Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων
και Μηχανικων Υπολογιστων
Τομεας Συστηματων Μεταδοσης Πληροφοριας και Τεχνολογιας Υλικων

Επιλογή Χαρακτηριστικών και Κατηγοριοποίηση Ιατρικών Δεδομένων από Υπέρυθρη Φασματοσκοπία

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Βασίλειος Χ. Κορφιάτης

Επιβλέπων : Γεώργιος Ματσόπουλος
Επ. Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2010

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Επιλογή Χαρακτηριστικών και Κατηγοριοποίηση Ιατρικών Δεδομένων από Υπέρυθρη Φασματοσκοπία

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Βασίλειος Χ. Κορφιάτης

Επιβλέπων : Γεώργιος Ματσόπουλος
Επ. Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22η Ιουλίου 2010.

Αθήνα, Ιούλιος 2010

.....	.....	.....
Γ.Ματσόπουλος	Ν.Ουζούνογλου	Αθ.Παναγόπουλος
Επ. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Λέκτορας Ε.Μ.Π.

.....

Βασίλειος Χ. Κορφιάτης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Βασίλειος Κορφιάτης, 2010
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Σκοπός της εργασίας αυτής ήταν η ανάπτυξη ενός πλήρους συστήματος κατηγοριοποίησης δεδομένων. Το σύστημα αυτό χρησιμοποιήθηκε για την κατηγοριοποίηση δεδομένων που έχουν προέλθει από εφαρμογή της υπέρυθρης (IR) φασματοσκοπίας. Χωρίζεται στα επιμέρους υποσυστήματα προεπεξεργασίας δεδομένων, επιλογής χαρακτηριστικών και κατηγοριοποίησης.

Για την υλοποίηση κάθε ενός εξ' αυτών χρησιμοποιήθηκε μια πληθώρα μεθόδων διαφορετικής φύσης, με στόχο την εύρεση του συνδυασμού που θα επιτύγχανε τα μεγαλύτερα ποσοστά επιτυχίας. Πιο συγκεκριμένα, στο τμήμα της προεπεξεργασίας έγινε χρήση κανονικοποίησης και παραγώγισης, στο τμήμα επιλογής χαρακτηριστικών έγινε υλοποίηση των αλγορίθμων Sequential Forward Floating Search (SFFS) και Hybrid Genetic Algorithm (HGA), ενώ για την κατηγοριοποίηση χρησιμοποιήθηκαν οι K-means, SVM, Νευρωνικά Δίκτυα, FCM, K-NN και μια επιτροπή που συνδύαζε αποτελέσματα από τους παραπάνω χρησιμοποιώντας βάρη.

Έπειτα από μια εισαγωγή στις έννοιες της IR φασματοσκοπίας και της εφαρμογής της σε ιατρικά δεδομένα, τοποθετήθηκε το θεωρητικό υπόβαθρο πάνω στο οποίο στηρίζεται κάθε μέθοδος και στη συνέχεια έγινε πλήρης αποτίμηση της ικανότητας κάθε αλγορίθμου να ανταπεξέλθει στο ρόλο του, για το συγκεκριμένο σύνολο δεδομένων, δίνοντας έμφαση στον τρόπο που η κάθε παράμετρος του επηρεάζει την απόδοση. Τέλος, αξιολογήθηκαν οι διάφοροι συνδυασμοί των παραπάνω στο πλαίσιο της σύνθεσης του γενικότερου συστήματος, αναφέρθηκαν τα συμπεράσματα που προέκυψαν και με βάση αυτά έγιναν προτάσεις για μελλοντική έρευνα.

Λέξεις-Κλειδιά:

Κατηγοριοποίηση, Επιλογή Χαρακτηριστικών, Ιατρικά Δεδομένα, IR φασματοσκοπία, SFFS, HGA, Kmeans, SVM, Νευρωνικά Δίκτυα, K-NN, Επιτροπή Κατηγοριοποίησης.

Abstract

The goal of this diploma was the creation of a complete classification system. This system was used for classifying IR spectroscopy data. It can be broken down to three different subsystems that are responsible for preprocessing, feature selection and classification of data.

For the implementation of each one of them many different methods, from the currently available, were used, with our main aim being the search of the combination that gave us the highest accuracy. Specifically, for the preprocessing part we selected between normalization and derivative, for the feature selection part between the SFFS method and the HGA and finally for the classification part amongst the Kmeans, SVM, Neural Networks, FCM, K-NN algorithms and we also built a committee that used the previous results with weights, in order to decide.

After a short introduction to the basics of IR spectroscopy and it's applications to medical data, the theoretical basis of all the used methods was established and then the ability of each to perform its part was evaluated accordingly, with much attention being put on the way the different parameters of each method affect the accuracy. Finally the various combinations of subsystem implementations were formed and evaluated, conclusions were drawn and suggestions for further investigation on promising aspects of this scientific area were proposed.

Key Words:

Classification, Feature Selection, Medical Data, IR spectroscopy, SFFS, Kmeans, SVM, Neural Network, K-NN, classification committee.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ. Γεώργιο Ματσόπουλο , Επίκουρο Καθηγητή-Ε.Μ.Π, για τις συμβουλές, τη βοήθεια και την υποστήριξη μου κατά τη διάρκεια της συγγραφής της εργασίας, αλλά και του φοιτητικού βίου μου γενικότερα.

Επίσης τον κ. Παντελή Ασβεστά, Καθηγητή Εφαρμογών του Τ.Ε.Ι. Αθήνας, για την καθοδήγηση, τη βοήθεια και τις καθοριστικές υποδείξεις του, δίχως τις οποίες δε θα ήταν δυνατή η ολοκλήρωση της εργασίας.

Στη συνέχεια θα ήθελα να ευχαριστήσω τους γονείς μου, Χρήστο και Δέσποινα Κορφιάτη, καταρχάς γιατί με έφεραν στη ζωή και έπειτα γιατί δημιούργησαν ένα ασφαλές, υγιές και γεμάτο αγάπη περιβάλλον για την ανάπτυξη μου. Επίσης το μικρό μου αδερφό Γρηγόρη, που με μαθαίνει τόσα πολλά για αυτόν τον κόσμο και ανεχόταν να μοιραζόμαστε το ίδιο δωμάτιο. Ύστερα, τους νονούς μου Παναγιώτη Φάκο και Αφροδίτη Κορφιάτη για την αγάπη, τις συμβουλές και την καθοδήγηση τους σε όλη τη διάρκεια του βίου μου, καθώς και τις δύο μου γιαγιάδες Αθανασία και Αντωνία για την αγάπη τους.

Για το τέλος αφήνω όλους τους λίγους καλούς μου φίλους που μοιράστηκαν, μοιράζονται και ελπίζω να συνεχίσουν να μοιράζονται τις χαρές μου καθώς και τις δυσκολίες μου, και να κάνουν αυτόν τον κόσμο ένα πιο όμορφο μέρος να ζει κανείς με την ξεχωριστή προσωπικότητα τους, τη ζεστασιά και την ειλικρίνεια τους.

Περιεχόμενα

Περίληψη 5
Abstract 6
1.ΕΙΣΑΓΩΓΗ 12
1.1 IR Φασματοσκοπία 13
1.2 Όργανα Μέτρησης 15
1.3 Σύστημα Κατηγοριοποίησης 16
2.ΔΕΔΟΜΕΝΑ ΕΙΣΟΔΟΥ 18
2.1 Προεπεξεργασία 18
Κανονικοποίηση 18
Παράγωγος με Γκαουσιανό Πυρήνα 21
3.ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ 25
Εισαγωγή 25
3.1. Προβλήματα Κατηγοριοποίησης 25
3.2 Σχετικότητα ενός χαρακτηριστικού 28
3.3 Αλγόριθμοι για την επιλογή χαρακτηριστικών 30
3.3.1 Ορισμός επιλογής χαρακτηριστικών 30
3.3.2 Χαρακτηρισμός των FSA 31
3.4 Τελεστές 33
3.4.1 Γενιά διαδόχων 33
3.4.2 Κριτήρια Αξιολόγησης 35
3.5 Πλαίσια που καθορίζουν τη σχέση μεταξύ κατηγοριοποιητή και FSA 39
3.6 Γενικός αλγόριθμος για την επιλογή χαρακτηριστικών 39
Χώρος χαρακτηριστικών του FSA 40
3.7 Περιγραφή των βασικών FSA 43
3.7.1 LVF αλγόριθμος 43
3.7.2 LVI αλγόριθμος 44
3.7.3 Αλγόριθμος Ανακούφισης (Relief) 45
3.7.4 SFG/SBG αλγόριθμοι 46

3.7.5 SFFS αλγόριθμος 46
3.7.6 Αλγόριθμος Συγκέντρωσης /Εξαντλητικός (Focus) 48
3.7.7 B&B αλγόριθμος 48
3.7.8 QBB αλγόριθμος 49
3.8 Αλγόριθμος SFFS για την επιλογή χαρακτηριστικών 50
4. ΕΞΕΛΙΚΤΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ 52
Εισαγωγή 52
Εξομοίωση της φυσικής εξέλιξης 53
Εισαγωγή στους Γενετικούς Αλγορίθμους 55
4.1 Γενετικοί αλγόριθμοι 56
4.1.1 Βασικοί Τελεστές 59
Επιλογή 59
Διασταύρωση 60
Μετάλλαξη 62
Αναπαράσταση Πολλών Μεταβλητών 62
Γράφημα Απόδοσης 65
4.1.2 Θεωρητικό Υπόβαθρο των ΓΑ 65
4.2. Στρατηγικές Εξέλιξης 69
5. ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΚΑΙ ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ 73
Εισαγωγή 73
5.1 Απλός Γενετικός Αλγόριθμος για την επιλογή χαρακτηριστικών 73
Κωδικοποίηση Χρωμοσωμάτων 74
Αρχικός πληθυσμός 74
Εκτίμηση καταλληλότητος, επιλογή, αντικατάσταση και λήξη 75
Διασταύρωση και Μετάλλαξη 76
Παράμετροι 77
5.2 ΥΒΡΙΔΙΚΟΣ ΓΑ 77
6. ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΕΣ 82
Εισαγωγή 82

6.1 Τεχνητά Νευρωνικά Δίκτυα 82
Multilayer Perceptrons 85
Βελτιώσεις Αλγορίθμου Εκπαίδευσης 89
Χρήση ορμής 89
Αναζήτηση Ελαχίστου σε ευθεία 90
Μέθοδος Συζυγούς Κατάβασης Δυναμικού 90
6.2 SUPPORT VECTOR MACHINES 93
Μη διαχωρίσιμα προβλήματα 97
6.3 K-MEANS 101
6.4 Fuzzy C-Means Clustering (FCM) 102
6.5 K-Nearest Neighbors Clustering (K-NN) 104
7. ΠΕΙΡΑΜΑΤΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ 106
Εισαγωγή 106
7.1 Χαρακτηριστικά Συστήματος Υλοποίησης 106
7.2 SFFS 106
7.3 Παράμετροι SFFS 108
Αριθμός αρχικών/ελαχίστων στοιχείων 108
7.4 HGA 111
7.5 Παράμετροι HGA 114
7.5.1 Αριθμός Γενεών 114
7.5.2 Αριθμός χρωμοσωμάτων P 116
7.5.3 Αριθμός Γονιδίων d 118
7.5.4 Συντελεστής Μετάλλαξης p_m 119
7.5.5 Πίεση Επιλογής 120
7.5.6 Ποινή w 122
7.5.7 Αριθμός Σημείων Διασταύρωσης 123
7.5.8 Βάθος Τοπικής Βελτιστοποίησης 124
7.6 Υποσύστημα Κατηγοριοποίησης 127
7.7 Παράμετροι SVM 128

7.8 Νευρωνικά Δίκτυα 130
Κανονικοποιημένα 130
Παραγωγιζόμενα 131
7.9 K-means 133
Κανονικοποιημένα 133
Παραγωγιζόμενα 133
7.10 FCM 134
7.11 K-NN 134
7.12 Επιτροπή 135
8. ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ 137
Βιβλιογραφία 140

1.ΕΙΣΑΓΩΓΗ

Κατά γενική ομολογία, μια ασθένεια είναι μια ανώμαλη κατάσταση ενός οργανισμού, η οποία επηρεάζει αρνητικά τις σωματικές λειτουργίες. Διάφοροι παράγοντες συντείνουν στην εμφάνισή της και χωρίζονται σε ενδογενείς, όπως οι γενετικές ατέλειες και οι τροφικές ελλείψεις και σε εξωγενείς, όπως το παθητικό κάπνισμα. Συνήθως αυτοί οι παράγοντες εμφανίζονται μαζί. Ο ευρύτερα γνωστός τρόπος αντιμετώπισης των ασθενειών είναι μέσω της φαρμακευτικής, αν και υπάρχουν και άλλες μέθοδοι, λιγότερο διαδεδομένες που δεν απαιτούν τη χρήση τεχνητών χημικών ουσιών.

Ο κλάδος της ιατρικής που ασχολείται με τη διάγνωση και τη θεραπεία των ασθενειών ονομάζεται παθολογία. Για να πετύχει το σκοπό της ερευνά τα όργανα, τους ιστούς, τα κύτταρα και τα σωματικά υγρά. Μπορεί να χωριστεί σε ανατομική παθολογία και κλινική παθολογία, γνωστή και ως εργαστηριακή φαρμακευτική. Η διάγνωση ασθενειών από την ανατομική παθολογία περιλαμβάνει γενικά και μακροσκοπική και μικροσκοπική οπτική έρευνα ιστών που έχουν εμπλουτισθεί με ειδικούς δείκτες, ώστε να γίνεται ευκολότερη η οπτικοποίηση συγκεκριμένων πρωτεϊνών και άλλων ουσιών. Ο Rudolf Virchow (1821-1902) θεωρείται ο πατέρας της μικροσκοπικής παθολογίας. Παρόλο που το μικροσκόπιο είχε εφευρεθεί παλαιότερα, αυτός ήταν ο πρώτος που επικεντρώθηκε στη μελέτη των δειγμάτων ύπαρξης ασθένειας που ήταν ορατά μόνο σε κυτταρικό επίπεδο. Ο μαθητής του Virchow, Julius Cohnheim (1839-1884) συνεχίζοντας το έργο του δασκάλου του, συνδύασε ιστολογικές τεχνικές με πειραματικές διαδικασίες και έγινε ο πρώτος πειραματικός παθολόγος. Στη βάση που έθεσε λοιπόν αναπτύχθηκαν μετέπειτα νέες τεχνικές έρευνας όπως η ηλεκτρονιακή μικροσκοπία, η ανοσοϊστοχημεία και η μοριακή βιολογία, βοηθώντας στη διάγνωση των ασθενειών.

Η ανάπτυξη που έχει παρουσιαστεί στις τεχνικές φασματοσκοπίας και στις υπολογιστικές τεχνικές την τελευταία δεκαετία, έχει δώσει ώθηση και στην πρόοδο της IR και της Raman φασματοσκοπίας για εφαρμογές στη βιολογία και τις φυσικές επιστήμες γενικότερα. Αυτές οι τεχνικές παρέχουν ένα πλούτο πληροφοριών από το κυτταρικό και μοριακό επίπεδο στερεών και υγρών δειγμάτων, χωρίς τη χρήση εξωτερικής βοήθειας από χρωματικούς δείκτες και τη ραδιενεργή σήμανση. Ο τρόπος λειτουργίας στηρίζεται στη μελέτη των μοριακών ταλαντώσεων που εξαρτώνται από τη δομή και τη σύνθεση των δειγμάτων. Οι ασθένειες και οι λοιπές παθολογικές ανωμαλίες προκαλούν μεταβολές στην χημική αλλά και στη δομική σύσταση των κυττάρων και κατ’ επέκταση μεταβάλλουν το φάσμα των ταλαντώσεων. Η μεταβολή αυτή καταγράφεται και μπορεί να χρησιμοποιηθεί για την εξαγωγή συμπερασμάτων και τη διάγνωση. Ένα άλλο χαρακτηριστικό που κάνει αυτές τις μεθόδους ελκυστικές είναι πως

και οι δύο μπορούν πλήρως να αυτοματοποιηθούν. Πρόσφατες μελέτες περιγράφουν αρκετές χρήσεις της IR και της Raman φασματοσκοπίας όπως για βιοχημικές εφαρμογές για τη διάγνωση ιστών, χημική εικονογράφηση του βιολογικού ιστού με σύχνοτρο IR ακτινοβολίας, κλινικές και διαγνωστικές εφαρμογές, καταγραφή του μεταβολισμού και μη-παρεμβατική βιοχημική ανάλυση επιμέρους κυττάρων. Αυτές μπορούν να εφαρμοστούν πρακτικά για τη διάγνωση προβλημάτων στους λεμφαδένες, σε επιμέρους ανθρώπινα κύτταρα, στον οισοφάγο και την ουροδόχο κύστη, στη διάγνωση της σπογγώδους εγκεφαλοπάθειας, του καρκίνου του λαιμού και του κεφαλιού, στην υψηλής απόδοσης ιστοπαθολογία, τη νευροπαθολογία, στη μελέτη των μικροοργανισμών κ.α.

Στις πιο συχνές ασθένειες που έχει γίνει εφαρμογή των φασματοσκοπικών τεχνικών Raman και IR με στόχο τη διάγνωσή τους είναι οι επιθηλιακοί όγκοι, οι εγκεφαλικοί όγκοι, οι πρωτεϊνικές ανωμαλίες, οι ανωμαλίες στα κόκκαλα, η αθηροσκλήρωση, οι πέτρες στα νεφρά, οι όγκοι του δέρματος, ο διαβήτης και η οστεοαρθρίτιδα.

1.1 IR Φασματοσκοπία

Τα δεδομένα εισόδου συλλέχθηκαν με τη μέθοδο της Υπέρυθρης Φασματοσκοπίας (Infrared Spectroscopy). Για το λόγο αυτό αφήνουμε στην άκρη τη φασματοσκοπία Raman και επικεντρωνόμαστε στην περιγραφή της IR φασματοσκοπίας.

Η υπέρυθρη φασματοσκοπία είναι το υποσύνολο της φασματοσκοπίας που ασχολείται με την υπέρυθρη περιοχή του ηλεκτρομαγνητικού φάσματος, δηλαδή το κομμάτι εκείνο μεταξύ του ορατού και της μικροκυματικής περιοχής. Η υπέρυθρη ακτινοβολία αυτή απορροφάται από τις οργανικές ενώσεις και μετατρέπεται σε ενέργεια που διοχετεύεται στις μοριακές ταλαντώσεις.

Η έννοια της φασματοσκοπίας καλύπτει ένα εύρος τεχνικών, με την πιο συνηθισμένη να είναι αυτή που λαμβάνει τη μορφή της φασματοσκοπίας απορρόφησης. Όπως και με όλες τις τεχνικές φασματοσκοπίας, μπορεί να χρησιμοποιηθεί για την ταυτοποίηση χημικών ενώσεων, την εύρεση της καθαρότητάς τους και τη διερεύνηση της σύνθεσης διαφόρων δειγμάτων. Ένα τυπικό όργανο που χρησιμοποιεί αυτήν την τεχνική είναι το υπέρυθρο φασματοφωτόμετρο.

