, 1979.
- [200] Davy Janssens, Tom Brijs, Koen Vanhoof, and Geert Wets, “Evaluating the performance of cost-based discretization versus entropy-and error-based discretization”, *Computers & Operations Research*, vol. 33, No. 11, 3107–3123, 2006.
- [201] JHM Janssens, Ferenc Huszár, EO Postma, and HJvan den Herik, “Stochastic outlier selection”, tech. rep., 2012.
- [202] José M Jerez, Ignacio Molina, Pedro J García-Laencina, Emilio Alba, Nuria Ribelles, Miguel Martín, and Leonardo Franco, “Missing data imputation using statistical and machine learning methods in a real breast cancer problem”, *Artificial Intelligence in Medicine*, vol. 50, No. 2, 105–115, 2010.

- [203] Ruoming Jin, Yuri Breitbart, and Chibuike Muoh, “Data discretization unification”, *Knowledge and Information Systems*, vol. 19, No. 1, 1–29, 2009.
- [204] Yaochu Jin and Bernhard Sendhoff, “Pareto-based multiobjective machine learning: An overview and case studies”, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, No. 3, 397–415, 2008.
- [205] Dylan F Jones, S Keyvan Mirrazavi, and Mehrdad Tamiz, “Multi-objective metaheuristics: An overview of the current state-of-the-art”, *European journal of operational research*, vol. 137, No. 1, 1–9, 2002.
- [206] TY Kam and JA Snyman, “Optimal design of laminated composite plates using a global optimization technique”, *Composite structures*, vol. 19, No. 4, 351–370, 1991.
- [207] Dervis Karaboga, “An idea based on honey bee swarm for numerical optimization”, Technical report, Technical report-tr06, Erciyes university, engineering faculty, computer . . . , 2005.
- [208] Dervis Karaboga and Bahriye Basturk, “A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm”, *Journal of global optimization*, vol. 39, No. 3, 459–471, 2007.
- [209] Spiridon A. Kazarlis, Stelios E Papadakis, JB Theocharis, and Vassilios Petridis, “Microgenetic algorithms as generalized hill-climbing operators for ga optimization”, *IEEE Transactions on Evolutionary Computation*, vol. 5, No. 3, 204–217, 2001.
- [210] Michael Kearns and Leslie Valiant, “Cryptographic limitations on learning boolean formulae and finite automata”, *Journal of the ACM (JACM)*, vol. 41, No. 1, 67–95, 1994.
- [211] James Kennedy and Russell Eberhart, “Particle swarm optimization”, In the proceedings of the Proceedings of ICNN’95-International Conference on Neural Networks, vol. 4, IEEE, 1942–1948, 1995.
- [212] Eik Fun Khor, Kay Chen Tan, Tong Heng Lee, and Chi Keong Goh, “A study on distribution preservation mechanism in evolutionary multi-objective optimization”, *Artificial Intelligence Review*, vol. 23, No. 1, 31–33, 2005.
- [213] Seung Kim, Nam Wook Cho, Bokyoung Kang, and Suk Ho Kang, “Fast outlier detection for very large log data”, *Expert Systems with Applications*, vol. 38, No. 8, 9587–9596, 2011.
- [214] S W Kim and B John Oommen, “A brief taxonomy and ranking of creative prototype reduction schemes”, *Pattern Analysis & Applications*, vol. 6, No. 3, 232–244, 2003.
- [215] Steven Orla Kimbrough, Gary J Koehler, Ming Lu, and David Harlan Wood, “On a feasible–infeasible two-population (fi-2pop) genetic algorithm for constrained optimization: Distance tracing and no free lunch”, *European Journal of Operational Research*, vol. 190, No. 2, 310–327, 2008.
- [216] Ralf Klinkenberg, “Learning drifting concepts: Example selection vs. example weighting”, *Intelligent Data Analysis*, vol. 8, No. 3, 281–300, 2004.
- [217] Levente Kocsis and Csaba Szepesvári, “Bandit based monte-carlo planning”, In the proceedings of the European conference on machine learning, Springer, 282–293, 2006.
- [218] Yves Kodratoff, *Introduction to machine learning*, Elsevier, 2014.
- [219] Mario Köppen, “Some technical remarks on the proof of the no free lunch theorem”, In the proceedings of the Proceedings of the Joint Conference on Information Sciences (JCIS 2000), Atlantic City, NJ, 1020–1024, 2000.

- [220] Mario Koppen, David H Wolpert, and William G. Macready, "Remarks on a recent paper on the "no free lunch" theorems", *IEEE Transactions on Evolutionary Computation*, vol. 5, No. 3, 295–296, 2001.
- [221] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas, "Supervised machine learning: A review of classification techniques", *Emerging artificial intelligence applications in computer engineering*, vol. 160, 3–24, 2007.
- [222] Hans Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek, "Outlier detection in axis-parallel subspaces of high dimensional data", In the proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 831–838, 2009.
- [223] Hans Peter Kriegel, Matthias Schubert, and Arthur Zimek, "Angle-based outlier detection in high-dimensional data", In the proceedings of the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 444–452, 2008.
- [224] Ajay Kumar and David Zhang, "Hand-geometry recognition using entropy-based discretization", *IEEE Transactions on Information Forensics and Security*, vol. 2, No. 2, 181–187, 2007.
- [225] Lukasz A Kurgan and Krzysztof J Cios, "Caim discretization algorithm", *IEEE transactions on Knowledge and Data Engineering*, vol. 16, No. 2, 145–153, 2004.
- [226] Elena C Laskari, Konstantinos E Parsopoulos, and Michael N Vrahatis, "Evolutionary operators in global optimization with dynamic search trajectories", *Numerical Algorithms*, vol. 34, No. 2-4, 393–403, 2003.
- [227] Aleksandar Lazarevic and Vipin Kumar, "Feature bagging for outlier detection", In the proceedings of the Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, 157–166, 2005.
- [228] Jon Lee, *A first course in combinatorial optimization*, ολ. 36, Cambridge University Press, 2004.
- [229] Deming Lei, "Multi-objective production scheduling: a survey", *The International Journal of Advanced Manufacturing Technology*, vol. 43, No. 9-10, 926, 2009.
- [230] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning", *Journal of Machine Learning Research*, vol. 18, No. 17, 1–5, 2017.
- [231] Yiu Wing Leung and Yuping Wang, "An orthogonal genetic algorithm with quantization for global numerical optimization", *IEEE Transactions on Evolutionary computation*, vol. 5, No. 1, 41–53, 2001.
- [232] Hui Li and Qingfu Zhang, "Multiobjective optimization problems with complicated pareto sets, moea/d and nsga-ii", *IEEE transactions on evolutionary computation*, vol. 13, No. 2, 284–302, 2008.
- [233] Jinyan Li, Simon Fong, Raymond K Wong, and Victor W Chu, "Adaptive multi-objective swarm fusion for imbalanced data classification", *Information Fusion*, vol. 39, 1–24, 2018.
- [234] Jin Li and Sope Taiwo, "Enhancing financial decision making using multi-objective financial genetic programming", In the proceedings of the 2006 IEEE International Conference on Evolutionary Computation, IEEE, 2171–2178, 2006.
- [235] Min Li, ShaoBo Deng, Shengzhong Feng, and Jianping Fan, "An effective discretization based on class-attribute coherence maximization", *Pattern Recognition Letters*, vol. 32, No. 15, 1962–1973, 2011.