Η υπέρυθρη φασματοσκοπία εκμεταλλεύεται το γεγονός πως τα μόρια απορροφούν συγκεκριμένες συχνότητες οι οποίες είναι χαρακτηριστικές της δομής τους. Αυτές οι απορροφήσεις είναι συντονισμένες συχνότητες, π.χ. η συχνότητα της απορροφούμενης ακτινοβολίας ταυτίζεται με τη συχνότητα του δεσμού ή της ομάδας που ταλαντώνεται. Οι ενέργειες καθορίζονται από το σχήμα των μοριακών πιθανών ενεργειακών επιφανειών, τις μάζες των ατόμων, και το σχετικό δεσμό ταλάντωσης. Όπως και αν έχει πάντως, η υπέρυθρη ακτινοβολία απορροφάται από οργανικά μόρια και μετατρέπεται σε ενέργεια μοριακής

ταλάντωσης, είτε μέσω της αύξησης του πλάτους τους είτε μέσω της αλλαγής της κατεύθυνσής τους (επέκταση και στροφή αντίστοιχα). Διαφορετικά είδη ενώσεων, και συνεπώς διαφορετικές λειτουργικές μονάδες, απορροφούν διαφορετικά μήκη κύματος της υπέρυθρης ακτινοβολίας. Στην IR φασματοσκοπία, ένα οργανικό μόριο εκτίθεται στην υπέρυθρη ακτινοβολία. Όταν η ακτινοβολούσα ενέργεια ταιριάξει με την ενέργεια μιας συγκεκριμένης μοριακής ταλάντωσης προκύπτει η απορρόφηση.

Η συχνότητα, *ν*(nu), είναι ο αριθμός των πλήρων κύκλων κύματος που περνάνε από ένα σημείο σε χρόνο ενός δευτερολέπτου. Μετράται σε Hz, όπου 1Hz=1κύκλος/sec. Το μήκος κύματος, *λ*, είναι το μήκος ενός πλήρους κύκλου. Μετράται συνήθως σε cm. Το μήκος κύματος και η συχνότητα συνδέονται μεταξύ τους ως εξής:

$$\nu = \frac{c}{\lambda} \; \kappa\alpha\iota \; \lambda = \frac{c}{\nu}, \qquad \text{όπου } c = 3 \times 10^{10} cm/sec$$

Η ενέργεια σχετίζεται με το μήκος κύματος και τη συχνότητα μέσω των παρακάτω σχέσεων:

$$E = h\nu = \frac{hc}{\lambda}, \qquad \text{όπου } h \text{ η σταθερά του } Planck, 6{,}6 \times 10^{-34} joules/sec$$

Το υπέρυθρο τμήμα του ηλεκτρομαγνητικού φάσματος συνήθως διαιρείται σε τρείς περιοχές. τη κοντινή, τη μέση και την μακρά υπέρυθρη, οι οποίες έχουν λάβει την ονομασία τους λόγω της σχέσης τους με το ορατό φάσμα. Η μακρά υπέρυθρη περιοχή η οποία βρίσκεται στα 400-10cm^{-1} (1000-30μm) και κοντά στη μικροκυματική περιοχή, έχει χαμηλή ενέργεια και μπορεί να χρησιμοποιηθεί για κυκλική φασματοσκοπία. Η μεσαία υπέρυθρη η οποία βρίσκεται στα 4000-400 cm^{-1} (30-2,5 μm) μπορεί να χρησιμοποιηθεί για τη μελέτη των θεμελιωδών δονήσεων και τη σχετική κυκλική-ταλαντωτική δομή. Η υψηλότερης ενέργειας κοντινές υπέρυθρες οι οποίες κυμαίνονται μεταξύ 14000-4000 cm^{-1} (2,5-0,8 μm) μπορούν να παράγουν υπερτονικές ή αρμονικές ταλαντώσεις . Τα ονόματα και η κατηγοριοποιήσεις αυτών των υποπεριοχών είναι απλές παραδοχές και δε βασίζονται ούτε σε αυστηρούς διαχωρισμούς ούτε σε ακριβείς μοριακές ή ηλεκτρομαγνητικές ιδιότητες. Η περιοχή που χρησιμοποιείται περισσότερο στις χημικές εφαρμογές είναι η μέση. Η ίδια περιοχή χρησιμοποιήθηκε και κατά την φασματογράφηση των δεδομένων της συγκεκριμένης εργασίας.

Η μονάδα μέτρησης που προτιμάται είναι το $\bar{\nu}$ (nu μπάρα ή συνηθέστερα *κυματαριθμός*). Ο κυματαριθμός είναι το αντίστροφο του μήκους κύματος σε cm:

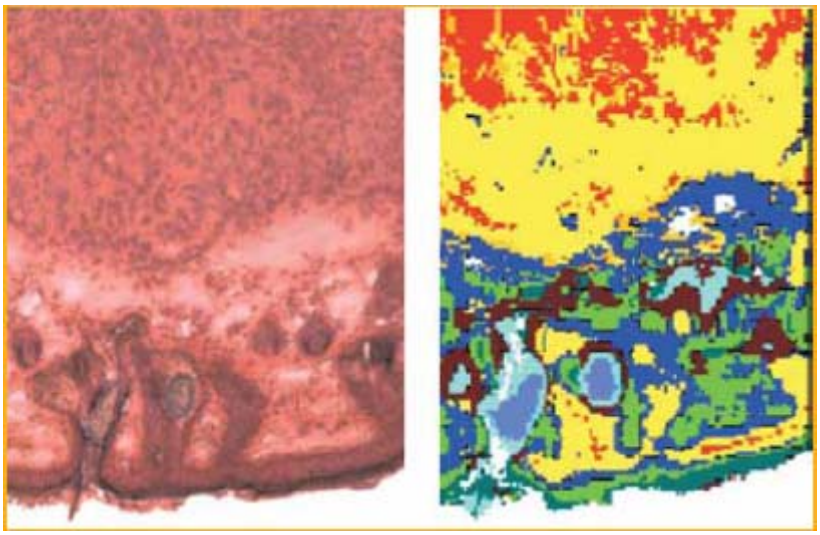
$$\bar{\nu} = \frac{1}{\lambda}$$

και συνεπώς για την ενέργεια έχουμε:

$$E = hc\bar{\nu}$$

Από τα παραπάνω είναι προφανές πως μια αύξηση στον κυματαριθμό συνεπάγεται αύξηση της ενέργειας.

Οι κυματαριθμοί (μερικές φορές αναφέρονται και ως *συχνότητες*) στους οποίους ένα οργανικό μόριο απορροφά ακτινοβολία δίνουν πληροφορίες για τις παρούσες λειτουργικές ομάδες στο μόριο. Συγκεκριμένες ομάδες ατόμων απορροφούν ενέργεια με αποτέλεσμα να προκαλούν αύξηση σε μπάντες σε περίπου τις ίδιες συχνότητες. Ο χημικός καλείται να αναλύσει το φάσμα με τη βοήθεια έτοιμων πινάκων που συσχετίζουν τις συχνότητες με τις λειτουργικές ομάδες και να καταλήξει μέσω αυτού στη σύνθεση των χημικών παραγώγων που μελετά.



Σχήμα 1.1: Αριστερά: Εικόνα καρκινοπαθούς ιστού με χρήση δεικτών
Δεξιά: Εικόνα του ίδιου ιστού με χρήση IR φασματοσκοπίας. Με κόκκινο και κίτρινο φαίνονται οι καρκινοπαθείς ιστοί, ενώ με τα υπόλοιπα οι φυσιολογικοί. Παρατηρούμε την αυξημένη ευκρίνεια που παρέχεται με τη μέθοδο αυτή δίχως επιβάρυνση του οργανισμού με χημικούς δείκτες

1.2 Όργανα Μέτρησης

Το όργανο μέτρησης του φάσματος, ονομάζεται φασματοσκόπιο απορρόφησης (*επίσης γνωστό και ως φασματογράφος*). Ο συνδυασμός της υψηλής περιεκτικότητας σε νερό των κυττάρων , των ιστών και των σωματικών υγρών με την υψηλή απορρόφηση των μεσαίων-IR ακτινοβολιών από το νερό, αναγκάζει το εύρος διείσδυσης των ακτινών να είναι περιορισμένο σε μερικά μικρόμετρα. Παρόλο που τα κύτταρα είναι αρκετά λεπτά και τα υγρά τους μπορούν να γεμίσουν κοιλότητες με μήκος 5-10μm, τα ιστολογικά δείγματα για τις περισσότερες έρευνες με χρήση ακτινοβολιών μεσαίων-IR κόβονται σε περιοχές με πάχος 5-20μm, και μεταφέρονται σε υποστρώματα διαφανή σε αυτό το είδος ακτινοβολίας όπως το φθοριούχο ασβέστιο και το φθοριούχο βάριο τα οποία ξηραίνονται και απεικονίζουν τα αποτελέσματα.

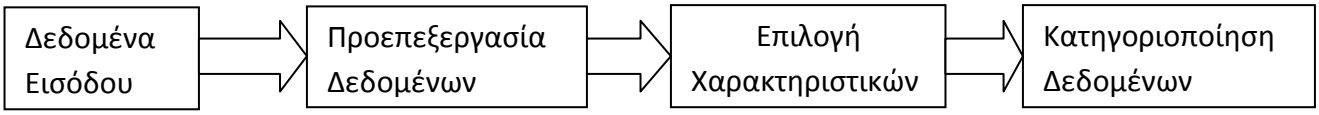
Μια εναλλακτική μέθοδος απεικόνισης είναι αυτή που ονομάζεται ανακλαστική-απορροφητική μέθοδος και χρησιμοποιεί υψηλής ανακλαστικότητας γυάλινες διαφάνειες με μεταλλική επένδυση, αποτελώντας μια πιο οικονομική λύση. Η μέθοδος Attenuated Total Reflection (ATR), προσφέρει μια ακόμη μέθοδο για την απεικόνιση φασμάτων από αδιαφανή βιοϊατρικά δείγματα, με χρήση IR φασματοσκοπίας.

Τα φασματοσκόπια μεσαίων-IR χρησιμοποιούν την αρχή του ιντερφερομετρικού μετασχηματισμού Fourier (FT), η οποία έχει τα πλεονεκτήματα του πολλαπλού ρυθμού αποστολής και της ακρίβειας του κυματικού αριθμού. Το βασικό πακέτο εξοπλισμού περιλαμβάνει μια ευρυζωνική πηγή ακτινοβολίας, ένα ιντερφερόμετρο, ένα θάλαμο δειγμάτων το οποίο μπορεί να είναι και μικροσκόπιο και ένα γρήγορο ανιχνευτή. Πολλές εφαρμογές της μικροσκοπικής παθολογίας απαιτούν τη λήψη εικόνων. Τα φασματοσκοπικά δεδομένα μπορούν να χρησιμοποιηθούν σε συνδυασμό με την πληροφορία από τα FTIR φασματοσκόπια με μονοκάναλους ανιχνευτές, μέσω του περιορισμού της ακτινοβολίας σε ένα οριοθετημένο επίπεδο και της σάρωσης πάνω από την περιοχή ενδιαφέροντος ,με ένα αυτοματοποιημένο στάδιο μετάφρασης. Σύμφωνα με το νόμο του Abbe, d=0,612 λ/NA η ανάλυση d στη φασματοσκοπική απεικόνιση περιορίζεται από τη περίθλαση (λ το μήκος κύματος, NA το αριθμητικό σφάλμα του μικροσκοπίου). Στην πράξη η πλευρική ανάλυση περιορίζεται συνήθως από την ακτινοβολία του φωτός παρά από το όριο περίθλασης. Για να βελτιστοποιηθεί η ευαισθησία υπό τους δύο αυτούς περιορισμούς χρησιμοποιείται η IR ακτινοβολία υψηλής λαμπρότητας από πηγές συχνοτρονίου, αντί για IR ακτινοβολία από πηγές θερμικού φωτισμού όπως γίνεται στα κλασσικά FTIR φασματοσκόπια. Τα FTIR φασματοσκόπια με πολυκάναλους ανιχνευτές, γνωστά και ως FPA, προσφέρουν μια άλλη μέθοδο συλλογής FTIR εικόνων η οποία προσφέρει υψηλές ταχύτητες, καθώς μπορεί να συλλέξει μια ολόκληρη εικόνα στον ίδιο χρόνο που ένα συμβατικό φασματοσκόπιο συλλέγει ένα φάσμα.

1.3 Σύστημα Κατηγοριοποίησης

Στην εργασία που ακολουθεί, σκοπός είναι η δημιουργία ενός πλήρους συστήματος κατηγοριοποίησης συγκεκριμένων ιατρικών δεδομένων εισόδου που έχουν προέλθει από IR φασματοσκοπία, με σκοπό το διαχωρισμό τους σε δύο κατηγορίες. Κύριος στόχος είναι η επίτευξη μιας ικανοποιητικά μεγάλης απόδοσης του συστήματος ή διαφορετικά την επίτευξη ελαχίστου σφάλματος κατηγοριοποίησης, χρησιμοποιώντας όσο το δυνατόν λιγότερους πόρους (επεξεργαστή, μνήμης κτλ.) του υπερκείμενου υπολογιστικού συστήματος, εντός του οποίου υλοποιείται.

Το σύστημα λοιπόν έχει την εξής γενική μορφή:



Τα παραπάνω μπορούν να αναλυθούν στις εξής συστατικά:

Προεπεξεργασία Δεδομένων:

- Κανονικοποίηση
- Παράγωγος με Γκαουσιανό Πυρήνα

Επιλογή Χαρακτηριστικών:

- SFFS
- HGA

Κατηγοριοποίηση Δεδομένων:

- Νευρωνικό Δίκτυο
- SVM
- K-means clustering
- Επιτροπή που συνδυάζει τα αποτελέσματα των παραπάνω με συντελεστές βάρους

Αρχικά παρουσιάζονται πληροφορίες σχετικά με τα δεδομένα εισόδου, τη φύση τους, τον τρόπο συλλογής τους και τα γενικά χαρακτηριστικά τους. Στη συνέχεια, δίνεται η θεωρητική περιγραφή των αλγορίθμων επιλογής χαρακτηριστικών. Παρουσιάζονται οι διάφορες κατηγορίες, γίνεται η θεωρητική περιγραφή τους και στη συνέχεια ακολουθεί η πιο λεπτομερής ανάλυση των αλγορίθμων που χρησιμοποιήθηκαν στην εργασία. Παρόμοια πορεία χαράσσεται και για τους κατηγοριοποιητές, οι οποίοι μπορούμε να πούμε πως αποτελούν το δεύτερο σημαντικό σκέλος του συστήματος. Ύστερα, με βάση το θεωρητικό υπόβαθρο που δόθηκε παραπάνω, γίνεται η παρουσίαση των εξειδικευμένων χαρακτηριστικών των συστημάτων της εργασίας και των τυχόν διαφοροποιήσεων από τις τυπικές θεωρήσεις, που έπρεπε να γίνουν ώστε να είναι αυτή υλοποιήσιμη. Τελικά λαμβάνει χώρα η αποτίμηση των πειραματικών αποτελεσμάτων, απόρροια αρκετών τρεξιμάτων και πειραματισμών με τις διαθέσιμες παραμέτρους, με απώτερο σκοπό να εξαχθούν τα συμπεράσματα της έρευνας γύρω από την αποτελεσματικότητα των διαφόρων τεχνικών στην κατηγοριοποίηση των ιατρικών δεδομένων εισόδου.

2.ΔΕΔΟΜΕΝΑ ΕΙΣΟΔΟΥ

Τα δεδομένα εισόδου προήλθαν από τμήμα ενδοθηλίου (εσωτερικός χιτώνας) καρωτίδων (αρτηριών) ύστερα από εφαρμογή της IR φασματοσκοπίας. Αποτελούνταν από ένα πίνακα 2 διαστάσεων, του οποίου οι γραμμές αποτελούσαν τα πρότυπα και οι στήλες τα χαρακτηριστικά (features) του κάθε προτύπου. Τα χαρακτηριστικά ήταν οι κυματαριθμοί των μετρήσεων, δηλαδή η απόκριση του προτύπου σε διαφορετικές συχνότητες. Στη συγκεκριμένη περίπτωση, μιας και χρησιμοποιήσαμε μεσαίες υπέρυθρες ακτίνες, το εύρος του μήκους κύματος ήταν 30-2,5 μm (κυματαριθμός από 400 έως4000 cm^{-1}) και μετρούνταν κατά τα γνωστά η απορρόφηση που υφίσταντο οι διάφορες ακτινοβολίες. Σκοπός μας είναι να διατηρήσουμε τις συχνότητες εκείνες (αντίστροφο του μήκους κύματος), στις οποίες οι διαφορές μεταξύ τους υποδεικνύουν και ταξινόμηση του προτύπου σε διαφορετική κατηγορία, δηλαδή τις συχνότητες που δίνουν στο σύστημα τη μέγιστη δυνατή ικανότητα διαχωρισμού των προτύπων σε κατηγορίες. Μεταφράζοντας το πρόβλημα σε πιο μαθηματική γλώσσα, σκοπός μας είναι η μείωση του αρχικού αριθμού των στηλών, διατηρώντας μόνο αυτές που συνεισφέρουν σημαντικά στη πληροφόρηση για την κατηγορία που ανήκει το κάθε πρότυπο. Ο αρχικός αριθμός των στηλών είναι 468. Όμως καταφέρνουμε να διατηρήσουμε ελάχιστες εξ' αυτών (χοντρικά 2-10 αναλόγως της αρχιτεκτονικής του συστήματος και της εκάστοτε επανάληψης). Αντίστοιχα, ο αρχικός αριθμός των προτύπων είναι 76. Όμως από αυτά χρησιμοποιούμε μόνο τα 68, καθώς για τα υπόλοιπα δεν γνωρίζουμε την κατηγορία που ανήκουν ώστε να μπορούμε να εκπαιδεύσουμε σωστά το σύστημα. Ας μην ξεχνάμε πως η μάθηση είναι επιβλεπόμενη, συνεπώς θα πρέπει να γνωρίζουμε από πριν που ανήκει το κάθε πρότυπο. Βέβαια θα μπορούσαμε μετά την ολοκλήρωση της εκπαίδευσης να τα δοκιμάσουμε στο σύστημα και να μας απαντήσει αυτό.

2.1 Προεπεξεργασία

Κανονικοποίηση

Τα δεδομένα έπρεπε να υποστούν κάποιας μορφής προεπεξεργασία καθώς οι τιμές τους κυμαίνονταν σε διαφορετικά δυναμικά εύρη και δεν υπήρχε η απαραίτητη ομοιομορφία που θα συντελέσει στην ευκολότερη από κάθε πλευρά, επεξεργασία τους. Πιο συγκεκριμένα, χαρακτηριστικά με μεγάλες τιμές μπορεί να έχουν μεγαλύτερη επίδραση στη συνάρτηση κόστους απ' ότι χαρακτηριστικά με μικρές τιμές, παρόλο που *αυτό δεν αντικατοπτρίζει απαραίτητα τη σημασία τους στη σχεδίαση του κατηγοριοποιητή*. Για το λόγο αυτό τα δεδομένα υπέστησαν μια τυπική κανονικοποίηση διαιρώντας όλα τα στοιχεία με τη μέγιστη εμφανιζόμενη τιμή, με αποτέλεσμα οι τιμές τους να κυμαίνονται τελικά στο διάστημα [0,1].