- [236] Wei Chao Lin, Chih Fong Tsai, Shih Wen Ke, Chia Wen Hung, and William Eberle, “Learning to detect representative data for large scale instance selection”, *Journal of Systems and Software*, vol. 106, 1–8, 2015.
- [237] Chuan Liu, Wenyong Wang, Meng Wang, Fengmao Lv, and Martin Konan, “An efficient instance selection algorithm to reconstruct training set for support vector machine”, *Knowledge-Based Systems*, vol. 116, 58–73, 2017.
- [238] Fei Tony Liu, Kai Ming Ting, and Zhi Hua Zhou, “Isolation-based anomaly detection”, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, No. 1, 3, 2012.
- [239] Hui Liu, Zixing Cai, and Yong Wang, “Hybridizing particle swarm optimization with differential evolution for constrained numerical and engineering optimization”, *Applied Soft Computing*, vol. 10, No. 2, 629–640, 2010.
- [240] Huan Liu, Farhad Hussain, Chew Lim Tan, and Manoranjan Dash, “Discretization: An enabling technique”, *Data mining and knowledge discovery*, vol. 6, No. 4, 393–423, 2002.
- [241] Huan Liu and Hiroshi Motoda, “On issues of instance selection”, *Data Mining and Knowledge Discovery*, vol. 6, No. 2, 115–130, 2002.
- [242] Huan Liu and Hiroshi Motoda, *Computational methods of feature selection*, CRC Press, 2007.
- [243] Xiaoyan Liu and Huaiqing Wang, “A discretization algorithm based on a heterogeneity criterion”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, No. 9, 1166–1173, 2005.
- [244] Zhun ga Liu, Quan Pan, Jean Dezert, and Arnaud Martin, “Adaptive imputation of missing values for incomplete pattern classification”, *Pattern Recognition*, vol. 52, 85–95, 2016.
- [245] Fabio Lobato, Claudomiro Sales, Igor Araujo, Vincent Tadaiesky, Lilian Dias, Leonardo Ramos, and Adamo Santana, “Multi-objective genetic algorithm for missing data imputation”, *Pattern Recognition Letters*, vol. 68, 126–131, 2015.
- [246] Victoria López, Alberto Fernández, Jose G Moreno-Torres, and Francisco Herrera, “Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. open problems on intrinsic data characteristics”, *Expert Systems with Applications*, vol. 39, No. 7, 6585–6608, 2012.
- [247] Viktor Losing, Barbara Hammer, and Heiko Wersing, “Incremental on-line learning: A review and comparison of state of the art algorithms”, *Neurocomputing*, vol. 275, 1261–1274, 2018.
- [248] Julián Luengo, Salvador García, and Francisco Herrera, “On the choice of the best imputation methods for missing values considering three groups of classification methods”, *Knowledge and Information Systems*, vol. 32, No. 1, 77–108, 2012.
- [249] Juanjuan Luo, Licheng Jiao, and Jose A Lozano, “A sparse spectral clustering framework via multiobjective evolutionary algorithm”, *IEEE Transactions on Evolutionary Computation*, vol. 20, No. 3, 418–433, 2015.
- [250] Afsaneh Mahanipour, Hossein Nezamabadi-pour, and Bahareh Nikpour, “Using fuzzy-rough set feature selection for feature construction based on genetic programming”, In the proceedings of the 2018 3rd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), IEEE, 1–6, 2018.

- [251] Sebastián Maldonado and Richard Weber, “A wrapper method for feature selection using support vector machines”, *Information Sciences*, vol. 179, No. 13, 2208–2217, 2009.
- [252] Shweta Malhotra, Vikram Bali, and KK Paliwal, “Genetic programming and k-nearest neighbour classifier based intrusion detection model”, In the proceedings of the 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, IEEE, 42–46, 2017.
- [253] KZ Mao, “Orthogonal forward selection and backward elimination algorithms for feature subset selection”, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, No. 1, 629–634, 2004.
- [254] Elena Marchiori, “Hit miss networks with applications to instance selection”, *Journal of Machine Learning Research*, vol. 9, No. Jun, 997–1017, 2008.
- [255] James AR Marshall and Thomas G Hinton, “Beyond no free lunch: realistic algorithms for arbitrary problem classes”, In the proceedings of the IEEE Congress on Evolutionary Computation, IEEE, 1–6, 2010.
- [256] D Martín, Alejandro Rosete, Jesús Alcalá-Fdez, and Francisco Herrera, “Qar-cip-nsga-ii: A new multi-objective evolutionary algorithm to mine quantitative association rules”, *Information Sciences*, vol. 258, 1–28, 2014.
- [257] Karl Mason, Jim Duggan, and Enda Howley, “A multi-objective neural network trained with differential evolution for dynamic economic emission dispatch”, *International Journal of Electrical Power & Energy Systems*, vol. 100, 201–221, 2018.
- [258] Daniel Mateos-García, Jorge García-Gutiérrez, and José C Riquelme-Santos, “On the evolutionary optimization of k-NN by label-dependent feature weighting”, *Pattern Recognition Letters*, vol. 33, No. 16, 2232–2238, 2012.
- [259] Eitan Menahem, Lior Rokach, and Yuval Elovici, “Troika—an improved stacking schema for classification tasks”, *Information Sciences*, vol. 179, No. 24, 4097–4122, 2009.
- [260] Zbigniew Michalewicz, Thomas Logan, and Swarnalatha Swaminathan, “Evolutionary operators for continuous convex parameter spaces”, In the proceedings of the Proceedings of the 3rd Annual conference on Evolutionary Programming, World Scientific, 84–97, 1994.
- [261] Athanasios Migdalas and PM Pardalos, “A note on open problems and challenges in optimization theory and algorithms”, In *Open Problems in Optimization and Data Analysis*, Springer, 1–8, 2018.
- [262] Behrouz Minaei-Bidgoli, R Barmaki, and Mahdi Nasiri, “Mining numerical association rules via multi-objective genetic algorithms”, *Information Sciences*, vol. 233, 15–24, 2013.
- [263] Péricles BC Miranda, Ricardo BC Prudêncio, André PLF De Carvalho, and Carlos Soares, “A hybrid meta-learning architecture for multi-objective optimization of svm parameters”, *Neurocomputing*, vol. 143, 27–43, 2014.
- [264] Supriya Mishra and Meenu Chawla, “A comparative study of local outlier factor algorithms for outliers detection in data streams”, In *Emerging Technologies in Data Mining and Information Security*, Springer, 347–356, 2019.
- [265] Marcin J Mizianty, Lukasz A Kurgan, and Marek R Ogiela, “Discretization as the enabling technique for the naive bayes and semi-naive bayes-based classification”, *The Knowledge Engineering Review*, vol. 25, No. 04, 421–449, 2010.

- [266] Jose G Moreno-Torres, Troy Raeder, RociO Alaiz-RodriGuez, Nitesh V Chawla, and Francisco Herrera, “A unifying view on dataset shift in classification”, *Pattern Recognition*, vol. 45, No. 1, 521–530, 2012.
- [267] Anirban Mukhopadhyay, Ujjwal Maulik, Sanghamitra Bandyopadhyay, and Carlos Artemio Coello Coello, “A survey of multiobjective evolutionary algorithms for data mining: Part i”, *IEEE Transactions on Evolutionary Computation*, vol. 18, No. 1, 4–19, 2013.
- [268] M Narasimha Murty and V Susheela Devi, “Combination of classifiers”, In *Pattern Recognition*, Springer, 188–206, 2011.
- [269] Kaustuv Nag and Nikhil R Pal, “A multiobjective genetic programming-based ensemble for simultaneous feature selection and classification”, *IEEE transactions on cybernetics*, vol. 46, No. 2, 499–510, 2015.
- [270] Loris Nanni and Alessandra Lumini, “Prototype reduction techniques: A comparison among different approaches”, *Expert Systems with Applications*, vol. 38, No. 9, 11820–11828, 2011.
- [271] Nasser M Nasrabadi, “Pattern recognition and machine learning”, *Journal of electronic imaging*, vol. 16, No. 4, 049901, 2007.
- [272] Subrat Kumar Nayak, Pravat Kumar Rout, and Alok Kumar Jagadev, “Automatic clustering by elitism-based multi-objective differential evolution”, *International Journal of Management and Decision Making*, vol. 17, No. 1, 50–74, 2018.
- [273] Mark EJ Newman, “Power laws, pareto distributions and zipf’s law”, *Contemporary physics*, vol. 46, No. 5, 323–351, 2005.
- [274] Tien Thanh Nguyen, Alan Wee Chung Liew, Xuan Cuong Pham, and Mai Phuong Nguyen, “Optimization of ensemble classifier system based on multiple objectives genetic algorithm”, In the proceedings of the 2014 International Conference on Machine Learning and Cybernetics, vol. 1, IEEE, 46–51, 2014.
- [275] Konstantinos Nikolaidis, John Yannis Goulermas, and QH Wu, “A class boundary preserving algorithm for data condensation”, *Pattern Recognition*, vol. 44, No. 3, 704–715, 2011.
- [276] Jinsung Oh and Heesoo Hwang, “Feature enhancement of medical images using morphology-based homomorphic filter and differential evolution algorithm”, *International Journal of Control, Automation and Systems*, vol. 8, No. 4, 857–861, 2010.
- [277] Varun Kumar Ojha, Ajith Abraham, and Václav Snášel, “Ensemble of heterogeneous flexible neural trees using multiobjective genetic programming”, *Applied Soft Computing*, vol. 52, 909–924, 2017.
- [278] J Arturo Olvera-López, J Ariel Carrasco-Ochoa, J Francisco Martínez-Trinidad, and Josef Kittler, “A review of instance selection methods”, *Artificial Intelligence Review*, vol. 34, No. 2, 133–143, 2010.
- [279] Francisco Javier Ordóñez, Agapito Ledezma, and Araceli Sanchis, “Genetic approach for optimizing ensembles of classifiers.” In the proceedings of the FLAIRS conference, 89–94, 2008.
- [280] H Allen Orr, “Review of no free lunch by william a dembski”, *Boston Review*. Available on-line at <http://bostonreview.net/BR27>, vol. 3, 2002.
- [281] Nikunj C Oza and Kagan Tumer, “Classifier ensembles: Select real-world applications”, *Information Fusion*, vol. 9, No. 1, 4–20, 2008.

- [282] Vassilis P Palgianakos, Michael N Vrahatis, and George D Magoulas, “Nonmonotone methods for backpropagation training with adaptive learning rate”, In the proceedings of the CNN’99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339), vol. 3, IEEE, 1762–1767, 1999.
- [283] Mrutyunjaya Panda, “Combining multi-objective evolutionary algorithm with averaged one-dependence estimators for big data analytics”, International Journal of Computational Intelligence Studies, vol. 7, No. 1, 1–18, 2018.
- [284] Deepak Panday, Renato Cordeiro de Amorim, and Peter Lane, “Feature weighting as a tool for unsupervised feature selection”, Information Processing Letters, vol. 129, 44–52, 2018.
- [285] Gisele L Pappa and Alex A Freitas, “Evolving rule induction algorithms with multi-objective grammar-based genetic programming”, Knowledge and information systems, vol. 19, No. 3, 283–309, 2009.
- [286] Panos M Pardalos and Pierre Hansen, Data mining and mathematical programming, ολ. 45, American Mathematical Soc., 2008.
- [287] Panos M Pardalos and Athanasios Migdalas, Open Problems in Optimization and Data Analysis, ολ. 141, Springer, 2018.
- [288] Panos M Pardalos, Antanas Žilinskas, and Julius Žilinskas, “Multi-objective branch and bound”, In Non-convex Multi-Objective Optimization, Springer, 45–56, 2017.
- [289] Vilfredo Pareto, “Manuale di economica politica, societa editrice libraria”, Manual of political economy, vol. 1971, 1906.
- [290] Dong Chul Park, “Centroid neural network with weighted features”, Journal of Circuits, Systems, and Computers, vol. 18, No. 08, 1353–1367, 2009.
- [291] Konstantinos E Parsopoulos, Dimitris K Tasoulis, Nicos G Pavlidis, Vassilis P Pagiakos, and Michael N Vrahatis, “Vector evaluated differential evolution for multiobjective optimization”, In the proceedings of the Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753), vol. 1, IEEE, 204–211, 2004.
- [292] Konstantinos E Parsopoulos and Michael N. Vrahatis, “Recent approaches to global optimization problems through particle swarm optimization”, Natural computing, vol. 1, No. 2-3, 235–306, 2002.
- [293] Konstantinos E Parsopoulos and Michael N Vrahatis, “On the computation of all global minimizers through particle swarm optimization”, IEEE Transactions on evolutionary computation, vol. 8, No. 3, 211–224, 2004.
- [294] Konstantinos E Parsopoulos and Michael N Vrahatis, “Parameter selection and adaptation in unified particle swarm optimization”, Mathematical and Computer Modelling, vol. 46, No. 1-2, 198–213, 2007.
- [295] Konstantinos E Parsopoulos and Michael N Vrahatis, “Multi-objective particles swarm optimization approaches”, In Multi-objective optimization in computational intelligence: Theory and practice, IGI global, 20–42, 2008.
- [296] Konstantinos E Parsopoulos and Michael N Vrahatis, “Particle swarm optimization and intelligence: advances and applications”, , 2010.
- [297] Konstantinos E Parsopoulos, Michael N Vrahatis, and others, “Particle swarm optimization method for constrained optimization problems”, Intelligent Technologies—Theory and Application: New Trends in Intelligent Technologies, vol. 76, No. 1, 214–220, 2002.

- [298] Sujoy Paul and Swagatam Das, “Simultaneous feature selection and weighting—an evolutionary multi-objective optimization approach”, *Pattern Recognition Letters*, vol. 65, 51–59, 2015.
- [299] Ronald K Pearson, *Mining imperfect data: Dealing with contamination and incomplete records*, SIAM, 2005.
- [300] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and others, “Scikit-learn: Machine learning in python”, *Journal of machine learning research*, vol. 12, No. Oct, 2825–2830, 2011.
- [301] Mark Perakh, “The no free lunch theorems and their application to evolutionary algorithms”, *Talk Reason*, 2003.
- [302] YG Petalas, DK Tasoulis, and MN Vrahatis, “Trajectory methods for neural network training”, In the proceedings of the Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA’04), vol. 1, 400–408, 2004.
- [303] Yiannis G Petalas, Dimitris K Tasoulis, and Michael N Vrahatis, “Dynamic search trajectory methods for neural network training”, In the proceedings of the International Conference on Artificial Intelligence and Soft Computing, Springer, 241–246, 2004.
- [304] Selwyn Piramuthu, “Evaluating feature selection methods for learning in data mining applications”, *European Journal of Operational Research*, vol. 156, No. 2, 483–494, 2004.
- [305] Selwyn Piramuthu and Riyaz T Sikora, “Iterative feature construction for improving inductive learning algorithms”, *Expert Systems with Applications*, vol. 36, No. 2, 3401–3406, 2009.
- [306] Elżbieta Pkekalska, Robert PW Duin, and Pavel Paclík, “Prototype selection for dissimilarity-based classifiers”, *Pattern Recognition*, vol. 39, No. 2, 189–208, 2006.
- [307] VP Plagianakos, GD Magoulas, and MN Vrahatis, “Nonmonotone learning rules for backpropagation networks”, In the proceedings of the ICECS’99. Proceedings of ICECS’99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No. 99EX357), vol. 1, IEEE, 291–294, 1999.
- [308] Vassilis P Plagianakos, George D Magoulas, and Michael N Vrahatis, “Deterministic nonmonotone strategies for effective training of multilayer perceptrons”, *IEEE Transactions on Neural Networks*, vol. 13, No. 6, 1268–1284, 2002.
- [309] Riccardo Poli and Mario Graff, “There is a free lunch for hyper-heuristics, genetic programming and computer scientists”, In the proceedings of the European conference on genetic programming, Springer, 195–207, 2009.
- [310] Riccardo Poli, Mario Graff, and Nicholas Freitag McPhee, “Free lunches for function and program induction”, In the proceedings of the Proceedings of the tenth ACM SIGEVO workshop on Foundations of genetic algorithms, 183–194, 2009.
- [311] Boris T Polyak, “Some methods of speeding up the convergence of iteration methods”, *USSR Computational Mathematics and Mathematical Physics*, vol. 4, No. 5, 1–17, 1964.
- [312] Zeinab Khatoun Pourtaheri, Seyed Hamid Zahiri, and Seyed Mohammad Razavi, “Stability investigation of multi-objective heuristic ensemble classifiers”, *International Journal of Machine Learning and Cybernetics*, vol. 10, No. 5, 1109–1121, 2019.