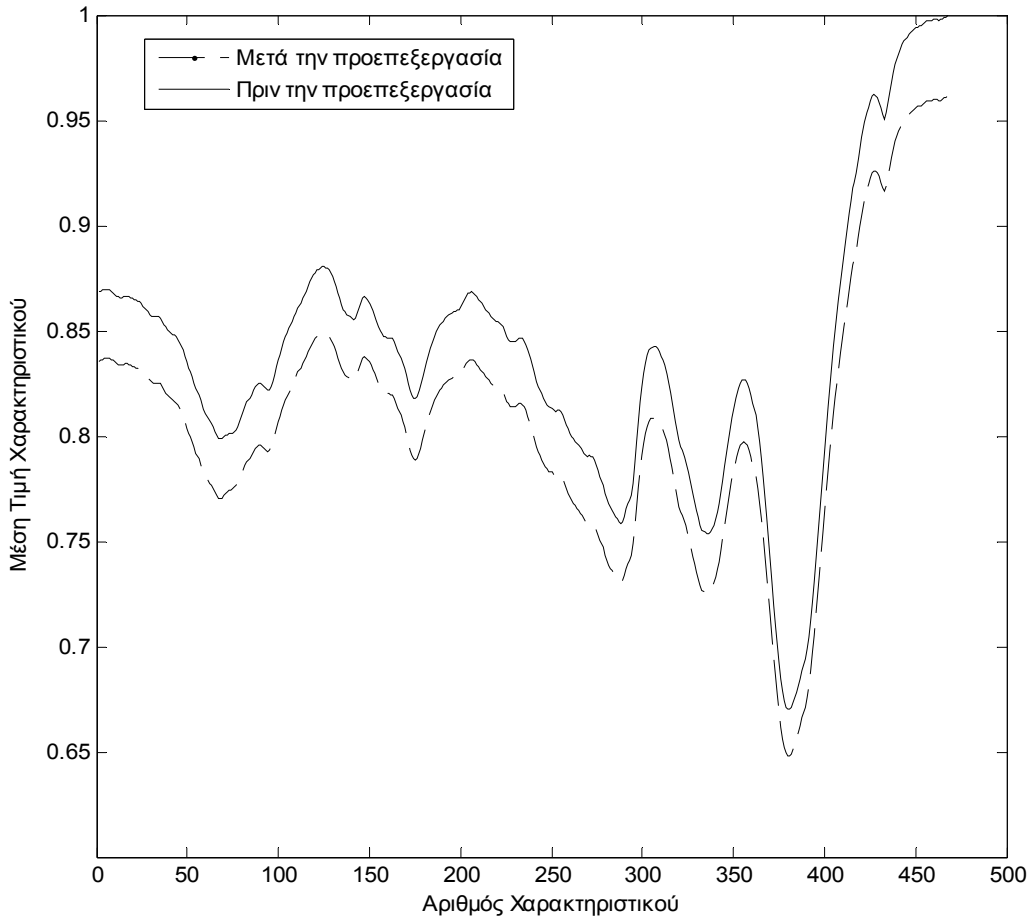
Στο σχήμα 2.1 φαίνεται το αποτέλεσμα της. Όμως πέραν αυτής υπάρχουν και διάφορα άλλα είδη κανονικοποίησης.

Μια προφανής τεχνική κανονικοποίησης είναι αυτή που χρησιμοποιεί τις εκτιμήσεις για το μέσο όρο και τη μεταβλητότητα. Για *N* διαθέσιμα δεδομένα του *k-στού* χαρακτηριστικού έχουμε:

$$\overline{x_k} = \frac{1}{N}\sum_{i=1}^{N} x_{ik}\,, \qquad k = 1,2,\dots,l$$

$$\sigma_k^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{ik} - \bar{x}_k)^2$$

$$\hat{x}_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k}$$



Σχήμα 2.1: Επίδραση κανονικοποίησης σε ένα από τα πρότυπα εισόδου. Παρατηρούμε την ομοιόμορφη μείωση του πλάτους του δίχως μεταβολή του σχήματός του

Με απλά λόγια, όλα τα κανονικοποιημένα χαρακτηριστικά που προκύπτουν θα έχουν τώρα μηδενικό μέσο όρο και μοναδιαία μεταβλητότητα. Αυτή είναι προφανώς μια γραμμική μέθοδος. Άλλες γραμμικές τεχνικές περιορίζουν τις τιμές των χαρακτηριστικών στο διάστημα [0,1] ή [-1,1] με κατάλληλη κλιμακοποίηση. Πέρα από τις γραμμικές μεθόδους, μη-γραμμικές μέθοδοι μπορούν επίσης να χρησιμοποιηθούν σε περιπτώσεις κατά τις οποίες τα δεδομένα δεν είναι ομοιόμορφα κατανεμημένα γύρω από το μέσο όρο. Σε αυτές τις περιπτώσεις μετασχηματισμοί βασισμένοι σε μη-γραμμικές (π.χ. λογαριθμικές ή σιγμοειδείς) συναρτήσεις μπορούν να χρησιμοποιηθούν για να τοποθετήσουν τα δεδομένα εντός συγκεκριμένων διαστημάτων. Το γνωστό softmax scaling είναι ένας δημοφιλής υποψήφιος. Αποτελείται από δύο βήματα:

$$y = \frac{x_{ik} - \bar{x}_k}{r\sigma_k},$$

$$\hat{x}_{ik} = \frac{1}{1 + \exp(-y)}$$

Αυτή βασικά είναι μια συνάρτηση που περιορίζει τα δεδομένα μεταξύ [0,1]. Χρησιμποιώντας μια σειρά προσέγγισης της επέκτασης, δεν είναι δύσκολο να δει κανείς ότι για μικρές τιμές του *y* αυτή είναι μια προσεγγιστικά γραμμική συνάρτηση σε σχέση με το x_{ik}. Το εύρος των τιμών του x_{ik} που αντιστοιχεί στο γραμμικό τμήμα εξαρτάται από την τυπική απόκλιση και τον παράγοντα *r*, που είναι καθορισμένος από τον χρήστη. Οι τιμές που απέχουν πολύ από το μέσο όρο διώχνονται εκθετικά.

Στην πράξη, είναι πιθανόν ο αριθμός των διαθέσιμων δεδομένων να μην είναι ίδιος για όλα τα χαρακτηριστικά. Εάν ο αριθμός των δεδομένων εκπαίδευσης είναι αρκετά μεγάλος, μπορούμε να διώξουμε μερικά από αυτά και να κρατήσουμε ένα μικρότερο αριθμό, οποίος θα είναι ίδιος για όλα τα χαρακτηριστικά, με σκοπό να μορφώσουμε τα διανύσματα χαρακτηριστικών. Παρόλα αυτά, στις περισσότερες περιπτώσεις είναι πολυτέλεια να πετάμε τα διαθέσιμα δεδομένα. Έτσι, τα χαμένα δεδομένα προβλέπονται ευριστικά. Μια προφανής σκέψη είναι να αντικαταστήσουμε τις χαμένες τιμές με τον αντίστοιχο μέσο όρο, όπως αυτός προκύπτει από τα υπόλοιπα διαθέσιμα χαρακτηριστικά. Επίσης μπορούν να χρησιμοποιηθούν και πιο περίτεχνες διαδικασίες που λαμβάνουν υπ' όψιν την κατανομή των δεδομένων. Περισσότερα μπορούν να αναζητηθούν στη σχετική βιβλιογραφία.

Παράγωγος με Γκαουσιανό Πυρήνα

Επιπλέον της κανονικοποίησης και πάνω σε αυτή υπολογίστηκε και η διακριτή παράγωγος των προτύπων. Έτσι υπολογίσαμε την παράγωγο για κάθε γραμμή του πίνακα (για κάθε πρότυπο με άλλα λόγια). Επειδή τα δεδομένα είναι διακριτά, χρησιμοποιήσαμε τον πυρήνα Gauss. Επίσης, λόγω της φύσης της Matlab η οποία δεν κάνει indexing στους πίνακες για μη θετικές τιμές, αναγκαστήκαμε να κάνουμε padding στα πρώτα και στα τελευταία στοιχεία του πίνακα. Αυτό σημαίνει πως αντιγράψαμε όσες φορές χρειάζονταν το πρώτο στοιχείο και τα αντίγραφα τα τοποθετήσαμε στην αρχή του πίνακα και το αντίστοιχο κάναμε για το τελευταίο στοιχείο.

Ο Γκαουσιανός πυρήνας πήρε το όνομά του από το διάσημο μαθηματικό Carl Friedrich Gauss (1777-1855), ο οποίος και τον ανακάλυψε. Μπορεί να οριστεί σε μία, δύο και κατ' επέκταση N διαστάσεις. Ακολουθούν οι τύποι που τον περιγράφουν:

$$G_{1D}(x,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}},$$

$$G_{2D}(x,y,\sigma) = \frac{1}{2\pi\,\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

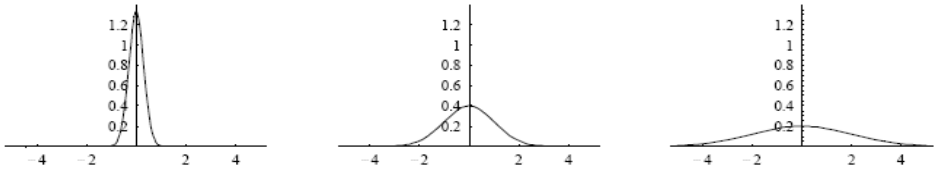
$$G_{ND}(\vec{x},\sigma) = \frac{1}{(\sqrt{2\pi}\,\sigma)^N} e^{-\frac{\overrightarrow{|x|}^2}{2\sigma^2}}$$

Το *σ* καθορίζει το πλάτος του Γκαουσιανού πυρήνα. Στη στατιστική, όταν υπολογίζουμε τη Γκαουσιανή συνάρτηση πυκνότητας πιθανότητας, αυτό ονομάζεται *τυπική απόκλιση*, και το τετράγωνό της, σ^2, ονομάζεται *μεταβλητότητα*. Επίσης το *σ* ονομάζεται και κλίμακα. Γενικά για το *σ* ισχύει $\sigma>0$.

Ο όρος $\frac{1}{\sqrt{2\pi}\,\sigma}$ στην αρχή του μονοδιάστατου πυρήνα είναι η σταθερά κανονικοποίησης. Προέρχεται από το γεγονός ότι το ολοκλήρωμα της εκθετικής συνάρτησης δεν είναι μονάδα:

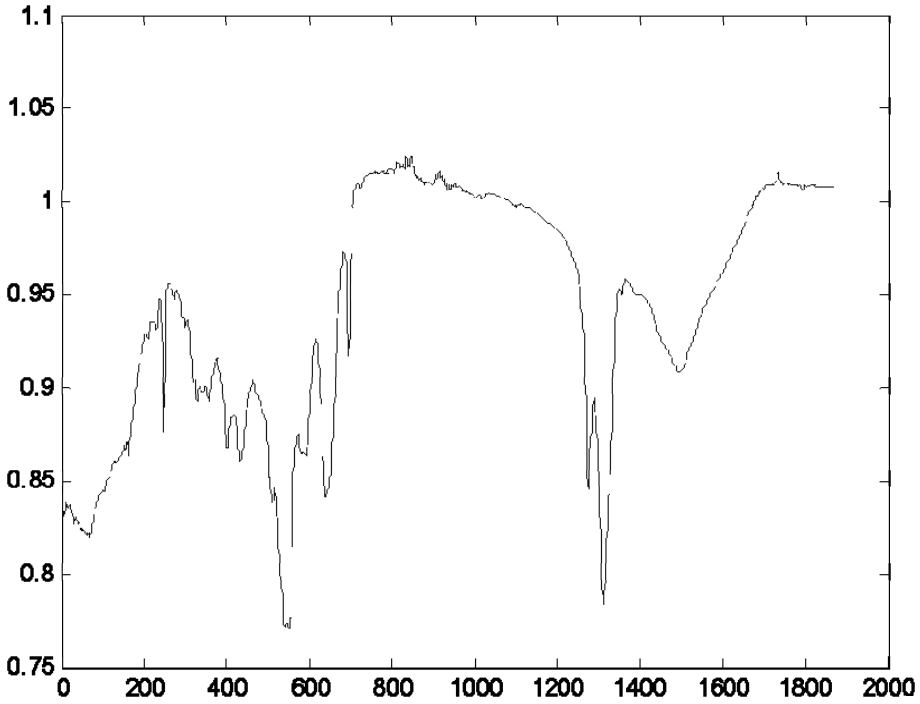
$$\int_{-\infty}^{\infty} e^{\frac{-x^2}{2\sigma^2}}\,dx = \sqrt{2\pi}\,\sigma$$

Με χρήση της σταθεράς κανονικοποίησης αυτός ο Γκαουσιανός πυρήνας είναι ένας κανονικοποιημένος πυρήνας, δηλαδή το ολοκλήρωμά του πάνω στο πλήρες εύρος του είναι μονάδα για οποιοδήποτε *σ*. Αυτό σημαίνει πως αύξηση του *σ* του πυρήνα μειώνει το πλάτος.

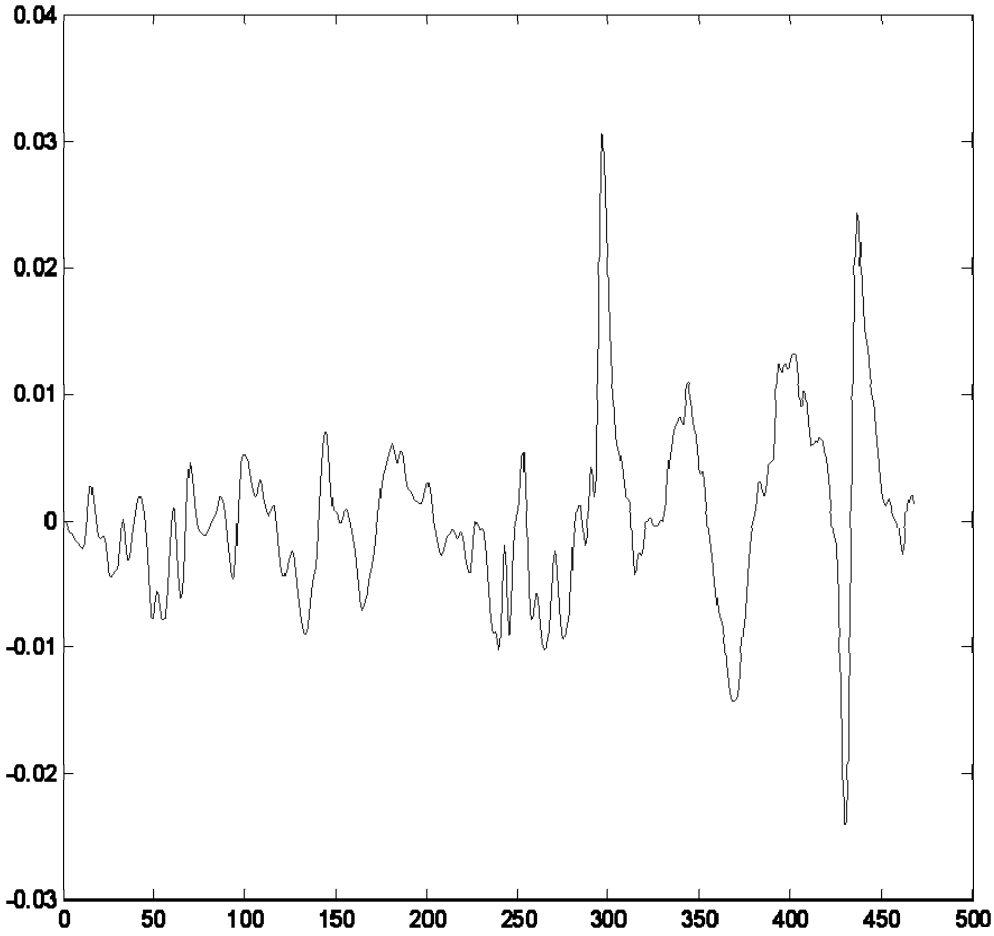


Σχήμα 2.2: Μορφή Γκαουσιανού πυρήνα για αυξανόμενο σ. Παρατηρούμε τη μείωση του πλάτους

Τελικά, για την προεπεξεργασία δώσαμε δύο επιλογές· είτε απλή κανονικοποίηση, είτε κανονικοποίηση και παραγώγιση μαζί. Όπως θα δούμε η επιλογή μας αυτή έχει σημασία για τη συνέχεια.



Σχήμα 2.3: Το φάσμα του πρώτου προτύπου όπου φαίνονται και οι περιοχές κυματαριθμών που παρουσιάζουν υψηλή απορρόφηση



Σχήμα 2.4: Το φάσμα του πρώτου προτύπου ύστερα από την παραγώγιση

3.ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Εισαγωγή

Η επιλογή των χαρακτηριστικών είναι το τμήμα του συστήματος με τη μεγαλύτερη ευαισθησία στο είδος των δεδομένων. Ο τύπος του αλγορίθμου που θα χρησιμοποιηθεί καθώς και οι παράμετροι του θα επηρεάσουν σε μεγάλο βαθμό το τελικό αποτέλεσμα, καθώς σε αυτό το στάδιο επιλέγονται οι είσοδοι του κατηγοριοποιητή. Άρα, όσο κατάλληλα κατασκευασμένος και αν είναι ο τελευταίος, αν δε του δοθεί η έγκυρη πληροφόρηση που θα προέλθει από αυτό το στάδιο, δεν πρόκειται να αποδώσει σωστά.

3.1. Προβλήματα Κατηγοριοποίησης

Ένα άμεσο ερώτημα που προκύπτει είναι για πιο λόγο είναι χρήσιμη και απαραίτητη η διαδικασία επιλογής χαρακτηριστικών για ένα πρόβλημα κατηγοριοποίησης, ενώ απλά μπορούμε να του δώσουμε όλα τα διαθέσιμα χαρακτηριστικά και να αφήσουμε το σύστημα να επιλέξει. Η απάντηση γίνεται σαφής με την παρουσίαση των προβλημάτων που εμφανίζονται στη διαδικασία κατηγοριοποίησης και σχετίζονται φυσικά με τον αριθμό και την φύση των χαρακτηριστικών εισόδου.

Κατά τη διαδικασία της κατηγοριοποίησης αντιμετωπίζουμε μια σειρά από ζητήματα που σχετίζονται με τις διάφορες φάσεις της και επηρεάζουν την ακρίβεια και την επιτυχία της εκάστοτε επιλεγμένης μεθόδου, ανεξαρτήτως ποια είναι αυτή.

Ένα από αυτά τα ζητήματα είναι η «κατάρα της διαστασιμότητας». Αυτό το φαινόμενο εμφανίζεται όταν ο αριθμός των χαρακτηριστικών που είναι διαθέσιμα στο σχεδιαστή του κατηγοριοποιητή είναι πολύ μεγάλος (δηλαδή σχεδόν πάντα στα πραγματικά προβλήματα). Υπάρχουν περισσότεροι του ενός λόγοι για να μειώσουμε τον αριθμό των χαρακτηριστικών σε ένα ικανοποιητικό ελάχιστο. Η υπολογιστική πολυπλοκότητα είναι το προφανές. Ένας άλλος λόγος είναι ότι παρόλο που δύο χαρακτηριστικά μπορούν να έχουν καλή πληροφορία κατηγοριοποίησης όταν αντιμετωπίζονται ξεχωριστά, όταν συνδυαστούν μπορούν να δίνουν λίγα οφέλη, λόγω της υψηλής αμοιβαίας πληροφορίας που μοιράζεται μεταξύ τους. Έτσι αν επιλεγούν και τα δύο, η πολυπλοκότητα του συστήματος αυξάνεται, χωρίς πολύ ουσιαστικό κέρδος.