- [313] Michael JD Powell, “A fast algorithm for nonlinearly constrained optimization calculations”, In *Numerical analysis*, Springer, 144–157, 1978.
- [314] WL1551847 Price, “Global optimization by controlled random search”, *Journal of Optimization Theory and Applications*, vol. 40, No. 3, 333–348, 1983.
- [315] Eloi Puertas, Sergio Escalera, and Oriol Pujol, “Generalized multi-scale stacked sequential learning for multi-class classification”, *Pattern Analysis and Applications*, vol. 18, No. 2, 247–261, 2015.
- [316] Dorian Pyle, *Data preparation for data mining*, ολ. 1, Morgan Kaufmann, 1999.
- [317] Yongsong Qin, Shichao Zhang, Xiaofeng Zhu, Jilian Zhang, and Chengqi Zhang, “POP algorithm: Kernel-based imputation to treat missing values in knowledge discovery from databases”, *Expert Systems with Applications*, vol. 36, No. 2, 2794–2804, 2009.
- [318] Joaquin Quionero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence, *Dataset shift in machine learning*, MIT Press, 2009.
- [319] Sergio Ramírez-Gallego, Salvador García, Héctor Mouriño-Talín, David Martínez-Rego, Verónica Bolón-Canedo, Amparo Alonso-Betanzos, José Manuel Benítez, and Francisco Herrera, “Data discretization: taxonomy and big data challenge”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 6, No. 1, 5–21, 2016.
- [320] Sergio Ramírez-Gallego, Bartosz Krawczyk, Salvador García, Michał Woźniak, and Francisco Herrera, “A survey on data preprocessing for data stream mining: Current status and future directions”, *Neurocomputing*, 2017.
- [321] Nalluri MadhuSudana Rao, Krithivasan Kannan, Xiao zhi Gao, and Diptendu Sinha Roy, “Novel classifiers for intelligent disease diagnosis with multi-objective parameter evolution”, *Computers & Electrical Engineering*, vol. 67, 483–496, 2018.
- [322] Sebastian Raschka, “Mlxtend: Providing machine learning and data science utilities and extensions to python’s scientific computing stack”, *Journal of open source software*, vol. 3, No. 24, 638, 2018.
- [323] Marcos Raydan, “The barzilai and borwein gradient method for the large scale unconstrained minimization problem”, *SIAM Journal on Optimization*, vol. 7, No. 1, 26–33, 1997.
- [324] Marek Reformat, Witold Pedrycz, and Nicolino J Pizzi, “Software quality analysis with the use of computational intelligence”, *Information and Software Technology*, vol. 45, No. 7, 405–417, 2003.
- [325] Thomas Reinartz, “A unifying view on instance selection”, *Data Mining and Knowledge Discovery*, vol. 6, No. 2, 191–210, 2002.
- [326] AHG Rinnooy Kan and GT Timmer, “Stochastic global optimization methods part ii: Multi level methods”, *Technical report*, 1985.
- [327] Isabelle Rivals and Léon Personnaz, “On cross validation for model selection”, *Neural computation*, vol. 11, No. 4, 863–870, 1999.
- [328] Juan D Rodríguez and Jose A Lozano, “Multi-objective learning of multi-dimensional bayesian classifiers”, In *the proceedings of the 2008 Eighth International Conference on Hybrid Intelligent Systems*, IEEE, 501–506, 2008.
- [329] Ismael Rodríguez-Fdez, Adrián Canosa, Manuel Mucientes, and Alberto Bugarín, “Stac: a web platform for the comparison of algorithms using statistical tests”, In *the proceedings of the 2015 IEEE international conference on fuzzy systems (FUZZ-IEEE)*, IEEE, 1–8, 2015.

- [330] Alejandro Rosales-Pérez, Salvador García, Jesus A Gonzalez, Carlos A Coello Coello, and Francisco Herrera, “An evolutionary multiobjective model and instance selection for support vector machines with pareto-based ensembles”, *IEEE Transactions on Evolutionary Computation*, vol. 21, No. 6, 863–877, 2017.
- [331] Louis B Rosenberg, “Human swarms, a real-time method for collective intelligence”, In the proceedings of the Artificial Life Conference Proceedings 13, MIT Press, 658–659, 2015.
- [332] Tim Roughgarden, “Beyond worst-case analysis”, *Communications of the ACM*, vol. 62, No. 3, 88–96, 2019.
- [333] Filip Rudziński, “A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers”, *Applied Soft Computing*, vol. 38, 118–133, 2016.
- [334] Nasser R Sabar, Xun Yi, and Andy Song, “A bi-objective hyper-heuristic support vector machines for big data cyber-security”, *IEEE Access*, vol. 6, 10421–10431, 2018.
- [335] José A Sáez, Mikel Galar, Julián Luengo, and Francisco Herrera, “INFFC: An iterative class noise filter based on the fusion of classifiers with noise sensitivity control”, *Information Fusion*, vol. 27, 19–32, 2016.
- [336] José A Sáez, Julián Luengo, and Francisco Herrera, “Predicting noise filtering efficacy with data complexity measures for nearest neighbor classification”, *Pattern Recognition*, vol. 46, No. 1, 355–364, 2013.
- [337] Luigi Salvadori, “Famiglie ad un parametro di funzioni di liapunov nello studio della stabilita”, In the proceedings of the Symposia Mathematica, vol. 6, 309–330, 1971.
- [338] Sandhya Samarasinghe, *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*, Crc Press, 2016.
- [339] Stefan Schäfler and Hubert Warsitz, “A trajectory-following method for unconstrained optimization”, *Journal of Optimization Theory and Applications*, vol. 67, No. 1, 133–140, 1990.
- [340] Robert J Schalkoff, *Artificial neural networks*, McGraw-Hill Higher Education, 1997.
- [341] Jürgen Schmidhuber, “Deep learning in neural networks: An overview”, *Neural networks*, vol. 61, 85–117, 2015.
- [342] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson, “Estimating the support of a high-dimensional distribution”, *Neural computation*, vol. 13, No. 7, 1443–1471, 2001.
- [343] Chris Schumacher, Michael D Vose, and L Darrell Whitley, “The no free lunch and problem description length”, In the proceedings of the Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001), 565–570, 2001.
- [344] Alexander K Seewald, “How to make stacking better and faster while also taking care of an unknown weakness”, In the proceedings of the Proceedings of the nineteenth international conference on machine learning, 554–561, 2002.
- [345] Nicola Segata, Enrico Blanzieri, Sarah Jane Delany, and Pádraig Cunningham, “Noise reduction for instance-based learning with a local maximal margin approach”, *Journal of Intelligent Information Systems*, vol. 35, No. 2, 301–331, 2010.
- [346] Saptarshi Sengupta, Sanchita Basak, and Richard Alan Peters, “Particle swarm optimization: A survey of historical and recent developments with hybridization perspectives”, *Machine Learning and Knowledge Extraction*, vol. 1, No. 1, 157–191, 2019.