Επιπλέον αυτών, ο αριθμός των χαρακτηριστικών επηρεάζει και τις ιδιότητες γενίκευσης του κατηγοριοποιητή, δηλαδή το πόσο καλά θα συμπεριφερθεί σε πρότυπα εισόδου που δεν έχει ξανασυναντήσει, αφού έχει ολοκληρώσει την εκπαίδευσή του. Πιο συγκεκριμένα, όσο μεγαλύτερος είναι ο λόγος των προτύπων εκπαίδευσης N προς τον αριθμό των ελευθέρων παραμέτρων ταξινόμησης l, τόσο καλύτερες είναι οι ιδιότητες γενίκευσης του τελικού

ταξινομητή. Όμως ο αριθμός των χαρακτηριστικών είναι ένα βασικό συστατικό των παραμέτρων ταξινόμησης. Συνεπώς ένας μεγάλος αριθμός χαρακτηριστικών μεταφράζεται άμεσα σε ένα μεγάλο αριθμό παραμέτρων ταξινόμησης (π.χ. τα συναπτικά βάρη σε ένα νευρωνικό, τα βάρη σε ένα γραμμικό ταξινομητή). Άρα, καθότι ο αριθμός *N* των προτύπων εκπαίδευσης είναι συγκεκριμένος και συνήθως περιορισμένος, το να κρατήσουμε τον αριθμό των χαρακτηριστικών τόσο μικρό όσο είναι απαραίτητο, είναι σε απόλυτη συμφωνία με την επιθυμία μας να σχεδιάσουμε κατηγοριοποιητές με υψηλό λόγο *N/l* και συνεπώς με καλές ικανότητες γενίκευσης. Τέλος, ένα σημαντικό βήμα στη σχεδίαση ενός συστήματος ταξινόμησης είναι η απόδοση του σταδίου αξιολόγησης, στο οποίο εκτιμάται η πιθανότητα του σφάλματος ταξινόμησης. Η πιθανότητα αυτή βελτιώνεται (μειώνεται), όσο ο λόγος αυτός αυξάνεται.

Για τους λόγους αυτούς επιθυμούμε τη μείωση των χαρακτηριστικών εισόδου του κατηγοριοποιητή στον απολύτως απαραίτητο. Αυτό επιτυγχάνεται με μια σειρά από αλγορίθμους που ονομάζονται γενικά αλγόριθμοι επιλογής χαρακτηριστικών (Feature Selection Algorithms-FSA).

Στόχος ενός αλγορίθμου επιλογής χαρακτηριστικών είναι η εύρεση του βέλτιστου υποσυνόλου χαρακτηριστικών εισόδου, που θα τροφοδοτηθεί στη συνέχεια στον κατηγοριοποιητή. Ο αριθμός των επιλεγμένων χαρακτηριστικών που αποτελούν το βέλτιστο υποσύνολο καθώς και η ταυτότητα τους, δηλαδή το ποια χαρακτηριστικά είναι αυτά, είναι τα δύο ζητήματα που καλείται να αντιμετωπίσει ο εκάστοτε αλγόριθμος. Το πόσο καλά θα τα καταφέρει εξαρτάται πάντα από το είδος της εργασίας καθώς ένας καλός αλγόριθμος για μια περίπτωση, είναι πιθανό να μη λειτουργεί καλά σε κάποια άλλη και αντιστρόφως. Γι' αυτό εξάλλου υπάρχει και μεγάλη ποικιλία.

Γενικά πάντως τείνουμε να διευκολύνουμε τη διαδικασία με το να προεπιλέγουμε τον αριθμό των επιθυμητών χαρακτηριστικών του τελικού υποσυνόλου. Συνεπώς ο αλγόριθμος δημιουργείται για να δώσει απάντηση στην εξής ερώτηση:

Δεδομένου ενός συγκεκριμένου αριθμού χαρακτηριστικών, πως μπορεί κάποιος να επιλέξει τα πιο σημαντικά από αυτά ώστε να μειώσει τον αριθμό τους και την ίδια στιγμή να διατηρήσει όσο περισσότερη πληροφορία σχετικά με το διαχωρισμό τους σε κατηγορίες;

Η διαδικασία αυτή είναι γνωστή ως **επιλογή χαρακτηριστικών** ή **μείωση**. Πρέπει να τονιστεί πως το βήμα αυτό είναι πολύ σημαντικό. Αν επιλέξουμε χαρακτηριστικά με μικρή διαχωριστική ικανότητα, ο επακόλουθος κατηγοριοποιητής θα έχει χαμηλή απόδοση. Από την άλλη πλευρά, εάν τα πλούσια σε πληροφορία χαρακτηριστικά επιλεγούν, η σχεδίαση του κατηγοριοποιητή μπορεί να απλοποιηθεί αρκετά. Σε μια πιο ποσοτική περιγραφή, σκοπός μας είναι η επιλογή χαρακτηριστικών που οδηγούν σε *μεγάλη δια-κατηγοριακή απόσταση και*

μικρή ενδοκατηγοριακή ποικιλομορφία στο χώρο επιλογής χαρακτηριστικών. Αυτό σημαίνει πως τα χαρακτηριστικά θα πρέπει να παίρνουν μακρινές τιμές για διαφορετικές κατηγορίες και κοντινές για την ίδια κατηγορία.

Λαμβάνοντας υπ' όψιν τα παραπάνω, το πρόβλημα της επιλογής χαρακτηριστικών υπό τη μορφή επιβλεπόμενης επαγωγικής μάθησης διατυπώνεται ως εξής,

Δεδομένου ενός συνόλου χαρακτηριστικών επέλεξε και παρουσίασε ένα υποσύνολο με βάση μία από τις 3 κάτωθι προσεγγίσεις:

- το υποσύνολο με συγκεκριμένο μέγεθος που βελτιστοποιεί ένα κριτήριο επιλογής
- το υποσύνολο με το μικρότερο μέγεθος που ικανοποιεί ένα συγκεκριμένο περιορισμό στη μέτρηση του κριτηρίου επιλογής (βρίσκεται δηλαδή πάνω ή κάτω από ένα κατώφλι ανάλογα το κριτήριο)
- το υποσύνολο με τον καλύτερο συνδυασμό των 2 παραπάνω περιπτώσεων (μέγεθος-συγκεκριμένος περιορισμός), η οποία αποτελεί και τη γενικότερη περίπτωση.

Ο γενικευμένος στόχος προς επίτευξη είναι η βελτίωση του υποκειμένου της επαγωγικής μάθησης, του κατηγοριοποιητή δηλαδή, είτε υπό την έννοια της ταχύτητας, είτε της δυνατότητας γενίκευσης ή της απλότητας της απεικόνισης. Αυτό συντείνει στην καλύτερη κατανόηση των αποτελεσμάτων που προκύπτουν, στη μείωση του μεγέθους του αποθηκευτικού χώρου που χρειάζεται για τη λειτουργία του, καθώς επίσης και στη μείωση του θορύβου που δημιουργείται από τα άσχετα ή περιττά χαρακτηριστικά, με τελικό όφελος να αποβληθεί τελικά όλη η περιττή πληροφορία.

Ένας αλγόριθμος επιλογής χαρακτηριστικών (Feature Selection Algorithm) είναι μια υπολογιστική λύση βασισμένη σε ένα συγκεκριμένο ορισμό της έννοιας της σχετικότητας (των χαρακτηριστικών). Η έννοια αυτή, όπως γίνεται αντιληπτή από το υποκείμενο της επαγωγικής μάθησης, μπορεί να έχει πολλούς ορισμούς ανάλογα με το αντικείμενο της εκάστοτε έρευνας. Ένα άσχετο (με τον καθορισμένο στόχο) χαρακτηριστικό σίγουρα δεν είναι χρήσιμο για τη διαδικασία της μάθησης, αλλά από την άλλη δεν είναι απαραίτητο όλα τα σχετικά χαρακτηριστικά να είναι και χρήσιμα.

Οι FSA μπορούν να κατηγοριοποιηθούν ανάλογα με το είδος της εξόδου που δίνουν:

- στους αλγορίθμους εκείνους που αποδίδουν τα χαρακτηριστικά σε μια γραμμική σειρά (σημαντικότητας)
- στους αλγορίθμους εκείνους που δίνουν ένα υποσύνολο των αρχικών χαρακτηριστικών

Στην πραγματικότητα το $2^{ο}$ μπορεί να θεωρηθεί πως αποτελεί υποπερίπτωση του $1^{ου}$, αν υποθέσουμε ότι χρησιμοποιούμε σε αυτό κβαντισμένα βάρη στο 0 και το 1, αντί για μια συνεχή κλίμακα ανάμεσα τους.

3.2 Σχετικότητα ενός χαρακτηριστικού

Ο σκοπός ενός FSA λοιπόν είναι να εντοπίσει τα σχετικά (με τη ζητούμενη έξοδο) χαρακτηριστικά του αρχικού συνόλου, σύμφωνα με έναν ορισμό αυτής της σχετικότητας. Όμως, η έννοια της σχετικότητας στη μηχανική μάθηση δεν έχει ακόμα οριστεί σαφώς. Συνεπώς έχουν προταθεί πολλοί ορισμοί, χωρίς όμως κάποιος ακόμα να υπερισχύει σε όλες τις περιπτώσεις, ώστε να θεωρηθεί ο ακριβέστερος.

Ας ορίσουμε τα Ei, με $1 \leq I \leq n$, ως χώρους χαρακτηριστικών $X=\{x_1, ..., x_n\}$. Ένας στιγμιαίος χώρος (ή στιγμιότυπο του χώρου)είναι ορισμένος ως $E=E_1*...*E_n$, όπου ένα στιγμιότυπο είναι ένα σημείο στο χώρο αυτό. Ας υποθέσουμε ότι το p είναι η κατανομή πιθανότητας στο E και το T ότι είναι ένας χώρος κατηγοριών (ταμπέλες). Το επιθυμητό είναι να μοντελοποιήσουμε ή να διακρίνουμε μια αντικειμενική συνάρτηση c: E->T, σύμφωνα με τα σχετικά χαρακτηριστικά του συνόλου. Αυτό σημαίνει δηλαδή ότι, η συνάρτηση αυτή θα δέχεται τα διάφορα πρότυπα (χώρος E) και θα τα μεταφέρει ανάλογα με τις τιμές των χαρακτηριστικών τους σε κάποιο σημείο στο χώρο T (χώρος κατηγοριών). Ένα σύνολο δεδομένων S συντεθειμένο από |S| στιγμιότυπα μπορεί να θεωρηθεί ως το αποτέλεσμα της δειγματοληψίας του E υπό τη p, για |S| φορές και η κατηγοριοποίηση των στοιχείων του σύμφωνα με τη συνάρτηση c.

Παρακάτω ακολουθούν οι πιο καλές προσπάθειες ορισμού της έννοιας της σχετικότητας ενός χαρακτηριστικού:

Ορισμός $1^{ος}$(Σχετικό με ένα στόχο) *ένα χαρακτηριστικό $x_i \in X$ είναι σχετικό με ένα στόχο c εάν υπάρχουν 2 παραδείγματα A,B προερχόμενα από τον χώρο στιγμιότυπων E τέτοια ώστε το A και το B να διαφέρουν μόνο στην εκχώρηση του x_i και c(A)≠c(B).*

Με άλλα λόγια, εάν υπάρχουν 2 στιγμιότυπα που να μπορούν να κατηγοριοποιηθούν μόνο χάρη στο x_i. Αυτός ο ορισμός έχει την ασυνέπεια ότι ο αλγόριθμος μάθησης δεν μπορεί πάντα να διακρίνει αν ένα χαρακτηριστικό x_i, είναι σχετικό ή όχι, χρησιμοποιώντας μόνο ένα δείγμα S του E. Επιπλέον, αν το πρόβλημα της απεικόνισης είναι περιττό(π.χ. κάποια χαρακτηριστικά επαναλαμβάνονται), δε θα μπορέσουν ποτέ 2 στιγμιότυπα να διαφέρουν μόνο κατά ένα χαρακτηριστικό. Μια πρόταση για την επίλυση αυτού του προβλήματος περιλαμβάνει 2 έννοιες της σχετικότητας, η μία με βάση ένα δείγμα και η άλλη με βάση την κατανομή.

Ορισμός 2ος (Ισχυρή σχετικότητα με το S).*Ένα χαρακτηριστικό $x_i \in X$ είναι ισχυρά σχετικό με ένα στόχο c μέσα σε μια κατανομή P, εάν υπάρχουν 2 παραδείγματα $A, B \in E$ με p(A)≠0 και p(B)≠0 που διαφέρουν μόνο στην εκχώρηση του x_i, και c(A)≠c(B).*

Με άλλα λόγια , ο ορισμός αυτός είναι παρόμοιος με τον ορισμό 1, αλλά τώρα $A, B \in E$ και ο ορισμός γίνεται με βάση το S.

Ορισμός 3ος (Ισχυρή σχετικότητα με το p) *Ένα χαρακτηριστικό $x_i \in X$ είναι ισχυρά σχετικό με ένα στόχο c εντός της κατανομής p εάν υπάρχουν 2 παραδείγματα $A, B \in E$ με p(A)≠0 και p(B)≠0 που μόνο διαφέρουν στην εκχώρηση του x_i και c(A)≠0 και c(B)≠0.*

Αυτός ο ορισμός είναι η φυσική επέκταση του ορισμού 2 και, αντιθέτως με αυτόν, η κατανομή p θεωρείται γνωστή.

Ορισμός 4 (Αδύναμη σχετικότητα με το S) *Ένα χαρακτηριστικό $x_i \in X$ είναι αδύναμα σχετικό με το δείγμα S εάν υπάρχει τουλάχιστον ένα X' υποσύνολο του X ($x_i \in X'$) όπου x_i είναι ισχυρά σχετικό με το S.*

Ένα αδύναμα σχετικό χαρακτηριστικό μπορεί να εμφανιστεί όταν ένα υποσύνολο που περιέχει τουλάχιστον ένα ισχυρά σχετικό χαρακτηριστικό έχει απομακρυνθεί.

Ορισμός 5 (Αδύναμη σχετικότητα με το p)*Ένα χαρακτηριστικό $x_i \in X$ είναι αδύναμα σχετικό με τον στόχο c σε μια κατανομή p εάν υπάρχει τουλάχιστον ένα X' υποσύνολο X ($x_i \in X'$) όπου x_i είναι ισχυρά σχετικό με το p.*

Αυτοί οι ορισμοί είναι σημαντικοί για να αποφασιστεί ποιά χαρακτηριστικά μπορούν να διατηρηθούν και πια οφείλουν να αφαιρεθούν. Τα ισχυρά σχετικά χαρακτηριστικά είναι, θεωρητικά, σημαντικά για τη διατήρηση μιας δομής στο χώρο, και θα πρέπει να διατηρούνται από οποιονδήποτε αλγόριθμο επιλογής χαρακτηριστικών με σκοπό την αποφυγή της προσθήκης ασάφειας στο δείγμα. Τα αδύναμα σχετικά χαρακτηριστικά μπορεί να είναι σημαντικά ή όχι, ανάλογα με τα υπόλοιπα χαρακτηριστικά που έχουν επιλεγεί και με τη μέθοδο αξιολόγησης που έχει επιλεγεί (ακρίβεια, απλότητα, συνέπεια κτλ.).

Από μια άλλη οπτική γωνία, αντί να επικεντρωνόμαστε στο ποιά χαρακτηριστικά είναι σχετικά, είναι πιθανό να χρησιμοποιήσουμε τη σχετικότητα σα μια «μέτρηση πολυπλοκότητας» με βάση τον στόχο c. Σε αυτή την περίπτωση, θα εξαρτάται από τον επιλεγμένο επαγωγέα.

Ορισμός 6 (Σχετικότητα ως μέτρηση πολυπλοκότητας) *Δεδομένου ενός συνόλου δεδομένων S και ενός στόχου c, όρισε r(S,c) ως το μικρότερο αριθμό σχετικών με το c χαρακτηριστικών χρησιμοποιώντας τον ορισμό 1 μόνο στο S, τέτοια ώστε το σφάλμα στο S να είναι το μικρότερο δυνατό από τον επαγωγέα.*

Με άλλα λόγια, αναφέρεται στο μικρότερο αριθμό χαρακτηριστικών που απαιτούνται από ένα συγκεκριμένο επαγωγέα για να φθάσουμε σε μια ιδανική απόδοση στη διαδικασία μοντελοποίησης του c χρησιμοποιώντας το S.

Ορισμός 7 (Αυξητική χρησιμότητα)*Δεδομένου ενός συνόλου δεδομένων S, ενός αλγορίθμου εκμάθησης L, και ενός υποσυνόλου χαρακτηριστικών X', το χαρακτηριστικό x_i είναι αυξητικά χρήσιμο στο L με βάση το X' εάν η ακρίβεια της υπόθεσης που το L παράγει χρησιμοποιώντας το σύνολο των χαρακτηριστικών $\{x_i\}UX'$ είναι καλύτερη από την ακρίβεια που επετεύχθη χρησιμοποιώντας μόνο το υποσύνολο X'.*

Αυτός ο ορισμός είναι εξαιρετικά φυσικός στους FSA που ψάχνουν τον υποχώρο των χαρακτηριστικών με αυξητικό τρόπο, προσθέτοντας ή αφαιρώντας χαρακτηριστικά από την υπάρχουσα λύση. Είναι επίσης συσχετισμένο με την παραδοσιακή έννοια της σχετικότητας, όπως αυτή εμφανίζεται στη βιβλιογραφία της φιλοσοφίας.

Ορισμός 8 (Εντροπική συσχέτιση) *Συμβολίζοντας την εντροπία (κατά Shannon) με H(x) και την αμοιβαία πληροφορία με I(x;y)=H(x)-H(x|y)(η διαφορά της εντροπίας στο x λόγω της γνώσης του y), η εντροπική συσχέτιση του x με το y είναι ορισμένη ως r(x;y)= I (x;y)|H(y).*

Ας ορίσουμε το X ως το αρχικό σύνολο χαρακτηριστικών και το C να είναι ο στόχος με τη μορφή χαρακτηριστικού, ένα σύνολο X' υποσύνολο του X είναι ικανοποιητικό εάν I(X';C)=I(X,C) (π.χ. εάν διατηρεί τη γνωστική πληροφορία). Για ένα ικανοποιητικό σύνολο X', φαίνεται πως r(X';C)=r(X,C). Το πιο σημαντικό σύνολο είναι αυτό το ικανοποιητικό σύνολο $X' \subset X$ για το οποίο η H(X') είναι η μικρότερη. Αυτό υπονοεί πως η r(C;X') είναι μεγαλύτερη. Εν τέλει, ο σκοπός είναι να έχουμε τις r(C;X') και r(X';C)ταυτόχρονα μεγιστοποιημένες.

3.3 Αλγόριθμοι για την επιλογή χαρακτηριστικών

Ένας FSA μπορεί να θεωρηθεί ως μια υπολογιστική λύση με βάση έναν ορισμό της σχετικότητας, παρόλο που σε αρκετές περιπτώσεις και οι υπόλοιποι ορισμοί ακολουθώνται, αλλά σε μια χαλαρότερη βάση.

3.3.1 Ορισμός επιλογής χαρακτηριστικών

Επανερχόμενοι ξανά στην έννοια της επιλογής χαρακτηριστικών έφθασε η ώρα να δώσουμε περισσότερες λεπτομέρειες για τη διαδικασία. Ας ορίσουμε το X ως το αρχικό σύνολο χαρακτηριστικών, με κύρια σχέση $|X|=n$ (δηλαδή το μέγεθος του είναι n στοιχεία). Το πρόβλημα συνεχούς επιλογής χαρακτηριστικών αναφέρεται στην ανάθεση βαρών w_i σε κάθε ένα χαρακτηριστικό $x_i \in X$ με τέτοιο τρόπο ώστε όσο μεγαλύτερη είναι η τιμή, τόσο πιο σχετικό να είναι το χαρακτηριστικό. Το δυαδικό πρόβλημα επιλογής χαρακτηριστικών

αναφέρεται στην ανάθεση δυαδικών βαρών, δηλαδή κβαντισμένων στο 0 και στο 1. Αυτό μπορεί να εκτελεσθεί άμεσα ή ύστερα από φιλτράρισμα της εξόδου της συνεχούς λύσης.

Τα δύο αυτά προβλήματα είναι εντελώς διαφορετικά και αντικατοπτρίζουν διαφορετικούς σχεδιαστικούς στόχους. Στη συνεχή περίπτωση, ενδιαφερόμαστε για τη διατήρηση όλων των χαρακτηριστικών αλλά τα χρησιμοποιούμε με άνισο τρόπο στη διαδικασία μάθησης. Από την άλλη πλευρά, στη δυαδική περίπτωση ενδιαφερόμαστε να διατηρήσουμε απλά ένα υποσύνολο των χαρακτηριστικών και να τα χρησιμοποιήσουμε εξ' ίσου στη διαδικασία μάθησης. Το δυαδικό πρόβλημα είναι αυτό που μας ενδιαφέρει στην εργασία μας, καθώς για τους λόγους που έχουμε προαναφέρει θέλουμε οπωσδήποτε να μειώσουμε τον αριθμό των χαρακτηριστικών που θα χρησιμοποιηθούν από τον κατηγοριοποιητή.

Το πρόβλημα επιλογής χαρακτηριστικών μπορεί να θεωρηθεί ως μια έρευνα στο χώρο της υπόθεσης (το σύνολο των πιθανών λύσεων). Στην περίπτωση του δυαδικού προβλήματος, ο αριθμός των πιθανών υποσυνόλων προς υπολογισμό είναι 2n. Σε αυτή την περίπτωση, ο γενικός ορισμός είναι:

Ορισμός 9 (Επιλογή Χαρακτηριστικών):*Έστω J(X') ένα κριτήριο μέτρησης προς βελτιστοποίηση (ας πούμε προς μεγιστοποίηση) ορισμένο ως J:X' (υποσύνολο του X)->R. Η επιλογή ενός υποσυνόλου χαρακτηριστικών μπορεί να ειδωθεί υπό το πρίσμα των κάτωθι 3 συλλογισμών:*

- *Ορίζουμε |X'|=m<n. Βρές το $X' \subset X$, τέτοιο ώστε J(X') να είναι μέγιστο.*
- *Ορίζουμε μια τιμή J_o, ως το ελάχιστο J το οποίο είναι αποδεκτό. Βρές το $X' \subset X$ με μικρότερο |X'|, τέτοιο ώστε $J(X') \geq J_o$.*
- *Βρές ένα συμβιβασμό ανάμεσα στην ελαχιστοποίηση του |X'| και στη μεγιστοποίηση του J(X') (γενική περίπτωση)*

Προσέξτε ότι, με αυτούς τους ορισμούς, ένα ιδανικό υποσύνολο χαρακτηριστικών δεν είναι απαραίτητα μοναδικό.

3.3.2 Χαρακτηρισμός των FSA

Υπάρχουν στη βιβλιογραφία αρκετοί χαρακτηρισμοί των αλγορίθμων επιλογής χαρακτηριστικών. Βλέποντας τους είναι δυνατόν να περιγράψουμε αυτό το χαρακτηρισμό ως ένα πρόβλημα έρευνας στο χώρο των υποθέσεων όπως παρακάτω:

Οργάνωση έρευνας. Η γενική στρατηγική με την οποία εξερευνάται ο χώρος των υποθέσεων. Αυτή η στρατηγική έχει σχέση με το μέγεθος της εξερευνούμενης υπόθεσης σε σχέση με το συνολικό αριθμό των στοιχείων της.

Γενιά επιτυχόντων. Ο μηχανισμός με τον οποίο προτείνονται οι πιθανές μεταβλητές (υποψήφιοι επιτυχόντες) της υπάρχουσας υπόθεσης.

Μέτρηση Αξιολόγησης. Συνάρτηση με την οποία οι υποψήφιοι επιτυχόντες υπολογίζονται, επιτρέποντας τη σύγκριση διαφορετικών υποθέσεων για την καθοδήγηση της διαδικασίας της έρευνας.

3.3.2.1 Οργάνωση της έρευνας

Ένας αλγόριθμος έρευνας είναι υπεύθυνος για την καθοδήγηση της διαδικασίας επιλογής χαρακτηριστικών χρησιμοποιώντας μια συγκεκριμένη στρατηγική. Κάθε κατάσταση στο χώρο έρευνας ορίζει ένα βάρος $w_1,\ldots,w_n$ των πιθανών χαρακτηριστικών του X, με $|X|=n$. Στη δυαδική διαδικασία $w_i \in \{0,1\}$(δηλαδή είναι είτε 0 είτε 1), ενώ στη συνεχή περίπτωση $w_i \in [0,1]$(δηλαδή παίρνει τιμές ανάμεσα στο 0 και το 1). Παρατηρήστε ότι αναφέρουμε πως η σχετικότητα πρέπει να είναι και άνω και κάτω φραγμένη. Επίσης στη δυαδική περίπτωση μια μερική σειρά είναι υπαρκτή στο χώρο έρευνας, με S1<S2 εάν S1 υποσύνολο του S2 (Σχήμα 1). ενώ στην συνεχή περίπτωση S1<S2 εάν, για όλα τα I, ισχύει $w_i(S1) \leq w_i(S2)$. (Σχήμα 2)

Γενικά, μια διαδικασία έρευνας εξετάζει μόνο ένα μέρος του χώρου έρευνας. Όταν πρέπει να εξετασθεί μια συγκεκριμένη κατάσταση, ο αλγόριθμος χρησιμοποιεί την πληροφορία από τις προηγούμενες καταστάσεις και τελικά ευρετική γνώση σχετικά με αυτές που δεν έχουν ακόμα εξετασθεί.

Έστω L μια (κατηγοριοποιημένη) λίστα υποσυνόλων τα οποία περιέχουν χαρακτηριστικά με βάρη. Συνεπώς η L περιέχει την ταξινομημένη λίστα των λύσεων. Οι ταμπέλες υποδεικνύουν την αξία των μετρήσεων αξιολόγησης. Θεωρούμε τρία είδη έρευνας: την εκθετική, τη σειριακή και την τυχαία. Οι περισσότεροι σειριακοί αλγόριθμοι χαρακτηρίζονται από |L|=1, ενώ οι εκθετικοί και οι τυχαίοι συνήθως χρησιμοποιούν |L|≥1.

Εκθετική έρευνα: Αναφέρεται σε αλγορίθμους που φέρουν εις πέρας έρευνες με κόστος $O(2^n)$. Η εξαντλητική έρευνα είναι μια βέλτιστη έρευνα, υπό την έννοια ότι η καλύτερη λύση είναι εγγυημένη. Μια βέλτιστη λύση δεν είναι απαραίτητο να είναι εξαντλητική, για παράδειγμα, εάν μια μέτρηση αξιολόγησης είναι μονότονη, ο αλγόριθμος Branch and Bound είναι ο καλύτερος. Μια μέτρηση J είναι μονότονη εάν για κάθε δύο υποσύνολα S1,S2 και S1 C S2, και J(S1)≥J(S2).

Σειριακή έρευνα: Αυτό το είδος έρευνας επιλέγει έναν από όλους τους διαδόχους της υπάρχουσας κατάστασης. Αυτό γίνεται με ένα επαναληπτικό τρόπο και όταν πλέον η επόμενη κατάσταση έχει επιλεγεί δεν είναι δυνατόν να υπάρξει γυρισμός. Παρόλο που δεν υπάρχει υποχρεωτικό πισωγύρισμα στη διαδικασία, ο αριθμός των βημάτων πρέπει να είναι περιορισμένος από το O(n) αν θέλουμε να χρησιμοποιήσουμε σειριακή έρευνα. Η πολυπλοκότητα καθορίζεται λαμβάνοντας υπ' όψιν τον αριθμό k των αξιολογημένων υποσυνόλων σε κάθε αλλαγή κατάστασης. Το κόστος αυτής της έρευνας είναι λοιπόν πολυωνυμικό $O(n^k+1)$. Συνεπώς, αυτές οι μέθοδοι δεν εγγυώνται τη βέλτιστη λύση, καθότι η

λύση αυτή μπορεί να είναι ένα μέρος του χώρου έρευνας που ο αλγόριθμος δεν έχει επισκεφτεί.

Τυχαία έρευνα: Η ιδέα πίσω από αυτόν τον τύπο έρευνας είναι να χρησιμοποιηθεί η τυχαιότητα ώστε να αποφευχθεί το κόλλημα του αλγορίθμου σε τοπικά ελάχιστα και να επιτρέπει την προσωρινή (αλλά και απαραίτητη) μετάβαση σε άλλες καταστάσεις με χειρότερες λύσεις. Αυτοί είναι αλγόριθμοι γενικής χρήσης και μπορούν να δώσουν αρκετά βέλτιστα υποσύνολα ως λύση.

3.4 Τελεστές

3.4.1 Γενιά διαδόχων

Μέχρι πέντε διαφορετικοί τελεστές μπορεί να θεωρηθεί ότι γεννούν διαδόχους για κάθε κατάσταση: *Εμπρόσθιος, Όπισθεν, Σύνθετος, Βαρύνων και Τυχαίος*.

Όλοι αυτοί οι τελεστές δρούν με το να μετατρέπουν με κάποιο τρόπο τα βάρη w_i των χαρακτηριστικών x_i, με $w_i \in R$(στην περίπτωση του βαρύνοντα τελεστή), ή $w_i \in \{0,1\}$ (στην περίπτωση των λοιπών τελεστών). Στις ακόλουθες περιγραφές, θεωρείται ως δεδομένο ότι η τιμή του κριτηρίου J θέλουμε να μεγιστοποιηθεί.

Εμπρός(Forward): Αυτός ο τελεστής προσθέτει χαρακτηριστικά στην υπάρχουσα λύση X', επιλέγοντας ανάμεσα από αυτά που δεν έχουν ακόμα επιλεγεί. Σε κάθε βήμα, το χαρακτηριστικό που αυξάνει την τιμή της J προστίθεται στη λύση. Ξεκινώντας με X'=0, το *εμπρόσθιο* βήμα αποτελείται από:

$$X' \coloneqq X' \cup \{x_i \in X \backslash X' | J(X' \cup \{x_i\}) \text{ } \textit{είναι μεγαλύτερο} \text{ } \}$$

Το κριτήριο τερματισμού της προώθησης μπορεί να είναι:

- |X'|=n' (εάν το n' είναι προκαθορισμένο),
- η τιμή της J δεν έχει αυξηθεί στα τελευταία j βήματα, ή
- έχει υπερβεί ένα κατώφλι J_o.

Το κόστος του τελεστή είναι O(n). Το κύριο μειονέκτημα είναι ότι δεν είναι δυνατόν να λαμβάνονται υπ' όψιν συγκεκριμένες βασικές σχέσεις ανάμεσα στα χαρακτηριστικά. Για παράδειγμα, εάν x_1, x_2 είναι τέτοια ώστε J({x_1, x_2}) >> J({x_1}), J({x_2}), κανένα από τα x_1 και x_2 δε θα μπορούσε να επιλεγεί, παρόλο που είναι αρκετά χρήσιμα.

Όπισθεν(Backward): Αυτός ο τελεστής απομακρύνει χαρακτηριστικά από τη συγκεκριμένη λύση X' ανάμεσα από αυτά που δεν έχουν ακόμα απομακρυνθεί. Σε κάθε βήμα, το

χαρακτηριστικό που με την απομάκρυνση του θα αυξήσει τη J βγαίνει από τη λύση. Ξεκινώντας με X'=X, ο *όπισθεν* τελεστής αποτελείται από:

$$X' \coloneqq X'\{x_i \in X' | J(X'\{x_i\}) \text{ } \textit{είναι μεγαλύτερο}\}$$

Το κριτήριο τερματισμού μπορεί να είναι κάποιο από τα εξής:

- |X'|=v',
- η τιμή του J δεν έχει αυξηθεί στα τελευταία j βήματα, ή
- πέφτει κάτω από μια προκαθορισμένη τιμή J_o.

Αυτός ο τελεστής αντιμετωπίζει με επιτυχία κάποια προβλήματα παρόλο που ακόμα θα υπάρχουν πολλές *κρυμμένες* σχέσεις. Το κόστος είναι όπως και προηγουμένως O(n), παρόλο που στην πράξη χρειάζεται περισσότερους υπολογισμούς από το εμπρόσθιο αντίστοιχο του.

Και οι δύο τελεστές (Εμπρόσθιος και Όπισθεν) μπορούν να γενικευθούν επιλέγοντας σε κάθε βήμα, υποσύνολα k στοιχείων X'' και επιλέγοντας αυτό το οποίο κάνει το J(X'UX'') ή J(X'\X'') μεγαλύτερο, αντίστοιχα. Το κόστος του τελεστή είναι τότε $O(n^k)$.

Επίσης οι τελεστές αυτοί είναι και οι βασικοί και χρησιμοποιούνται σε διαφορετικούς συνδυασμούς ώστε να δώσουν μορφή και στους υπόλοιπους τελεστές. Μπορούμε να πούμε πως είναι τα ανάλογα της πρόσθεσης και της αφαίρεσης των αλγεβρικών τελεστών.

Σύνθετος: η ιδέα αυτής της τακτικής είναι απλή: εφάρμοσε *f* συνεχόμενα εμπρόσθια βήματα και *b* συνεχόμενα όπισθεν. Εάν το αποτέλεσμα είναι *f>b*, το τελικό αποτέλεσμα είναι ένας εμπρόσθιος τελεστής, ειδάλλως είναι όπισθεν. Μια ενδιαφέρουσα προσέγγιση είναι να εκτελέσουμε τα εμπρόσθια ή τα όπισθεν βήματα, ανάλογα με τις αντίστοιχες τιμές της J. Αυτό επιτρέπει την ανακάλυψη νέων σχέσεων ανάμεσα στα χαρακτηριστικά. Ένας ενδιαφέρων μηχανισμός αντίστροφης πορείας καθιερώνεται, παρόλο που άλλες συνθήκες τερματισμού πρέπει να θεσπιστούν αν *f=b*. Για παράδειγμα, για *f=b=1*, εάν x_i προστίθεται και x_j απομακρύνεται, αυτό μπορεί να αναιρεθεί στα βήματα που ακολουθούν. Ένα πιθανό κριτήριο τερματισμού είναι $x_i=x_j$. Στο σειριακό FSA, η συνθήκη f≠b σιγουρεύει ένα μέγιστο v βημάτων, με συνολικό κόστος $O(n^{f+b+1})$.

Βαρύνων: Στους τελεστές με βάρη, ο χώρος έρευνας είναι συνεχής, και όλα τα χαρακτηριστικά είναι παρόντα στη λύση μέχρι ενός βαθμού. Μια κατάσταση επιτυχόντων είναι μια κατάσταση με διαφορετικά βάρη. Αυτό γίνεται συνήθως με αναδρομική δειγματοληψία των διαθέσιμων στιγμιότυπων.

Τυχαίος: Αυτή η ομάδα περιλαμβάνει αυτούς τους τελεστές που μπορούν να δημιουργήσουν οποιαδήποτε νέα κατάσταση μέσα σε ένα βήμα. Οι προηγούμενοι τελεστές μπορούν να έχουν

κάποια τυχαία στοιχεία, αλλά περιορίζονται σε ένα κριτήριο «προόδου» στον αριθμό των χαρακτηριστικών ή στη βελτίωση της τιμής της J σε κάθε βήμα.

3.4.2 Κριτήρια Αξιολόγησης

Υπάρχουν αρκετές προσεγγίσεις στην εκτίμηση της ποιότητας της J(X') ενός υποσυνόλου χαρακτηριστικών X'. Είναι προφανές πως η σχετικότητα ενός χαρακτηριστικού είναι απλά μια συνάρτηση του τύπου μέτρησης και όχι ίδιον του χαρακτηριστικού αυτού καθ' εαυτού. Αυτό σημαίνει πως ένα χαρακτηριστικό θεωρείται κατάλληλο αν μας βοηθά να προσεγγίσουμε τον συγκεκριμένο στόχο μας, ενώ διαφορετικά θεωρείται ακατάλληλο. Μια άλλη σημαντική σκέψη είναι το ότι το εύρος και η κλίμακα του J δεν είναι σημαντικά. Αυτό που μετράει είναι, οι σχετικές αξίες που έχουν ανατεθεί στα διαφορετικά υποσύνολα να αντικατοπτρίζουν τη μεγαλύτερη ή μικρότερη σχέση τους με την αντικειμενική συνάρτηση αξιολόγησης. Ανάμεσα στα μελετημένα μέτρα το πιθανοτικό και οι διακατηγοριακές αποστάσεις, μαζί με τη συνέπεια, είναι μεγέθη διαχωρισμού των κατηγοριών. Επιπλέον, η διακατηγοριακή απόσταση, συνέπεια, εντροπία και εκτίμηση της πιθανότητας λάθους μπορεί να μη χρειάζονται ακριβή μοντελοποίηση των πιθανοτήτων κατανομής.

Έστω J: $X' \subset X$->R είναι ένα κριτήριο αξιολόγησης προς μεγιστοποίηση, όπου X' είναι ένα υποσύνολο χαρακτηριστικών, με βάρη.

Πιθανότητα λάθους: Δεδομένου ότι ο κύριος στόχος είναι το κτίσιμο ενός κατηγοριοποιητή ικανό να βάζει σωστές ταμπέλες στα στιγμιότυπα όπως αυτά έχουν δημιουργηθεί από την ίδια πιθανότητα κατανομής, η ελαχιστοποίηση της Μπαγεσιανής πιθανότητας σφάλματος P του κατηγοριοποιητή φαίνεται να είναι η πιο φυσική επιλογή. Συνεπώς, αποτελεί και μια καθαρή επιλογή για τη J.

Έστω $x \in R$, να παρουσιάζει τα μη-κατηγοριοποιημένα στιγμιότυπα, και Ω=(ω_1, ...,ω_m) ένα σύνολο κλάσεων, τέτοιο ώστε c:R^n->Ω.

Αυτή η πιθανότητα ορίζεται ως:

$$P_e = \int \left[1 - \max_i P(\omega_i|\vec{x})\right] p(\vec{x}) d\vec{x}$$

όπου $p(\vec{x}) = \sum_{i=1}^{m} p(\vec{x}|\omega_i)P(\omega_i)$ είναι η () κατανομή πιθανότητας των στιγμιότυπων, και $P(\omega_i|\vec{x})$ είναι η *a posteriori* πιθανότητα του ω_i να ανήκει στην κατηγορία του $\vec{x}$.

Από τη στιγμή που οι εξαρτώμενες από την κατηγορία πυκνότητες είναι συνήθως άγνωστες, μπορούν είτε να μοντελοποιηθούν άμεσα (χρησιμοποιώντας παραμετρικές ή μη-παραμετρικές μεθόδους) ή έμμεσα μέσω της σχεδίασης ενός κατηγοριοποιητή ο οποίος χτίζει τα αντίστοιχα

όρια αποφάσεων μεταξύ των κλάσεων. Μερικοί από αυτούς τους κατηγοριοποιητές, όπως και στον κανόνα του κοντινότερου γείτονα, έχουν άμεση σχέση με την πιθανότητα λάθους.

Η χρήση (μιας προσέγγισης $\widehat{\widehat{P}_e}$) της πιθανότητας για την κατασκευή ενός κατηγοριοποιητή, χρησιμοποιώντας ένα δείγμα του συνόλου δεδομένων, είναι η βάση των μεθόδων *wrapper*. Δεδομένου ότι ο κατηγοριοποιητής έχει κατασκευαστεί χρησιμοποιώντας μόνο ένα υποσύνολο X' C X των χαρακτηριστικών, προκύπτει:

$$\widehat{P_e} = 1 - \frac{|S_{TE}^{+}|}{|S_{TE}|}$$

έτσι ώστε J=1-$\widehat{P_e}$, με το S_{TE} να αποτελεί το δείγμα ελέγχου και το S_{TE}^{+} το υποσύνολο εκείνο του S_{TE} στο οποίο ο κατηγοριοποιητής απέδωσε σωστά (και εδώ χρησιμοποιώντας μόνο μια μερική περιγραφή X').

Για την προσέγγιση $\widehat{P_e}$, μπορεί να χρειάζονται πιο εξειδικευμένες μεθόδοι ώστε να επιτύχει το σύστημα μια πιο αξιόπιστη τιμή.