- [347] M Paz Sesmero, Agapito I Ledezma, and Araceli Sanchis, “Generating ensembles of heterogeneous classifiers using stacked generalization”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, No. 1, 21–34, 2015.
- [348] Yi Shang and Benjamin W Wah, “Global optimization for neural network training”, *Computer*, vol. 29, No. 3, 45–54, 1996.
- [349] Chuanhe Shen, Xiangrong Wang, and Di Yu, “Feature weighting of support vector machines based on derivative saliency analysis and its application to financial data mining”, *International Journal of Advancements in Computing Technology*, vol. 4, No. 1, 199–206, 2012.
- [350] Wenhao Shu and Hong Shen, “Multi-criteria feature selection on cost-sensitive data with missing values”, *Pattern Recognition*, vol. 51, 268–280, 2016.
- [351] Palanisamy Shunmugapriya and S Kanmani, “Optimization of stacking ensemble configurations through artificial bee colony algorithm”, *Swarm and Evolutionary Computation*, vol. 12, 24–32, 2013.
- [352] Mei Ling Shyu, Shu Ching Chen, Kanoksri Sarinnapakorn, and LiWu Chang, “A novel anomaly detection scheme based on principal component classifier”, Technical report, MIAMI UNIV CORAL GABLES FL DEPT OF ELECTRICAL AND COMPUTER ENGINEERING, 2003.
- [353] Esther Lydia Silva-Ramírez, Rafael Pino-Mejías, Manuel López-Coello, and María Dolores Cubiles de la Vega, “Missing value imputation on missing completely at random data using multilayer perceptrons”, *Neural Networks*, vol. 24, No. 1, 121–129, 2011.
- [354] Jaemun Sim, Ohbyung Kwon, and Kun Chang Lee, “Adaptive pairing of classifier and imputation methods based on the characteristics of missing values in data sets”, *Expert Systems with Applications*, vol. 46, 485–493, 2016.
- [355] Karanjit Singh and Shuchita Upadhyaya, “Outlier detection: applications and techniques”, *International Journal of Computer Science Issues (CSI)*, vol. 9, No. 1, 307, 2012.
- [356] David B Skillicorn and Sabine M McConnell, “Distributed prediction from vertically partitioned data”, *Journal of Parallel and Distributed Computing*, vol. 68, No. 1, 16–36, 2008.
- [357] Matthew G Smith and Larry Bull, “Genetic programming with a genetic algorithm for feature construction and selection”, *Genetic Programming and Evolvable Machines*, vol. 6, No. 3, 265–281, 2005.
- [358] Michael R Smith and Tony Martinez, “Improving classification accuracy by identifying and removing instances that should be misclassified”, In the proceedings of the The 2011 International Joint Conference on Neural Networks, IEEE, 2690–2697, 2011.
- [359] JA Snyman and LP Fatti, “A multi-start global minimization algorithm with dynamic search trajectories”, *Journal of Optimization Theory and Applications*, vol. 54, No. 1, 121–141, 1987.
- [360] JA Snyman and KA Geerthsen, “The practical application of a dynamic search-trajectory method for constrained global optimization”, In the proceedings of the IUTAM Symposium on Optimization of Mechanical Systems, Springer, 285–292, 1996.
- [361] JA Snyman and AM Hay, “The dynamic-q optimization method: An alternative to sqp?” *Computers & Mathematics with Applications*, vol. 44, No. 12, 1589–1598, 2002.

- [362] JA Snyman, Nielen Stander, and WJ Roux, “A dynamic penalty function method for the solution of structural optimization problems”, *Applied Mathematical Modelling*, vol. 18, No. 8, 453–460, 1994.
- [363] Johannes Arnoldus Snyman, “A new and dynamic method for unconstrained minimization”, *Applied Mathematical Modelling*, vol. 6, No. 6, 449–462, 1982.
- [364] Johannes Arnoldus Snyman, “An improved version of the original leap-frog dynamic method for unconstrained minimization: Lfop1 (b)”, *Applied mathematical modelling*, vol. 7, No. 3, 216–218, 1983.
- [365] Jan A Snyman, “New gradient-based trajectory and approximation methods”, *Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms*, 97–150, 2005.
- [366] Jan A Snyman and Schalk Kok, “A reassessment of the snyman–fatti dynamic search trajectory method for unconstrained global optimization”, *Journal of Global Optimization*, vol. 43, No. 1, 67–82, 2009.
- [367] Jan A. Snyman and Daniel N. Wilke, *NEW GRADIENT-BASED TRAJECTORY AND APPROXIMATION METHODS*, Springer International Publishing, Cham, 2018, 197–250.
- [368] Jan A Snyman and Daniel N Wilke, “New gradient-based trajectory and approximation methods”, In *Practical Mathematical Optimization*, Springer, 197–250, 2018.
- [369] Jan A Snyman and Daniel N Wilke, *Practical Mathematical Optimization: Basic Optimization Theory and Gradient-Based Algorithms*, ολ. 133, Springer, 2018.
- [370] Petr Somol and Pavel Pudil, “Feature selection toolbox”, *Pattern Recognition*, vol. 35, No. 12, 2749–2759, 2002.
- [371] Kenneth Sörensen, “Metaheuristics—the metaphor exposed”, *International Transactions in Operational Research*, vol. 22, No. 1, 3–18, 2015.
- [372] Ömer M Soysal, “Association rule mining with mostly associated sequential patterns”, *Expert Systems with applications*, vol. 42, No. 5, 2582–2592, 2015.
- [373] James C Spall, *Introduction to stochastic search and optimization: estimation, simulation, and control*, ολ. 65, John Wiley & Sons, 2005.
- [374] Sujatha Srinivasan and Sivakumar Ramakrishnan, “Evolutionary multi objective optimization for rule mining: a review”, *Artificial Intelligence Review*, vol. 36, No. 3, 205, 2011.
- [375] Rainer Storn and Kenneth Price, “Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces”, *Journal of global optimization*, vol. 11, No. 4, 341–359, 1997.
- [376] Marc Sumner, Eibe Frank, and Mark Hall, “Speeding up logistic model tree induction”, In *the proceedings of the European conference on principles of data mining and knowledge discovery*, Springer, 675–683, 2005.
- [377] Yanmin Sun, Andrew KC Wong, and Mohamed S Kamel, “Classification of imbalanced data: a review”, *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, No. 04, 687–719, 2009.
- [378] Marzieh Hajizadeh Tahan and Shahrokh Asadi, “Memod: a novel multivariate evolutionary multi-objective discretization”, *Soft Computing*, vol. 22, No. 1, 301–323, 2018.
- [379] Choo Jun Tan, Chee Peng Lim, and Yu N Cheah, “A multi-objective evolutionary algorithm-based ensemble optimizer for feature selection and classification with neural network models”, *Neurocomputing*, vol. 125, 217–228, 2014.

- [380] Daniel R Tauritz and others, “A no-free-lunch framework for coevolution”, In the proceedings of the Proceedings of the 10th annual conference on Genetic and evolutionary computation, ACM, 371–378, 2008.
- [381] Georgios S Temponeras, Stamatios Aggelos N Alexandropoulos, Sotiris B Kotsiantis, and Michael N Vrahatis, “Financial fraudulent statements detection through a deep dense artificial neural network”, In the proceedings of the 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), IEEE, 1–5, 2019.
- [382] Olivier Teytaud and Sébastien Flory, “Upper confidence trees with short term partial information”, In the proceedings of the European Conference on the Applications of Evolutionary Computation, Springer, 153–162, 2011.
- [383] K Ting and I Witten, “Issues in stacked generalization, ai research 10, 271–289”, 1999.
- [384] Aimo Torn and Antanas Zilinskas, Global optimization, Springer-Verlag, 1989.
- [385] Isaac Triguero, Joaquín Derrac, Salvador García, and Francisco Herrera, “Integrating a differential evolution feature weighting scheme into prototype generation”, Neurocomputing, vol. 97, 332–343, 2012.
- [386] Isaac Triguero, Joaquín Derrac, Salvador Garcia, and Francisco Herrera, “A taxonomy and experimental study on prototype generation for nearest neighbor classification”, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, No. 1, 86–100, 2012.
- [387] Chih Fong Tsai and Fu Yu Chang, “Combining instance selection for better missing value imputation”, Journal of Systems and Software, vol. 122, 63–71, 2016.
- [388] Cheng Jung Tsai, Chien I Lee, and Wei Pang Yang, “A discretization algorithm based on class-attribute contingency coefficient”, Information Sciences, vol. 178, No. 3, 714–731, 2008.
- [389] Lin Yu Tseng and Chun Chen, “Multiple trajectory search for multiobjective optimization”, In the proceedings of the 2007 IEEE congress on evolutionary computation, IEEE, 3609–3616, 2007.
- [390] Lin Yu Tseng and Chun Chen, “Multiple trajectory search for large scale global optimization”, In the proceedings of the 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), IEEE, 3052–3059, 2008.
- [391] Alper Unler, Alper Murat, and Ratna Babu Chinnam, “mr 2 PSO: a maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification”, Information Sciences, vol. 181, No. 20, 4625–4641, 2011.
- [392] Milagros Van Grieken, Optimisation pour l’apprentissage et apprentissage pour l’optimisation, Ph.D. thesis, Université Paul Sabatier-Toulouse III, 2004.
- [393] Jason Van Hulse and Taghi Khoshgoftaar, “Knowledge discovery from imbalanced and noisy data”, Data & Knowledge Engineering, vol. 68, No. 12, 1513–1542, 2009.
- [394] Jason D Van Hulse, Taghi M Khoshgoftaar, and Haiying Huang, “The pairwise attribute noise detection algorithm”, Knowledge and Information Systems, vol. 11, No. 2, 171–190, 2007.
- [395] David A Van Veldhuizen and Gary B Lamont, “Multiobjective evolutionary algorithms: Analyzing the state-of-the-art”, Evolutionary computation, vol. 8, No. 2, 125–147, 2000.