Απόκλιση: Αυτές οι μετρήσεις υπολογίζουν μια πιθανοτική απόσταση γνωστή και ως *απόκλιση* ανάμεσα στην class-conditional πυκνότητες πιθανότητας $p(x|\omega_i)$, χρησιμοποιώντας τη γενική μέθοδο:

$$J = \int f[p(\vec{x}|\omega_1), p(\vec{x}|\omega_2)]d\vec{x}$$

Για να θεωρηθεί η μέτρηση έγκυρη, θα πρέπει η συνάρτηση *f* να είναι τέτοια ώστε η τιμή του J να ικανοποιεί τις παρακάτω συνθήκες:

- J≥0
- J=0 μόνο όταν οι $p(x|\omega_i)$ είναι ίσες και
- η J γίνεται μέγιστη όταν δεν επικαλύπτονται

Αν τα χαρακτηριστικά που χρησιμοποιούνται σε μια λύση X'C X είναι καλά, η απόκλιση ανάμεσα στις εξαρτημένες πιθανότητες θα είναι σημαντική. Κακά χαρακτηριστικά θα δώσουν παρόμοιες πιθανότητες. Μερικές κλασικές επιλογές τύπων μέτρησης είναι:

Chernoff

$$f(a,b) = a^s b^{1-s}, s \in [0,1] \kappa\alpha\iota\ \tau\acute{o}\tau\varepsilon\ J_{Che} = -lnJ$$

Bhattacharyya

$$f(a,b) = \sqrt{ab}\ \kappa\alpha\iota\ \tau\acute{o}\tau\varepsilon\ J_{Bha} = -lnJ$$

Kullback-Liebler

$$f(a,b) = (a-b)(lna - lnb) \; \kappa\alpha\iota \; \tau\acute{o}\tau\varepsilon \; J_{KL} = J$$

Kolmogorov

$$f(a,b) = |a-b| \; \kappa\alpha\iota \; \tau\acute{o}\tau\varepsilon \; J_{Kol} = J$$

Matusita

$$f(a,b) = (\sqrt{a} - \sqrt{b})^2 \; \kappa\alpha\iota \; \tau\acute{o}\tau\varepsilon \; J_{Mat} = \sqrt{J}$$

Patrick-Fisher

$$f(a,b) = (a-b)^2 \kappa\alpha\iota \; \tau\acute{o}\tau\varepsilon \; J_{PF} = \sqrt{J}$$

Αυτοί οι τύποι μέτρησης αιτιολογούν τις προηγούμενες συνθήκες και μπορούν να χρησιμοποιηθούν και με τη μορφή των βαρών, λαμβάνοντας υπ' όψιν τις πρώτης τάξεως πιθανότητας $P(\omega_i)$ έτσι ώστε το $f[p(\vec{x}|\omega_1), p(\vec{x}|\omega_2)]$ να γίνει $f[p(\vec{x}|\omega_1)P(\omega_1), p(\vec{x}|\omega_2)P(\omega_2)]$. Μπορούν επίσης να σχετισθούν και με το P_e υπό τη μορφή άνω συνόρου.

Εξάρτηση: Αυτές οι μετρήσεις ποσοτικοποιούν το πόσο ισχυρά σχετίζονται 2 χαρακτηριστικά μεταξύ τους, υπό την έννοια ότι γνωρίζοντας την τιμή του ενός είναι πιθανό να προβλεφθεί η τιμή του άλλου. Στο πλαίσιο της επιλογής χαρακτηριστικών, ένα χαρακτηριστικό λαμβάνει μεγαλύτερη αξία, όσο καλύτερα προβλέπει την κατηγορία. Ο παράγοντας συσχέτισης είναι μια κλασσική μέτρηση που ακόμα βρίσκει εφαρμογή. Μια κάπως διαφορετική προσέγγιση αποτελεί η εκτίμηση της απόκλισης μεταξύ των εξαρτώμενων από την κατηγορία πιθανοτήτων και των ανεξάρτητων. Συγκεκριμένα, έχουμε μετρήσεις της μορφής $f[p(\vec{x}|\omega_i), p(\vec{x})], i = 1,2$.

Διακλασική απόσταση: Αυτές οι μετρήσεις βασίζονται στην υπόθεση ότι στιγμιότυπα από διαφορετικές κατηγορίες είναι απομακρυσμένα στο χώρο των στιγμιότυπων. Τότε είναι αρκετό να ορίσουμε ένα μέτρο μεταξύ των κατηγοριών και να το χρησιμοποιήσουμε ως μέτρηση:

$$D(\omega_i, \omega_j) = \frac{1}{N_i N_j} \sum_{k_1}^{N_i} \sum_{k_2=k_1+1}^{N_j} d(x_{(i,k_1)}, x_{(j,k_2)})$$

$$J = \sum_{i=1}^{m} P(\omega_i) \sum_{j=i+1}^{m} P(\omega_j) D(\omega_i, \omega_j)$$

με το $x_{(i,j)}$ να είναι το στιγμιότυπο j της κατηγορίας ω_i, και N_i ο αριθμός των στιγμιότυπων της κλάσης ω_i. Οι πιο συνηθισμένες αποστάσεις d ανήκουν στην οικογένεια των Ευκλείδειων

αποστάσεων. Αυτές οι μετρήσεις δε χρειάζονται τη μοντελοποίηση οποιασδήποτε συνάρτησης πυκνότητας, αλλά η σχέση τους με τη πιθανότητα σφάλματος μπορεί να είναι πολύ χαλαρή.

Πληροφορία ή Αβεβαιότητα: Ομοίως με την πιθανοτική εξάρτηση, μπορούμε να παρατηρήσουμε το $\vec{x}$ και να υπολογίσουμε τις a posteriori πιθανότητες $P(\omega_i|\vec{x})$ για να αποφασίσουμε πόση πληροφορία από την κατηγορία $\vec{x}$ έχει ανακτηθεί, με βάση την προηγούμενη πιθανότητά της. Εάν όλες οι κατηγορίες γίνουν χοντρικά το ίδιο πιθανές, τότε το κέρδος πληροφορίας είναι μηδαμινό και η αβεβαιότητα (εντροπία) είναι μέγιστη. Πολλές μετρήσεις μπορούν τότε να εφαρμοσθούν κάνοντας χρήση των $p(\vec{x})$ και του συνόλου $\{P(\omega_i|\vec{x})\}, i = 1, \dots, n$. Για παράδειγμα, χρησιμοποιώντας την εντροπία κατά Shannon έχουμε:

$$J_{Sha} = -\int p(\vec{x}) \sum_{i=1}^{m} P(\omega_i|\vec{x})\, log_2\, P(\omega_i|\vec{x}) d\vec{x}$$

Η εντροπία μπορεί να χρησιμοποιηθεί και χωρίς τη γνώση των πυκνοτήτων όπως γίνεται στα επαγωγικά δέντρα, όπου το *κέρδος πληροφορίας* υπολογίζεται συνήθως ξεχωριστά για κάθε χαρακτηριστικό στην επαγωγική διαδικασία.

Συνέπεια: Μια *ασυνέπεια* μεταξύ X' και S ορίζεται όταν έχουμε δύο στιγμιότυπα στο S τα οποία είναι ίσα όταν λαμβάνονται υπ' όψιν μόνο τα χαρακτηριστικά του X' που ανήκουν σε διαφορετικές κατηγορίες. Ο σκοπός αυτού είναι να βρεθεί το *ελάχιστο* υποσύνολο χαρακτηριστικών που οδηγεί σε μηδενική ασυνέπεια. Ο *βαθμός ασυνέπειας* ενός στιγμιότυπου $A \in S$ είναι τότε:

$$IC_{X'}(A) = X'(A) - \max_k X'_k(A)$$

όπου X'(A) είναι ο αριθμός των στιγμιότυπων στο S ίσα με A χρησιμοποιώντας μόνο τα χαρακτηριστικά στο X' και X'_k είναι ο αριθμός των στιγμιότυπων στο S της κατηγορίας k ίσα με A χρησιμοποιώντας μόνο τα χαρακτηριστικά του X'. Ο ρυθμός ασυνέπειας ενός συνόλου χαρακτηριστικών σε ένα δείγμα S είναι τότε:

$$IR(X') = \frac{\sum_{A \in S} IC_{X'}(A)}{|S|}$$

Αυτό είναι ένα μονότονο μέγεθος, υπό την έννοια ότι:

$$X_1 \subset X_2 \Longrightarrow IR(X_1) \geq IR(X_2)$$

Μια πιθανή μέτρηση αξιολόγησης είναι τότε $J(X') = \frac{1}{IR(X')+1}$. Αυτό το μέγεθος βρίσκεται μεταξύ [0,1] και μπορεί να υπολογισθεί σε χρόνο O(|S|) χρησιμοποιώντας ένα πίνακα κατακερματισμού.

3.5 Πλαίσια που καθορίζουν τη σχέση μεταξύ κατηγοριοποιητή και FSA

Η σχέση μεταξύ ενός FSA και του επιλεγμένου κατηγοριοποιητή/επαγωγέα για την εκτίμηση της χρησιμότητας της διαδικασίας επιλογής χαρακτηριστικών μπορεί να τοποθετηθεί σε 3 βασικά πλαίσια: *ενσωματωμένο, φιλτραρισμένο* και *αναδιπλούμενο.*

Ενσωματωμένο(Embedded): Ο κάθε επαγωγέας έχει το δικό του FSA (είτε άμεσα είτε έμμεσα) και αυτά τα δύο είναι αλληλοεξαρτούμενα. Οι μέθοδοι που χρησιμοποιούνται για την επαγωγή λογικών ενώσεων αποτελούν ένα καλό παράδειγμα. Άλλα παραδοσιακά εργαλεία μηχανικής μάθησης όπως τα δέντρα αποφάσεων ή τα τεχνητά νευρωνικά δίκτυα συμπεριλαμβάνονται εδώ.

Φιλτραρισμένο(Filter): Αν η διαδικασία επιλογής χαρακτηριστικών λαμβάνει χώρα πριν την επαγωγή, μπορεί να θεωρηθεί ως φίλτρο μη χρήσιμων χαρακτηριστικών πριν την επαγωγή. Γενικότερα, η επιλογή χαρακτηριστικών μπορεί να ειδωθεί ως μια προεπεξεργασία. Η βασική ιδιότητα πάντως των φιλτραρισμένων πλαισίων είναι η ανεξαρτησία τους από τον μετέπειτα αλγόριθμο επαγωγής.

Αναδιπλούμενο(Wrapper): Σε αυτήν την περίπτωση η σχέση αντιστρέφεται: είναι ο FSA που χρησιμοποιεί τον αλγόριθμο μάθησης ως υπορουτίνα. Το γενικό επιχείρημα προς υποστήριξη αυτού του πλαισίου είναι η ισορροπημένη πόλωση τόσο του FSA όσο και του αλγορίθμου μάθησης που θα χρησιμοποιηθεί αργότερα για την εκτίμηση της ποιότητας της λύσης. Το βασικό μειονέκτημα είναι το υπολογιστικό κόστος που προέρχεται από την κλήση του επαγωγικού αλγορίθμου για την εκτίμηση κάθε υποσυνόλου επιλεγμένων χαρακτηριστικών.

3.6 Γενικός αλγόριθμος για την επιλογή χαρακτηριστικών

Ένας γενικός αλγόριθμος για την επιλογή χαρακτηριστικών, που φανερώνει σε μια ενοποιημένη μορφή, τη συμπεριφορά οποιουδήποτε FSA παρουσιάζεται στον ακόλουθο ψευδοκώδικα:

Είσοδος:

S-σύνολο δεδομένων νε χαρακτηριστικά $X, |X|=n$

J- κριτήριο αξιολόγησης προς μεγιστοποίηση

GS- τελεστής γενιάς επιτυχόντων

Είσοδος:

Λύση- (με βάρη) σύνολο χαρακτηριστικών

L:= Σημείο_Εκκίνησης(*X*);
Λύση:= { καλύτερα του *L* με βάση τη *J*};

Επανέλαβε
L:= Στρατηγική_΄Ερευνας(*L,GS(J),X*);
X':= {καλύτερα του *L* σύμφωνα με τη *J*};
εάν *J(X')≥J(Λύση)* **ή** (*J(X')=J(Λύση))* **και** |X'|<|Λύση|)
τότε *Λύση*:=*X'*;
μέχρι Σταμάτα(*J,L*)

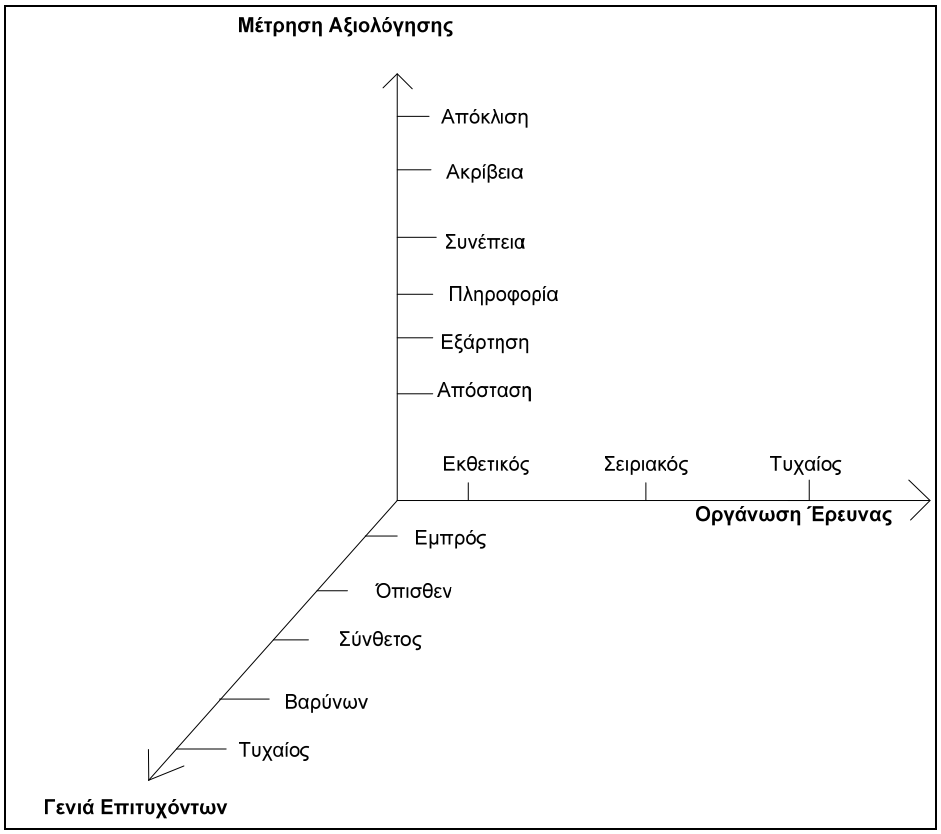
Συγκεκριμένα, το L είναι μια λίστα η οποία περιέχει υποσύνολα χαρακτηριστικών χαρακτηριζόμενα από βάρη και διατηρεί το ταξινομημένο σύνολο λύσεων. Οι εκθετικοί αλγόριθμοι χαρακτηρίζονται συνήθως από |L|≥1 (π.χ. ο BRANCH AND BOUND ή ο A*). Η παρουσία στη λίστα είναι μια συνάρτηση του κριτηρίου αξιολόγησης και καθορίζει τη σειρά επέκτασης. Οι ευριστικοί αλγόριθμοι επίσης διατηρούν αυτή τη λίστα (ανοιχτών βρόγχων), και η ανάθεση των βαρών γίνεται ευριστικά. Οι τυχαίες μέθοδοι αναζήτησης όπως οι Εξελικτικοί Αλγόριθμοι χαρακτηρίζονται από |L|≥1 (η λίστα είναι ο πληθυσμός και τα βάρη είναι η τιμή της καταλληλότητας των υποκειμένων). Οι σειριακοί αλγόριθμοι διατηρούν |L|=1, παρόλο που υπάρχουν εξαιρέσεις.

Η αρχική λίστα *L* είναι γενικά κατασκευασμένη από το αρχικό σύνολο χαρακτηριστικών και ο αλγόριθμος διατηρεί την καλύτερη λύση κάθε φορά. Σε κάθε βήμα, ένας FSA με δεδομένη οργάνωση αναζήτησης τροποποιεί τη λίστα με συγκεκριμένο τρόπο και καλεί το μηχανισμό για τη γενιά των διαδόχων ο οποίος με τη σειρά του χρησιμοποιεί τη *J*. Το αποτέλεσμα είναι μια ανανεωμένη λίστα και η ενημέρωση της καλύτερης μέχρι στιγμής λύσης. Παρατηρείστε ότι το δείγμα δεδομένων *S* θεωρείται γνωστό σε όλο τον αλγόριθμο.

Χώρος χαρακτηριστικών του FSA
Γενικεύοντας τις γνώσεις μας σχετικά με τα FSA, θεωρούμε πως όλα μπορούν να αναπαρασταθούν σε ένα χώρο χαρακτηριστικών του οποίου οι άξονες αντιστοιχούν στα εξής κριτήρια:

- οργάνωση αναζήτησης (Org),
- γενιά διαδόχων (GS) και
- τιμή συνάρτησης αξιολόγησης (J)

σύμφωνα με τα όσα έχουμε πει μέχρι στιγμής. Αυτά τα κριτήρια δομούν λοιπόν έναν νέο τρισδιάστατο χώρο, τον <Org,GS,J> . Ο χώρος αυτός ανοίγει ένα καινούργιο σύνολο πιθανοτήτων για ένα FSA. Νέες προτάσεις για εκτιμήσεις μετρήσεων θα διατείνουν τον κάθετο άξονα.



Σχήμα3.1 Χώρος χαρακτηριστικών FSA

Παρατηρείστε ότι η οργάνωση της αναζήτησης και η γενιά των καταστάσεων διαδοχής είναι συντεταγμένες των οποίων τα στοιχεία είναι, κατά κανόνα, αποκλειόμενα μεταξύ τους, υπό την έννοια ότι ένα συγκεκριμένο FSA θα χρησιμοποιήσει μόνο ένα συνδυασμό οργάνωσης/γενεάς. Όταν ένα FSA χρειάζεται περισσότερα από ένα σημεία στην ίδια συντεταγμένη για να χαρακτηριστεί, θα κάνουμε λόγο για ένα υβριδικό FSA. Αυτό είναι ασυνήθιστο στη βιβλιογραφία, παρόλο που πρόσφατες έρευνες φαίνεται να δείχνουν προς αυτή την κατεύθυνση.

Από την άλλη πλευρά, είναι εφικτό να *συνδυάσουμε* διάφορες μετρήσεις αξιολόγησης σε ένα μόνο FSA. Επιπλέον, μια μέτρηση θα μπορούσε να θεωρηθεί σα να ανήκει σε περισσότερες από μια κατηγορίες (π.χ. οι μετρήσεις απόκλισης, εξάρτησης και πληροφορίας είναι πολύ συσχετισμένες). Υπό αυτό το πέπλο, η κατηγοριοποίηση στον κάθετο άξονα θα πρέπει να θεωρείται ως εννοιολογική (δηλαδή το τι μετράται είναι πιο σημαντικό από το πως).

Μη-κλασσικοί αλγόριθμοι (στο πλαίσιο της επιλογής χαρακτηριστικών), μπορούν επίσης να βρεθούν σε αυτόν τον χώρο. Για παράδειγμα, οι εξελικτικές προσεγγίσεις με καταλληλότητα *J* αντιστοιχούν σε <*τυχαίο,τυχαίο,οποιοσδήποτε*> και τα τεχνητά νευρωνικά δίκτυα σε <*σειριακό,βαρύνων,ακρίβεια*>. Μια ευρεία βιβλιογραφική επανάληψη των FSA φαίνεται στον πίνακα 1.

<u>Πίνακας 3.1</u>: Είδη FSA και συντεταγμένες τους στον χώρο χαρακτηριστικών. Πιο συγκεκριμένα ισχύει:
Εκθ.=εκθετικός,Σ=σειριακός,Τ=τυχαίος,Ο=όπισθεν,Ε=εμπρόσθιος,ΣΥΝ.=σύνθετος,Β=βαρύνων

Algorithm	Οργάνωση Αναζήτησης	Γενιά Διαδόχων	J (Συνάρτηση Αξιολόγησης)
ABB	Εκθ.	Ο	Συνέπεια
B&B	Εκθ.	Ο	Συνέπεια
BDS	Σ	Ε/Ο	Ακρίβεια
BEAM	Εκθ.	Ε	οποιοσδήποτε
BFF	Εκθ.	Ε	Απόσταση
BOBRO	Εκθ.	Ο	Απόσταση
BSE	Σ	Ε/Ο	Πληροφορία/Ακρίβεια
CARDIE	Εκθ.	Ε	Πληροφορία/Ακρίβεια
CFS	Εκθ.	Ε	Εξάρτηση
DTM	Σ	Ο	Πληροφορία
Εκθ.	Εκθ.	Ε	Συνέπεια
GA	Τ	Τ	Ακρίβεια
K2-AS	Σ	Ε	Πιθανότητα/Ακρίβεια
KOLLER	Σ	Ο	Πληροφορία
LVF	Τ	Τ	Συνέπεια
LVI	Τ	Τ	Συνέπεια
LVW	Τ	Τ	Ακρίβεια
MDLM	Εκθ.	Ο	Πληροφορία
MIFES-1	Εκθ.	Ο	Συνέπεια
OBLIVION	Εκθ.	Ο	Απόσταση/Ακρίβεια
POE-ACC	Σ	Ε	Εξάρτηση
PQSS	Σ	ΣΥΝ.	Ακρίβεια
PRESET	Σ	Ο	Εξάρτηση
QBB	Τ/Εκθ.	Τ/Ο	Συνέπεια
RACE	Σ	Ε/Ο	Πιθανότητα/Ακρίβεια
RC	Σ	Ο	Ακρίβεια
RELIEF	Τ	Ο	Απόσταση
RGSS	Τ	Ε/Ο	Ακρίβεια
RMHC-PF1	Τ	Τ	Απόσταση/Ακρίβεια
SA	Τ	Τ	Ακρίβεια
SBG	Σ	Ο	οποιοσδήποτε
SBS	Σ	Ο	Απόσταση
SBS-W	Σ	Ο	Ακρίβεια
SBS-SLASH	Σ	Ο	Ακρίβεια
SCHLIMMER	Εκθ.	Ε	Συνέπεια
SEGEN	Σ	Ε	Απόσταση

SFG	Σ	Ε	οποιοσδήποτε
Εκθ.	Εκθ.	ΣΥΝ.	οποιοσδήποτε
SFBS	Εκθ.	ΣΥΝ.	οποιοσδήποτε
WINNOW	Σ	Ο	Συνέπεια
W-SBG	Σ	Ο	Ακρίβεια
W-SFG	Σ	Ε	Ακρίβεια

3.7 Περιγραφή των βασικών FSA

Σε αυτή την παράγραφο αρκετοί από τους πιο δημοφιλείς FSA στον τομέα της μηχανικής μάθησης περιγράφονται και σχολιάζονται. Για τη συνέχεια ας υποθέσουμε ξανά ότι η τιμή της συνάρτησης αξιολόγησης J θέλουμε να μεγιστοποιηθεί. Ας παρατηρήσουμε επίσης και την αναφορά των συντεταγμένων του κάθε αλγορίθμου στο χώρο επιλογής χαρακτηριστικών, αμέσως μετά το όνομα του αλγορίθμου.

3.7.1 LVF αλγόριθμος

LVF (Las Vegas Filter) (<τυχαίος,τυχαίος,οποιοσδήποτε>) επαναλαμβανόμενα δημιουργεί τυχαία υποσύνολα χαρακτηριστικών και στη συνέχεια υπολογίζει το μέτρο της αξίας τους. Ήταν αρχικά υλοποιημένο με τη συνέπεια του δείγματος, ως μέτρο αξίας. Ο αλγόριθμος περιγράφεται παρακάτω:

Είσοδος:

μέγιστο-η μέγιστη τιμή των επαναλήψεων

J- μέτρηση εκτίμησης

S(X)-ένα δείγμα S περιγραφόμενο από το *X*, $|X|=n$

Έξοδος:

L- όλες οι σχετικές λύσεις που έχουν βρεθεί

L :=[] */Αποθηκεύει τα ισοδύναμα καλά σύνολα

Best :=*X* */Αρχικοποίησε την καλύτερη λύση

J_0 := J(*S(X)*) */Η μικρότερη επιτρεπτή τιμή για τη J

επανέλαβε max **φορές**

X' := Τυχαίο_Υποσύνολο(*Best*)

Εάν J(*S(X')*)>J_0 **τότε**

Εάν $|X'|<|Best|$ **τότε**

Best := *X'*

L := [*X'*] */Το L επαναρχικοποιείται

τότε εάν $|X'|=|Best|$ **τότε**

L := πρόσθεσε(*L,X'*)
τέλος
τέλος
τέλος
τέλος

LVW (Las Vegas Wrapper) είναι ένας αναδιπλούμενος αλγόριθμος για τη δημιουργία υποψήφιων υποσυνόλων και χρησιμοποιεί την ακρίβεια του επαγωγέα ως μέτρο της αξίας.

3.7.2 LVI αλγόριθμος

LVI (Las Vegas Incremental) (<τυχαίος,τυχαίος,συνεπής>) είναι βασισμένος στο ότι δε χρειάζεται να χρησιμοποιηθεί ολόκληρο το δείγμα *S* για να εκτιμηθεί η αξία *J*. Ο αλγόριθμος ξεκινά από ένα τμήμα S_o του *S*: εάν ο LVF βρει μια ικανοποιητική λύση στο S_o τότε ο LVI σταματά. Ειδάλλως το σύνολο από δείγματα $S\backslash S_o$ που κάνει το S_o ασυνεπές προστίθεται σε αυτό, αυτό το νέο τμήμα περνά στον LVF και η διαδικασία επαναλαμβάνεται. Το κριτήριο της αξίας μπορεί να είναι οποιοδήποτε. Ο αλγόριθμος περιγράφεται παρακάτω:

Είσοδος:
μέγιστο-ο μέγιστος αριθμός επαναλήψεων
J-το κριτήριο αξιολόγησης
S(X)-ένα δείγμα *S* που περιγράφεται από *X*, $|X|=n$
p-αρχικό ποσοστό
Έξοδος:
X'-η προτεινόμενη λύση

S_0 := τμήμα(*S,p*) */Αρχικό τμήμα
S_1 := S/S_0 */Σύνολο ελέγχου
J_0 := J(*S(X)*) */Ελάχιστη επιτρεπόμενη τιμή για το J
επανέλαβε για πάντα
X':= LVF(*max,J,*$S_0(X)$)
εάν J(*S(X')*)≥J_o **τότε σταμάτα**
αλλιώς
C := {στοιχεία του S1 με χαμηλή προσφορά στη J χρησιμοποιώντας το X'};
$S_0 := S_0 \cup C$
$S_1 := S_1 \backslash C$
τέλος
τέλος

Ενστικτωδώς, το τμήμα δεν μπορεί να είναι ούτε πολύ μεγάλο ούτε πολύ μικρό. Αν είναι πολύ μικρό, μετά την πρώτη επανάληψη πολλές ασυνέπειες θα προκύψουν και θα προστεθούν στο υπάρχον υποσύνολο του δείγματος, το οποίο θα καταλήξει να είναι παρόμοιο με το *S*. Αν είναι πολύ μεγάλο, τα κέρδη από την οικονομία σε υπολογισμούς θα είναι μικρά. Οι συγγραφείς προτείνουν p=10% ή μια τιμή ανάλογη με τον αριθμό των χαρακτηριστικών.

3.7.3 Αλγόριθμος Ανακούφισης (Relief)

Ο αλγόριθμος ανακούφισης (<τυχαίος,βαρύνων,απόσταση>) επιλέγει τυχαία ένα στιγμιότυπο *A* του *S* και αποφασίζει τη *γειτονική επιτυχία* και τη *γειτονική αποτυχία* σε σχέση με το *S*. Το πρώτο είναι το κοντινότερο στιγμιότυπο στο *A* ανάμεσα από όλα τα στιγμιότυπα της ίδιας κατηγορίας με αυτή του *A.* Το δεύτερο είναι κοντινότερο στιγμιότυπο στο *A* ανάμεσα προερχόμενο από διαφορετική κατηγορία. Η βασική ιδέα είναι ότι ένα χαρακτηριστικό είναι πιο σχετικό με το *A* όσο περισσότερο διαχωρίζει το *A* από την κοντινή του αποτυχία, και όσο λιγότερο διαχωρίζει το *A* από την κοντινή του επιτυχία. Το αποτέλεσμα είναι μια έκδοση με βάρη του αρχικού συνόλου χαρακτηριστικών. Ο βασικός αλγόριθμος περιγράφεται παρακάτω.

Μια βελτιωμένη έκδοση προτείνεται (RELIEF-F) όπου τα k πιο όμοια στιγμιότυπα επιλέγονται (που ανήκουν στην ίδια ή διαφορετική κατηγορία, αντίστοιχα) και οι μέσοι όροι τους υπολογίζονται.

Είσοδος:

p-ποσοστό δείγματος

d-μέτρηση απόστασης

S(X)-ένα δείγμα *S* που περιγράφεται από *X*, $|X|=n$

Έξοδος:

W-πίνακας βαρών των χαρακτηριστικών

Αρχικοποίησε *W[]* στο μηδέν

κάνε $p|S|$ **φορές**

A :=Τυχαίο_Στιγμιότυπο(*S*);

A_{nh} := Κοντινή_Επιτυχία(*A,S*);

A_{nm} := Κοντινή_Αποτυχία(*A,S*);

για κάθε $i \in [1, \dots, n]$ **κάνε**

$W[i] := W[i]+d_i(A,A_{nm})-d_i(A,A_{nh})$

τέλος

τέλος

3.7.4 SFG/SBG αλγόριθμοι

SFG (Sequential Forward Generation) (<σειριακός, εμπρόσθιος, οποιοσδήποτε>) επαναληπτικά προσθέτει χαρακτηριστικά σε ένα αρχικό υποσύνολο, με τέτοιο τρόπο ώστε να βελτιώνει μια δεδομένη τιμή J λαμβάνοντας υπ' όψιν αυτά τα χαρακτηριστικά που ήδη υπάρχουν στη λύση. Επιπλέον, μια ταξινομημένη λίστα μπορεί επίσης να ληφθεί ως αποτέλεσμα. Ο SBG (Sequential Backward Generation)(<σειριακός,οπίσθιος,οποιοσδήποτε>) είναι το όπισθεν ανάλογο. Οι αλγόριθμοι περιγράφονται μαζί παρακάτω:

Είσοδος:

S(X)-ένα δείγμα *S* που περιγράφεται από *X*, $|X|=n$

J- το κριτήριο αξιολόγησης

Έξοδος:

X'-η προτεινόμενη λύση

X':=∅ //εμπρόσθιος

X':=*X* //όπισθεν

επανέλαβε

x' :=*argmax{J(S(X'U{x}))|*$x \in X \backslash X'$*}*//εμπρόσθιος

x' :=*argmax{J(S(X'\{x}))|* $x \in X'$*}*//όπισθεν

X' :=*X'*∪*{x'}* //εμπρόσθιος

X' :=*X'**{x'}* //όπισθεν

μέχρι καμία βελτίωση στη J στα τελευταία *j* βήματα

ή *X'*=*X* //εμπρόσθιος

ή *X'*=∅ //όπισθεν

Οι αλγόριθμοι W-SFG και W-SBG (όπου w συμβολίζει το wrapper) χρησιμοποιούν την ακρίβεια ενός εξωτερικού κατηγοριοποιητή/επαγωγέα ως εκτιμητή της αξίας.

3.7.5 SFFS αλγόριθμος

SFFS (Sequential Floating Forward Search)(<εκθετικός,σύνθετος,οποιοσδήποτε>) είναι ένας αλγόριθμος λογαριθμικού κόστους που λειτουργεί με σειριακή χροιά. Σε κάθε βήμα επιλογής ο SFFS κάνει ένα εμπρόσθιο βήμα ακολουθούμενο από ένα μεταβλητό αριθμό (πιθανόν και μηδέν) οπίσθιων βημάτων. Επί της ουσίας, ένα χαρακτηριστικό προστίθεται αρχικά χωρίς να ικανοποιεί κάποιο κριτήριο και στη συνέχεια χαρακτηριστικά απομακρύνονται όσο τα υποσύνολα που δημιουργούνται είναι τα καλύτερα για το μέγεθός τους. Το οπίσθιο ανάλογο

κάνει ένα πίσω βήμα ακολουθούμενο από ένα μεταβλητό αριθμό (πιθανόν και μηδέν) εμπρόσθιων βημάτων.

Είσοδος:

S(X)-ένα δείγμα *S* που περιγράφεται από *X*, |*X*|=*n*

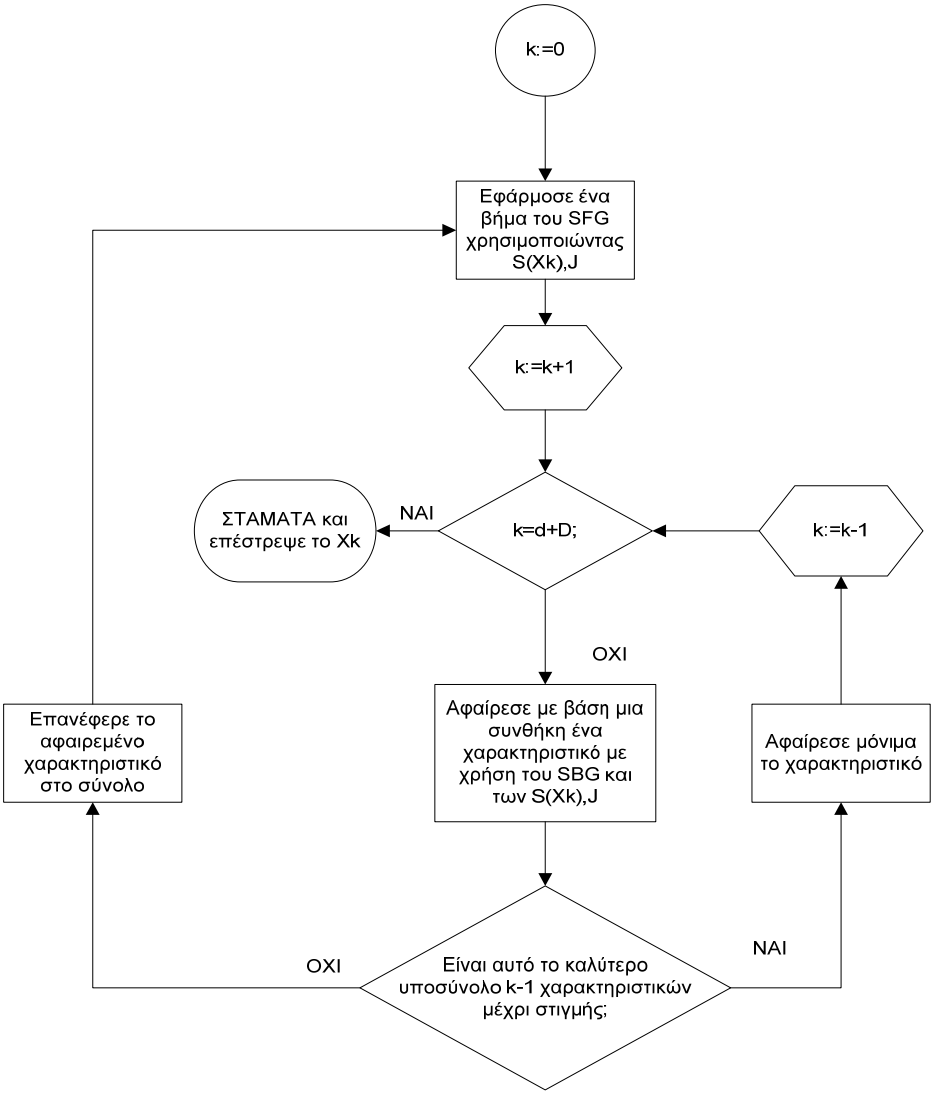
J- το κριτήριο αξιολόγησης

d-επιθυμητό μέγεθος της λύσης

D-μέγιστη επιτρεπόμενη απόκλιση σε σχέση με το *d*

Έξοδος:

λύση μεγέθους *d±D*



3.7.6 Αλγόριθμος Συγκέντρωσης /Εξαντλητικός (Focus)

Ο βασικός αλγόριθμος συγκέντρωσης (<εκθετικός,εμπρόσθιος,συνεπής>) ξεκινά την εκτίμηση κάθε υποσυνόλου ενός χαρακτηριστικού, ύστερα όλων των υποσυνόλων δύο χαρακτηριστικών και πάει λέγοντας. Σταματά όταν μια ικανοποιητικά συνεπής λύση έχει βρεθεί. Ο βασικός αλγόριθμος περιγράφεται παρακάτω.

Είσοδος:

S(X)-ένα δείγμα *S* που περιγράφεται από *X*, $|X|=n$

J- το κριτήριο αξιολόγησης

J_0- η ελάχιστη επιτρεπόμενη τιμή της J

Έξοδος:

X'-η προτεινόμενη λύση

για $i \in [1, \dots, n]$ **κάνε**

για κάθε $X' \subset X$, με $|X'|=i$ **κάνε**

εάν $J(S(X'))\geq J_0$ **τότε σταμάτα**

τέλος

τέλος

Ο αλγόριθμος συγκέντρωσης-2 (FOCUS-2) παρουσιάζει την έννοια της *σύγκρουσης* μεταξύ θετικών και αρνητικών παραδειγμάτων για να κλαδέψει την έρευνα.

3.7.7 B&B αλγόριθμος

B&B (BRANCH & BOUND) (<εκθετικός,οπίσθιος,οποιαδήποτε μονοτονική>) είναι ένας βέλτιστος αλγόριθμος έρευνας. Δεδομένου ενός κατωφλίου *β* (καθορισμένου από τον χρήστη), η έρευνα σταματά σε κάθε βρόγχο η αξία του οποίου είναι κάτω από το *β,* ώστε οι διαφεύγοντες βρόγχοι να κλαδεύονται.