- [396] Charlie Vanaret, François Gallard, and Joaquim Martins, “On the consequences of the “no free lunch” theorem for optimization on the choice of an appropriate mdo architecture”, In the proceedings of the 18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, p. 3148, 2017.
- [397] Robert J Vanderbei and others, Linear programming, Springer, 2015.
- [398] Stephen A Vavasis, “Complexity issues in global optimization: a survey”, In Handbook of global optimization, Springer, 27–41, 1995.
- [399] William T Vetterling, William H Press, Saul A Teukolsky, and Brian P Flannery, Numerical Recipes Example Book (C++): The Art of Scientific Computing, Cambridge University Press, 2002.
- [400] TL Vincent, BS Goh, and KL Teo, “Trajectory-following algorithms for min-max optimization problems”, Journal of optimization theory and applications, vol. 75, No. 3, 501–519, 1992.
- [401] Marco Virgolin, Tanja Alderliesten, Arjan Bel, Cees Witteveen, and Peter AN Bosman, “Symbolic regression and feature construction with gp-gomea applied to radiotherapy dose reconstruction of childhood cancer survivors”, In the proceedings of the Proceedings of the Genetic and Evolutionary Computation Conference, ACM, 1395–1402, 2018.
- [402] Thomas P Vogl, JK Mangis, AK Rigler, WT Zink, and DL Alkon, “Accelerating the convergence of the back-propagation method”, Biological cybernetics, vol. 59, No. 4-5, 257–263, 1988.
- [403] Christian Von Lücken, Benjamín Barán, and Carlos Brizuela, “A survey on multi-objective evolutionary algorithms for many-objective problems”, Computational optimization and applications, vol. 58, No. 3, 707–756, 2014.
- [404] MN Vrahatis, GD Magoulas, and VP Plagianakos, “From linear to nonlinear iterative methods”, Applied Numerical Mathematics, vol. 45, No. 1, 59–77, 2003.
- [405] David J Wales and Jonathan PK Doye, “Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms”, The Journal of Physical Chemistry A, vol. 101, No. 28, 5111–5116, 1997.
- [406] W Walter, “Gewöhnliche differentialgleichungen springer verlag”, 2000.
- [407] BX Wang and N Japkowicz, “Imbalanced data set learning with synthetic samples”, In the proceedings of the Proc. IRIS Machine Learning Workshop, p. 19, 2004.
- [408] Rui Wang, Shiming Lai, Guohua Wu, Lining Xing, Ling Wang, and Hisao Ishibuchi, “Multi-clustering via evolutionary multi-objective optimization”, Information Sciences, vol. 450, 128–140, 2018.
- [409] S K Wang, J P Chiou, and C W Liu, “Non-smooth/non-convex economic dispatch by a novel hybrid differential evolution algorithm”, IET Generation, Transmission & Distribution, vol. 1, No. 5, 793–803, 2007.
- [410] Zhichun Wang, Minqiang Li, and Juanzi Li, “A multi-objective evolutionary algorithm for feature selection based on mutual information with a new redundancy measure”, Information Sciences, vol. 307, 73–88, 2015.
- [411] Dietrich Wettschereck, David W Aha, and Takao Mohri, “A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms”, Artificial Intelligence Review, vol. 11, No. 1-5, 273–314, 1997.
- [412] D Randall Wilson and Tony R Martinez, “Reduction techniques for instance-based learning algorithms”, Machine learning, vol. 38, No. 3, 257–286, 2000.

- [413] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2016.
- [414] Philip Wolfe, “Convergence conditions for ascent methods”, *SIAM review*, vol. 11, No. 2, 226–235, 1969.
- [415] Philip Wolfe, “Convergence conditions for ascent methods. ii: Some corrections”, *SIAM review*, vol. 13, No. 2, 185–188, 1971.
- [416] DH Wolpert and WG Macready, “No free lunch theorems for search-tech. rep”, Technical report, SFI-TR-95-02-010, Santa Fe Institute, 1995.
- [417] D Wolpert and W Macready, “What makes an optimization problem hard”, The Santa Fe Institute. Santa Fe, NM, 1996.
- [418] David H Wolpert, “On the connection between in-sample testing and generalization error”, *Complex Systems*, vol. 6, No. 1, 47, 1992.
- [419] David H Wolpert, “Stacked generalization”, *Neural networks*, vol. 5, No. 2, 241–259, 1992.
- [420] David H Wolpert, “The existence of a priori distinctions between learning algorithms”, *Neural Computation*, vol. 8, No. 7, 1391–1420, 1996.
- [421] David H Wolpert, “The lack of a priori distinctions between learning algorithms”, *Neural computation*, vol. 8, No. 7, 1341–1390, 1996.
- [422] David H Wolpert, “The supervised learning no-free-lunch theorems”, In *Soft computing and industry*, Springer, 25–42, 2002.
- [423] David H Wolpert, “What the no free lunch theorems really mean; how to improve search algorithms”, In *the proceedings of the Santa Fe Institute*, vol. 7, 2012.
- [424] David H Wolpert and William G Macready, “No free lunch theorems for optimization”, *IEEE transactions on evolutionary computation*, vol. 1, No. 1, 67–82, 1997.
- [425] David H Wolpert and William G Macready, “Coevolutionary free lunches”, *IEEE Transactions on evolutionary computation*, vol. 9, No. 6, 721–735, 2005.
- [426] Tzu Tsung Wong, “A hybrid discretization method for naïve bayesian classifiers”, *Pattern Recognition*, vol. 45, No. 6, 2321–2325, 2012.
- [427] Xindong Wu and Xingquan Zhu, “Mining with noise knowledge: error-aware data mining”, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, No. 4, 917–932, 2008.
- [428] Bing Xue, Mengjie Zhang, and Will N Browne, “Particle swarm optimization for feature selection in classification: A multi-objective approach”, *IEEE transactions on cybernetics*, vol. 43, No. 6, 1656–1671, 2012.
- [429] Xin She Yang, “Firefly algorithm, stochastic test functions and design optimisation”, *arXiv preprint arXiv:1003.1409*, 2010.
- [430] Xin She Yang, “A new metaheuristic bat-inspired algorithm”, In *Nature inspired cooperative strategies for optimization (NISCO 2010)*, Springer, 65–74, 2010.
- [431] Xin She Yang, “Swarm-based metaheuristic algorithms and no-free-lunch theorems”, *Theory and new applications of swarm intelligence*, vol. 9, 1–16, 2012.
- [432] Xin She Yang and Suash Deb, “Cuckoo search via lévy flights”, In *the proceedings of the 2009 World congress on nature & biologically inspired computing (NaBIC)*, IEEE, 210–214, 2009.
- [433] Ying Yang, Geoffrey I Webb, and Xindong Wu, “Discretization methods”, In *Data mining and knowledge discovery handbook*, Springer, 101–116, 2009.

- [434] Zeld B Zabinsky, Stochastic adaptive search for global optimization, ολ. 72, Springer Science & Business Media, 2013.
- [435] Shichao Zhang, “Shell-neighbor method and its application in missing data imputation”, Applied Intelligence, vol. 35, No. 1, 123–133, 2011.
- [436] Shichao Zhang, Zhi Jin, and Xiaofeng Zhu, “Missing data imputation by utilizing information within incomplete instances”, Journal of Systems and Software, vol. 84, No. 3, 452–459, 2011.
- [437] Shichao Zhang, Jilian Zhang, Xiaofeng Zhu, Yongsong Qin, and Chengqi Zhang, “Missing value imputation based on data clustering”, In Transactions on Computational Science I, Springer, 128–138, 2008.
- [438] Huimin Zhao, “A multi-objective genetic programming approach to developing pareto optimal decision trees”, Decision Support Systems, vol. 43, No. 3, 809–826, 2007.
- [439] Jiaqi Zhao, Vitor Basto Fernandes, Licheng Jiao, Iryna Yevseyeva, Asep Maulana, Rui Li, Thomas Bäck, Ke Tang, and Michael TM Emmerich, “Multiobjective optimization of classifiers by means of 3d convex-hull-based evolutionary algorithms”, Information Sciences, vol. 367, 80–104, 2016.
- [440] Yue Zhao, Zain Nasrullah, Maciej K Hryniewicki, and Zheng Li, “Lscp: Locally selective combination in parallel outlier ensembles”, In the proceedings of the Proceedings of the 2019 SIAM International Conference on Data Mining, SIAM, 585–593, 2019.
- [441] Yue Zhao, Zain Nasrullah, and Zheng Li, “Pyod: A python toolbox for scalable outlier detection”, arXiv preprint arXiv:1901.01588, 2019.
- [442] Kangfeng Zheng and Xiujuan Wang, “Feature selection method with joint maximal information entropy between features and class”, Pattern Recognition, vol. 77, 20–29, 2018.
- [443] Aimin Zhou, Bo Yang Qu, Hui Li, Shi Zheng Zhao, Ponnuthurai Nagaratnam Suganthan, and Qingfu Zhang, “Multiobjective evolutionary algorithms: A survey of the state of the art”, Swarm and Evolutionary Computation, vol. 1, No. 1, 32–49, 2011.
- [444] Ya’nan Zhou, Yuehong Chen, Li Feng, Xin Zhang, Zhanfeng Shen, and Xiaocheng Zhou, “Supervised and adaptive feature weighting for object-based classification on satellite images”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, No. 9, 3224–3234, 2018.
- [445] Huaiyu Zhu and Richard Rohwer, “No free lunch for cross-validation”, Neural Computation, vol. 8, No. 7, 1421–1426, 1996.
- [446] Eckart Zitzler and Lothar Thiele, “Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach”, IEEE transactions on Evolutionary Computation, vol. 3, No. 4, 257–271, 1999.
- [447] G Zoutendijk, “Nonlinear programming, computational methods”, Integer and nonlinear programming, 37–86, 1970.

Κατάλογος Δημοσιεύσεων Υποψηφίου

Δημοσιεύσεις σε Επιστημονικά Περιοδικά με Σύστημα Κριτών :

- J1. Alexandropoulos S.-A.N., Pardalos P.M., Vrahatis M.N., Dynamic search trajectory methods for global optimization, *Annals of Mathematics and Artificial Intelligence*, 88(1-3), pp.3-37, 2020.
- J2. Alexandropoulos S.-A.N., Kotsiantis S.B., Vrahatis M.N., Data preprocessing in predictive data mining, *Knowledge Engineering Review*, 34, Article No. e1, pp.1-33, 2019.

Άρθρα και Κεφάλαια σε Βιβλία και Συλλογικούς Τόμους με Σύστημα Κριτών :

- V1. Alexandropoulos S.-A.N., Aridas C.K., Kotsiantis S.B., Vrahatis M.N., A deep dense neural network for bankruptcy prediction, *Communications in Computer and Information Science (CCIS)*, vol.1000, pp.435-444, Springer Nature Switzerland AG, Switzerland, 2019.
- V2. Alexandropoulos S.-A.N., Aridas C.K., Kotsiantis S.B., Vrahatis M.N., Stacking strong ensembles of classifiers, *IFIP Advances in Information and Communication Technology*, vol.559, pp.545-556, Springer Nature Switzerland AG, Switzerland, 2019.
- V3. Aridas C.K., Alexandropoulos S.-A.N., Kotsiantis S.B., Vrahatis M.N., Random re-sampling in the one-versus-all strategy for handling multi-class problems, *Communications in Computer and Information Science (CCIS)*, vol.744, pp.111-121, Springer, 2017.

Δημοσιεύσεις σε Διεθνή Επιστημονικά Συνέδρια με Σύστημα Κριτών :

- C1. Alexandropoulos S.-A.N., Kotsiantis S.B., Piperigou V.E., Vrahatis M.N., A new ensemble method for outlier identification, *Proceedings of the IEEE Tenth International Conference on Cloud Computing, Data Science Engineering (Confluence 2020)*, January 29-31, 2020, Noida, Uttar Pradesh, India, pp.786-791, IEEE 2020.
- C2. Temponeras G.S., Alexandropoulos S.-A.N., Kotsiantis S.B., Vrahatis M.N., Financial fraudulent statements detection through a deep dense artificial neural network, *Proceedings of the IEEE Tenth International Conference on Information, Intelligence, Systems and Applications (IISA 2019)*, July 15-17, 2019, Patras, Greece, art.no.8900741, pp.1-5, IEEE 2019.
- C3. Meletiou G.C., Alexandropoulos S.-A.N., Vrahatis M.N., Comparative e-evaluations gathering in open and distance learning, in Greek, *Proceedings of the Seventh International Conference in Open and Distance Learning (ICODL 2013)*, Learning

Methodologies', November 8-10, 2013, Athens, Greece, A. Lionarakis (ed.), Vol.5, Part B, pp.159-164, Hellenic Open University, OPENET, Athens, Greece, December 2013 [ISBN: 978-618-81051-5-7].

Δημοσιεύσεις σε Κεφάλαια Βιβλίων μετά από Πρόσκληση και Κρίση:

- Ch1. Alexandropoulos S.-A.N., Aridas C.K., Kotsiantis S.B., Vrahatis M.N., Multi-objective evolutionary optimization algorithms for machine learning: A recent survey, *Approximation and Optimization*, I.C. Demetriou and P.M. Pardalos (eds.), Chapter 4, pp.35-55, Springer Optimization and Its Applications 145, Springer Nature Switzerland AG 2019 [ISBN: 978-3-030-12766-4, ISBN: 978-3-030-12767-1 (eBook)].
- Ch2. Adam S.P., Alexandropoulos S.-A.N., Pardalos P.M., Vrahatis M.N., No free lunch theorem: A review, *Approximation and Optimization*, I.C. Demetriou and P.M. Pardalos (eds.), Chapter 5, pp.57-82, Springer Optimization and Its Applications 145, Springer Nature Switzerland AG 2019 [ISBN: 978-3-030-12766-4, ISBN: 978-3-030-12767-1 (eBook)].
- Ch3. Alexandropoulos S.-A.N., Meletiou G.C., Triantafyllou D.S., Vrahatis M.N., Transformations of cryptographic schemes through interpolation techniques, *Computation, Cryptography, and Network Security*, N.J. Daras and M.Th. Rassias (eds.), Chapter 1, pp.1-17, Springer International Publishing, Switzerland, 2015 [ISBN: 978-3-319-18274-2, ISBN: 978-3-319-18275-9 (eBook)].

Άλλες επιστημονικές δημοσιεύσεις:

- O1. Alexandropoulos S.-A.N., Meletiou G.C., Triantafyllou D.S., Vrahatis M.N., Bottom-up hierarchical ramp secret sharing scheme, abstract, Fourth International Conference on Operational Planning, Technological Innovations and Mathematical Applications (OPTIMA 2017), May 25-26, 2017, Hellenic Military Academy, Athens, Greece, N.J. Daras (ed.), pp.132-134, 2017.
- O2. Alexandropoulos S.-A.N., Meletiou G.C., Triantafyllou D.S., Vrahatis M.N., Secret sharing schemes through structured matrices, abstract, Third International Conference on Cryptography, Cyber Security and Information Warfare (CryCybiW 2016), May 26-27, 2016, Hellenic Military Academy, Athens, Greece, N.J. Daras (ed.), pp.139-140, 2016.
- O3. Meletiou G.C., Alexandropoulos S.-A.N., Vrahatis M.N., Boolean functions and neural networks, abstract, Third International Conference on Technology Trends and Scientific Applications in Artillery and other Military Science (TTSAAMS 2015), May 5-6, 2015, Hellenic Artillery School, Athens, Greece, N.J. Daras (ed.), p.34, 2015.

Διακρίσεις:

- Υποτροφία εκπόνησης Διδακτορικής Διατριβής: « Το έργο συγχρηματοδοτείται από την Ελλάδα και την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) μέσω του Επιχειρησιακού Προγράμματος «Ανάπτυξη Ανθρώπινου Δυναμικού, Εκπαίδευση και Διά Βίου Μάθηση», στο πλαίσιο της Πράξης «Ενίσχυση του ανθρώπινου ερευνητικού δυναμικού μέσω της υλοποίησης διδακτορικής έρευνας» (MIS-5000432), που υλοποιεί το Ίδρυμα Κρατικών Υποτροφιών (ΙΚΥ) »

- k Nearest Neighbors (kNN) Detector, 123
- Άμεσοι Μέθοδοι, 8
- Έμμεσοι Μέθοδοι, 8
- έγκαιρη διακοπή, 34
- Αντικειμενική συνάρτηση, 147
- Εξόρυξη Δεδομένων, 10, 81
- Εξόρυξη Δεδομένων (ΕΔ), 81
- Μέθοδοι εγγυημένης ακρίβειας, 8
- Μηχανική Μάθηση, 10
- Φυσικός Υπολογισμός, 3
- Προεπεξεργασία Δεδομένων (ΠΔ), 81
- Προετοιμασία Δεδομένων, 81
- Υπολογιστικά Μαθηματικά, 6
- Υπολογιστική Νοημοσύνη, 3, 5
- ακραία τιμή, 121
- ανωμαλίες, 86
- αντικειμενικές συναρτήσεις, 150
- αρχή του Hamilton, 47
- αρχή Pareto, 149
- αυτόνομο αρχικό πρόβλημα τιμών, 65
- δέντρα απόφασης, 155
- δειγματοληψία, 104
- δωρεάν γεύματα, 15
- δυναμικές τροχιές αναζήτησης, 40
- επιλογή παραδειγμάτων, 104
- ευρετικές μέθοδοι, 40
- εξελικτικοί αλγόριθμοι, 151
- κανόνες συσχέτισης, 147
- κανονικοποίηση δεδομένων, 97
- κατασκευή χαρακτηριστικών, 114
- κατασκευή/μετασχηματισμός χαρακτηριστικών, 113
- μέτωπο Pareto, 150
- μέθοδοι γενικευμένης μείωσης, 8
- μέθοδοι κάλυψης, 8
- μέθοδοι τυχαίας αναζήτησης, 8
- μέθοδος του Euler, 65
- μέθοδος του Heun, 67
- μέθοδος Euler-Cauchy, 67
- μέθοδος Ralston, 67
- μαθηματικός προγραμματισμός, 159
- μη μονοτονική στρατηγική σύγκλισης, 70
- μηχανή διανυσματικής υποστήριξης, 157
- μονο-κριτηριακή βελτιστοποίηση, 9
- περιορισμοί, 150
- πολλαπλών αντικειμένων, 148
- πολυ-αντικειμενική βελτιστοποίηση, 10
- προσέγγιση Pareto, 148
- σύνολο βέλτιστων λύσεων Pareto, 150
- στοίβαξη, 135
- στρωματοποιημένη δειγματοληψία, 104
- συνεργατική μέθοδο, 121
- τιμή καταλληλότητας, 150
- τυχαία δειγματοληψία, 104
- k nearest neighbor (k NN), 158
- (Almost) No Free Lunch, 31
- Bootstrap aggregating, 134
- Cluster-Based Local Outlier Factor, 125
- Computational Intelligence, 3
- Constraints, 147
- Continuous Free Lunch, 29
- Data Mining (DM), 81
- Data mining, 10
- Data preparation, 81
- Data preprocessing, 81
- Direct Methods, 8
- Ensemble noise Filter approach, 86
- Evolutionary algorithms, 9
- Free Lunches, 15
- Indirect Methods, 8
- Irrelevant, 111, 151
- Machine Learning, 10, 121, 147
- Methods with guaranteed accuracy, 8
- Natural Computing, 3
- Newton trajectory method, 42
- No Free Lunch theorems for supervised learning, 33
- No Free Lunch theorem, 15
- Objective function, 147
- Pareto front, 150
- Pareto optimal solutions, 150
- Principal Component Analysis (PCA), 124
- Redundant, 111, 152
- Relevant, 111, 151
- Runge-Kutta δεύτερης τάξης, 60
- Stochastic Outlier Selection, 124
- Subspace Outlier Detection, 124
- Variables, 147
- anomalies, 86, 121
- anomaly, 121
- association rule mining, 160
- attributes, 151
- backward stepwise selection, 112
- bagging technique, 134
- bagging, 134
- classification, 147

- cluster analysis, 159
 - clustering, 8, 147
 - condensation methods, 104
 - constraints, 150
 - covering methods, 8
 - data normalization, 97
 - data preparation, 82
 - data preprocessing, 147
 - dataset shift, 118
 - decision support system, 133
 - decision trees, 155
 - decremental methods, 104
 - density-based methods, 84
 - descriptive model, 147
 - discretization, 98, 154
 - dissimilar values, 121
 - distance-based methods, 84
 - edition methods, 104
 - ensemble methods, 125, 134
 - ensemble method, 121
 - ensemble of noise filtering schemes, 87
 - evolutionary algorithms, 151
 - feature construction/transformation, 113
 - feature construction, 114
 - feature scaling, 97
 - feature selection, 111, 151
 - features, 151
 - filter methods, 104
 - fitness value, 150
 - forward stepwise selection, 112
 - generalized descent methods, 8
 - gradient boosting, 138
 - heuristic methods, 40
 - hybrid methods, 104
 - incremental methods, 104
 - instance selection, 153
 - instance selection, 104
 - leap-frog method, 50
 - lexicographical approach, 148
 - mathematical programming, 159
 - meta-learning, 136
 - missing feature values, 92
 - mixed methods, 104
 - multi-objective evolutionary optimization, 147
 - multi-objective feature selection, 152
 - multi-objective optimization, 10
 - multi-objective, 148
 - multiple criteria optimization, 150
 - noise filters, 84
 - noise reduction, 86
 - noise, 84
 - nonmonotone convergence strategy, 70
 - objective functions, 150
 - outliers, 84, 121
 - outlier, 121
 - population based methods, 40
 - random sampling, 104
 - random search methods, 8
 - sampling, 104
 - single-objective problem, 148
 - single-objective, 9
 - stacking, 133, 135
 - statistics-based methods, 84
 - stratified sampling, 104
 - supervised discretization methods, 98
 - support vector machine, 157
 - uni-objective, 148
 - unsupervised discretization methods, 98
 - weighted approach, 148
 - wrapper methods, 104, 112
- Εξελικτικοί Αλγόριθμοι, 9