ABB (Automatic Branch & Bound)(<εκθετικός,οπίσθιος,οποιοσδήποτε μονότονος>) είναι μια εκδοχή του B&B στην οποία το κατώφλι τίθεται αυτόματα. Ο αλγόριθμος περιγράφεται παρακάτω:

Είσοδος:

S(X)-ένα δείγμα *S* που περιγράφεται από *X*,$|X|=n$

J- το κριτήριο αξιολόγησης

Έξοδος:

L-όλες οι προτεινόμενες λύσεις

Διαδικασία ABB (S(X)): δείγμα; **μεταβλητή** L': λίστα του συνόλου)

για κάθε *x* στο *X* **κάνε**

βάλε_στην_ουρά(*Q,X\{x}*) //αφαίρεσε ένα χαρακτηριστικό τη φορά

τέλος

όσο όχι άδεια(*Q*) **κάνε**

X':= βγάλε_από_την_ουρά(Q) //X' είναι αποδεκτό εάν δεν είναι υποσύνολο μιας περιορισμένης κατάστασης

εάν αποδεκτό(*X'*) **και** J(*S(X')*)≥J_o **τότε**

L':=πρόσθεσε(L',X');

ABB(*S(X'),L'*);

τέλος

τέλος

τέλος

ξεκίνα

Q:=0

L':=[X]

J_0:=J(S(X))

ABB (*S(X),L'*)

k:= το μικρότερο μέγεθος υποσυνόλου στο L'

L:= το σύνολο των στοιχείων του L' μεγέθους κ

τέλος

3.7.8 QBB αλγόριθμος

QBB (Quick Branch & Bound) (<τυχαίος/-εκθετικός,τυχαίος/όπισθεν, συνέπεια>) είναι ένας υβριδικός αλγόριθμος που συνίσταται από τον LVF και τον ABB. Η βασική ιδέα είναι να χρησιμοποιηθεί ο LVF για να βρεθούν μερικά καλά σημεία για τον ABB. Είναι αναμενόμενο ότι ο ABB μπορεί να εξερευνήσει τον εναπομείναντα χώρο έρευνας ικανοποιητικά. Ο αλγόριθμος περιγράφεται παρακάτω. Οι συγγραφείς παρατήρησαν ότι ο QBB μπορεί να είναι, εν γένει, πιο αποδοτικός από τον LVF, FOCUS και ABB σε ότι αφορά στο μέσο χρόνο εκτέλεσης και στις επιλεγμένες λύσεις.

Είσοδος:

μέγιστο-ο μέγιστος αριθμός επαναλήψεων
S(X)-ένα δείγμα *S* που περιγράφεται από *X*, $|X|=n$
J- το κριτήριο αξιολόγησης

Έξοδος:

L-όλες οι προτεινόμενες λύσεις

L_ABB := []
L_LVF := LVF(*μέγιστο,J,S(X)*)
για κάθε $X' \in L_LVF$ **κάνε**
L_ABB := ένωσε(*L_ABB,ABB(S(X'),J*))
τέλος
k :=το μικρότερο μέγεθος ενός υποσυνόλου στο *L_ABB*
L := το σύνολο των στοιχείων του *L_ABB* μεγέθους k

3.8 Αλγόριθμος SFFS για την επιλογή χαρακτηριστικών

Για τη συγκεκριμένη εργασία από όλους τους προαναφερθέντες αλγορίθμους, χρησιμοποιήσαμε τον αλγόριθμο SFFS. Ο αλγόριθμος αυτός είναι κατά βάση μια μέθοδος έρευνας από κάτω προς τα πάνω (δηλαδή από το 0 προς κάποιο μέγιστο αριθμό χαρακτηριστικών), η οποία συμπεριλαμβάνει νέα χαρακτηριστικά μέσω της εφαρμογής της βασικής SFS διαδικασίας, ξεκινώντας από το υπάρχον σύνολο χαρακτηριστικών, και συνεχίζοντας με μια σειρά από συνεχόμενες αφαιρέσεις των χειρότερων χαρακτηριστικών υπό συνθήκη, υπό την προϋπόθεση πως μπορεί να επέλθει περαιτέρω βελτίωση στα προηγούμενα σύνολα.

Υποθέτουμε ότι *k* χαρακτηριστικά έχουν ήδη επιλεγεί από ολόκληρο το σύνολο μετρήσεων *Y* και συγκροτούν το υποσύνολο X_k με την αντίστοιχη συνάρτηση αξιολόγησης $J(X_k)$. Επιπλέον, οι τιμές $J(X_i)$ όλων των προηγούμενων υποσυνόλων μεγέθους *i= 1, 2, …, k-1,* έχουν διατηρηθεί και είναι γνωστές. Ο αλγόριθμος έχει ως εξής:

- *Βήμα 1° (Εισαγωγή χωρίς συνθήκη)*
  Χρησιμοποιώντας τη βασική SFS διαδικασία, επέλεξε το πιο σημαντικό χαρακτηριστικό x_{k+1}, σε σχέση με το υποσύνολο X_k, δηλαδή το χαρακτηριστικό εκείνο το οποίο όταν προστεθεί στο υπάρχον υποσύνολο βελτιώνει περισσότερο την τιμή της συνάρτηση αξιολόγησης *J*. Πρόσθεσε το στο υπάρχον υποσύνολο. Συνεπώς δημιουργούμε το νέο υποσύνολο,

$$X_{k+1} = X_k + x_{k+1}$$

- *Βήμα 2° (Εξαγωγή υπό συνθήκη)*
  Βρές το λιγότερο σημαντικό χαρακτηριστικό του υποσυνόλου X_{k+1}, δηλαδή το χαρακτηριστικό εκείνο το οποίο όταν αφαιρεθεί από το υποσύνολο δίνει τη μεγαλύτερη τιμή στη *J*, σε σχέση με τις τιμές που επιτυγχάνονται από την αφαίρεση των υπολοίπων χαρακτηριστικών.
  - Αν το χαρακτηριστικό αυτό είναι το x_{k+1}, δηλαδή το χαρακτηριστικό που προστέθηκε κατά το βήμα 1, τότε διατήρησέ το (μην το αφαιρέσεις) και πήγαινε στο βήμα 1 ολοκληρώνοντας μια επανάληψη. Σε αυτή την περίπτωση λοιπόν έχουμε μονιμοποίηση της αύξησης του αριθμού k.
  - Αν είναι οποιοδήποτε άλλο χαρακτηριστικό αφαίρεσέ το, όρισε το νέο υποσύνολο X'_k και πήγαινε στο βήμα 3, εκτός αν ισχύει k=minfeatures, οπότε πήγαινε στο βήμα 1.
- *Βήμα 3° (Συνέχιση της εξαγωγής υπό συνθήκη)*
  Βρές το λιγότερο σημαντικό χαρακτηριστικό στο υποσύνολο X'_k.
  - Αν η τιμή της *J* που επιτυγχάνει είναι μικρότερη ή ίση από την $J(X_{k-1})$ τότε μην κάνεις τίποτα και ολοκλήρωσε πηγαίνοντας στο βήμα 1.
  - Αν η τιμή της *J* που επιτυγχάνει είναι μεγαλύτερη από την $J(X_{k-1})$ τότε αφαίρεσε το στοιχείο, δημιουργώντας έτσι ένα νέο υποσύνολο X'_{k-1}
    - Αν το k=minfeatures ολοκλήρωσε πηγαίνοντας στο βήμα 1, αλλιώς επανέλαβε το βήμα 3

Ο αλγόριθμος αρχικοποιείται θέτοντας k=0 και προσθέτοντας χαρακτηριστικά μέσω της μεθόδου SFS μέχρι k=minfeatures, όπου minfeatures προκαθορισμένο από τον χρήστη. Στη συνέχεια πηγαίνει στο βήμα 1, μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού, συνήθως ένας αριθμός επαναλήψεων.

4. ΕΞΕΛΙΚΤΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ

Εισαγωγή

Στο κεφάλαιο αυτό ασχολούμαστε με τον εξελικτικό υπολογισμό, μια ευρύτερη οικογένεια αλγορίθμων η οποία περιλαμβάνει τους γενετικούς αλγορίθμους, τις εξελικτικές στρατηγικές και το γενετικό προγραμματισμό, καθώς και τις εφαρμογές τους στη μηχανική μάθηση.

Η πρώτη έννοια που πρέπει να προσδιορίσουμε είναι αυτή της ευφυΐας ενός συστήματος. Ως ευφυΐα μπορεί να οριστεί η ικανότητα του συστήματος να προσαρμόζει τη συμπεριφορά του σε ένα συνεχώς μεταλλασσόμενο περιβάλλον. Σύμφωνα με τον Άλαν Τούρινγκ, η μορφή ή η εμφάνιση ενός συστήματος είναι άσχετη με την ευφυΐα του. Ωστόσο, από την καθημερινή μας εμπειρία γνωρίζουμε πως αποδείξεις ευφυούς συμπεριφοράς εύκολα παρατηρούνται στους ανθρώπους. Οι άνθρωποι όμως είναι προϊόντα της εξέλιξης, και συνεπώς με το να μοντελοποιήσουμε τη διαδικασία αυτή, υποθέτουμε ότι μπορούμε να δημιουργήσουμε ευφυή συμπεριφορά.

Ο εξελικτικός υπολογισμός εξομοιώνει την εξέλιξη σε έναν υπολογιστή. Το αποτέλεσμα μιας τέτοιας εξομοίωσης είναι μια σειρά από αλγορίθμους βελτιστοποίησης, βασισμένους συνήθως σε μια σειρά από απλούς κανόνες. Η βελτιστοποίηση, αυξάνει με επαναληπτικό τρόπο την ποιότητα των λύσεων μέχρι να βρεθεί μια βέλτιστη, ή τουλάχιστον αποδεκτή με βάση κάποιο κριτήριο λύση. Αλλά είναι η εξέλιξη πραγματικά ευφυής;

Μπορούμε να θεωρήσουμε πως ο τρόπος που συμπεριφέρεται ένας οργανισμός, βασίζεται σε κάποια επαγωγικά συμπεράσματα για τον τρόπο λειτουργίας του περιβάλλοντός του. Με πιο απλά λόγια, η συμπεριφορά του είναι αποτέλεσμα της προσπάθειας προσαρμογής του στις συνθήκες του περιβάλλοντος του, όπως όμως τις αντιλαμβάνεται αυτός με βάση τις εμπειρίες του και την αντιληπτική του ικανότητα. Η ταυτότητα των στοιχείων που καθορίζουν την ποιότητα του περιβάλλοντος δράσης και ζωής είναι γενικά άγνωστη με αντικειμενικά κριτήρια στον οργανισμό και αυτός καλείται να γενικεύσει τις εμπειρίες του (επαγωγικό κομμάτι) και μέσα από αυτή τη γενίκευση να προσεγγίσει την αντικειμενική πραγματικότητα των χαρακτηριστικών. Στη συνέχεια, με βάση αυτή τη πληροφορία προσαρμόζει τη συμπεριφορά και τον τρόπο λειτουργίας του, ώστε να εκμεταλλεύεται τα πλεονεκτήματα και να ελαχιστοποιεί τα μειονεκτήματα της διαβίωσής του.

Λαμβάνοντας υπ’ όψιν τα παραπάνω κάνουμε τη ρεαλιστική υπόθεση πως, εάν ύστερα από συνεχόμενες γενεές ο οργανισμός επιβιώσει, μπορούμε να πούμε ότι ο οργανισμός είναι ικανός να μάθει να προβλέπει τις αλλαγές στο περιβάλλον του, καθώς έχει αντιληφθεί έστω και ασυνείδητα τον τρόπο λειτουργίας του. Η εξέλιξη είναι μια πολύπλοκη και αργή

διαδικασία από την πλευρά του ανθρώπου, αλλά η εξομοίωση της σε ένα υπολογιστή δε χρειάζεται δισεκατομμύρια χρόνια.

Η εξελικτική προσέγγιση της μηχανικής μάθησης είναι βασισμένη σε υπολογιστικά μοντέλα φυσικής επιλογής και γενετικής. Συνολικά τα ονομάζουμε **εξελικτικό υπολογισμό**, ένας υπερκείμενος όρος που συνδυάζει τους **γενετικούς αλγορίθμους, τις στρατηγικές εξέλιξης και το γενετικό προγραμματισμό.** Όλες αυτές οι τεχνικές εξομοιώνουν την εξέλιξη μέσω της χρήσης των διαδικασιών της επιλογής, μετάλλαξης και αναπαραγωγής, οι οποίοι αποτελούν και τους βασικούς τελεστές των αλγορίθμων.

Εξομοίωση της φυσικής εξέλιξης

Στη 1 Ιουλίου του 1858, ο Δαρβίνος παρουσίασε τη θεωρία του σχετικά με την εξέλιξη στη Linnean Society του Λονδίνου. Αυτή η μέρα σημαδεύει την αρχή μιας επανάστασης στη βιολογία. Η κλασσική θεωρία εξέλιξης του Δαρβίνου, σε συνδυασμό με τη θεωρία της φυσικής επιλογής του Weismann και την έννοια της γενετικής κατά τον Mendel, συγκροτούν τον νέο-Δαρβινισμό.

Ο νέο-Δαρβινισμός βασίζεται σε διαδικασίες αναπαραγωγής, μετάλλαξης, ανταγωνισμού και επιλογής. Η δυνατότητα αναπαραγωγής φαίνεται να είναι ουσιαστικό κομμάτι της ζωής. Η δυνατότητα μετάλλαξης είναι εξ' ίσου εγγυημένη σε κάθε ζωντανό οργανισμό που αναπαράγεται σε ένα διαρκώς μεταβαλλόμενο περιβάλλον. Οι διαδικασίες του ανταγωνισμού και της επιλογής λαμβάνουν χώρα στον φυσικό κόσμο κανονικά, καθώς οι αναπτυσσόμενοι πληθυσμοί διαφορετικών ειδών περιορίζονται από συγκεκριμένο χώρο, στον οποίο καλούνται να επιβιώσουν.

Αν λοιπόν η διαδικασία εξέλιξης πρέπει να εξομοιωθεί σε έναν υπολογιστή, πρέπει να απαντηθεί το ερώτημα, τί είναι αυτό που βελτιστοποιείται από την εξέλιξη στη πραγματική φυσιολογική ζωή; Η εξέλιξη μπορεί να θεωρηθεί ως μια διαδικασία που οδηγεί στη διατήρηση ή αύξηση της ικανότητας ενός πληθυσμού να επιβιώνει και να αναπαράγεται σε ένα συγκεκριμένο περιβάλλον. Αυτή η ικανότητα ονομάζεται **εξελικτική καταλληλότητα (evolutionary fitness)** ή πιο απλά, καταλληλότητα **(fitness)**. Παρόλο που η καταλληλότητα δεν μπορεί να μετρηθεί άμεσα, μπορεί να προσεγγισθεί με βάση την οικολογία και τη λειτουργική μορφολογία του οργανισμού στο περιβάλλον του. Η εξελικτική καταλληλότητα μπορεί να θεωρηθεί και ως μια μέτρηση της ικανότητας του οργανισμού να προβλέπει αλλαγές στο περιβάλλον του. Συνεπώς, η καταλληλότητα, ή η αριθμητική μέτρηση της ικανότητας πρόβλεψης περιβαλλοντικών αλλαγών και ανάλογης προσαρμογής, μπορεί να θεωρηθεί ως η ποιότητα που βελτιστοποιείται στην κανονική ζωή.

Για να οπτικοποιήσουμε την καταλληλότητα, μπορούμε να χρησιμοποιήσουμε την έννοια της **προσαρμοστικής τοπολογίας**. Μπορούμε να αναπαραστήσουμε ένα δοσμένο περιβάλλον

μέσω ενός τοπίου όπου κάθε κορυφή αντιστοιχεί στη βελτιστοποιημένη καταλληλότητα του είδους. Καθώς η εξέλιξη λαμβάνει χώρα, κάθε είδος ενός συγκεκριμένου πληθυσμού ανεβαίνει τις πλαγιές του τοπίου προς τις κορυφές. Το περιβάλλον αλλάζει με το πέρας του χρόνου και συνεπώς αλλάζει η θέση αλλά και το μονοπάτι προσέγγισης των κορυφών. Λόγω αυτού, τα είδη οφείλουν να προσαρμόζουν συνεχώς την πορεία τους, με αποτέλεσμα μόνο τα ικανότερα μπορούν να φθάσουν στις κορυφές.

Η προσαρμοστική τοπολογία είναι μια συνεχής συνάρτηση και εξομοιώνει το γεγονός ότι το περιβάλλον ή η φυσική τοπολογία στον πραγματικό κόσμο δεν είναι στατική. Το σχήμα της τοπολογίας μεταβάλλεται με το χρόνο, και κάθε είδος συνεχώς υπόκειται των νόμων της φυσικής επιλογής. Ο στόχος της εξέλιξης είναι να δημιουργήσει ένα πληθυσμό από άτομα με αυξημένη καταλληλότητα για το τρέχον περιβάλλον. Αλλά πως δημιουργείται ένας πληθυσμός με αυξημένη καταλληλότητα;

Ο Michalewicz προτείνει μια απλή εξήγηση βασισμένη σε ένα πληθυσμό λαγών. Μερικοί λαγοί είναι γρηγορότεροι από άλλους, και μπορούμε να πούμε ότι αυτοί οι λαγοί έχουν μεγαλύτερη καταλληλότητα καθώς έχουν μεγαλύτερη πιθανότητα να αποφεύγουν αλεπούδες, και συνεπώς να επιβιώσουν και να αναπαραχθούν. Προφανώς και μερικοί αργοί λαγοί θα επιβιώσουν. Ως αποτέλεσμα, μερικοί αργοί λαγοί αναπαράγονται με γρήγορους λαγούς, μερικοί γρήγοροι με άλλους γρήγορους και μερικοί αργοί με άλλους αργούς. Με άλλα λόγια, η αναπαραγωγή δημιουργεί ένα μείγμα γονιδίων λαγών. Αν δύο γονείς έχουν καλή καταλληλότητα, υπάρχει σοβαρή πιθανότητα ότι ο συνδυασμός τους θα παράγει έναν απόγονο με ακόμα μεγαλύτερη καταλληλότητα. Σε βάθος χρόνου όλος ο πληθυσμός των λαγών γίνεται γρηγορότερος με σκοπό να αντιμετωπίσει την περιβαλλοντική πρόκληση του κυνηγιού από τις αλεπούδες. Ωστόσο, οι περιβαλλοντικές συνθήκες μπορούν να μεταβληθούν ώστε να ευνοούνται οι χοντροί αλλά έξυπνοι λαγοί. Σε αυτήν την περίπτωση, για να βελτιστοποιηθούν οι πιθανότητες επιβίωσης, η γενετική δομή του πληθυσμού των λαγών θα μεταβληθεί ανάλογα. Από την άλλη πλευρά, οι γρήγοροι και έξυπνοι λαγοί βοηθούν στη γένεση ταχύτερων και γρηγορότερων αλεπούδων, καθώς αυτές προσπαθούν με τη σειρά τους να επιβιώσουν και να αναπαραχθούν. Η φυσική εξέλιξη είναι μια συνεχής και αδιάκοπη διαδικασία.

Περνώντας τώρα στο πρακτικό κομμάτι της εξομοίωσης της εξέλιξης στον Η/Υ, αρκετές διαφορετικές μέθοδοι υλοποίησης του εξελικτικού υπολογισμού είναι πλέον γνωστές. Όλες εξομοιώνουν τη φυσική εξέλιξη γενικά, με το να δημιουργούν ένα τυχαίο αρχικό πληθυσμό από υποκείμενα, να εκτιμούν την καταλληλότητα τους, να δημιουργούν ένα νέο πληθυσμό μέσω της χρήσης γενετικών τελεστών, και να επαναλαμβάνουν αυτή τη διαδικασία αρκετές φορές μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού. Παρόλη τη φαινομενική ομοιότητα τους πάντως, υπάρχουν αρκετοί τρόποι να υλοποιήσουμε τον εξελικτικό

υπολογισμό. Θα ξεκινήσουμε με τους γενετικούς αλγορίθμους, καθώς οι περισσότεροι από τους άλλους εξελικτικούς αλγορίθμους μπορούν να θεωρηθούν ως παραλλαγές των ΓΑ (γενετικών αλγορίθμων).

Εισαγωγή στους Γενετικούς Αλγορίθμους

Στις αρχές του 1970, ο John Holland, ένας από τους ανθρώπους που πρωτοεισήγαγαν την έννοια του εξελικτικού υπολογισμού, παρουσίασε την έννοια των γενετικών αλγορίθμων. Ο στόχος του ήταν να δείξει στους υπολογιστές πως να κάνουν ότι έκανε η φύση. Ως μηχανικός υπολογιστών, ο Holland ασχολήθηκε με αλγορίθμους που επεξεργάζονται σειρές δυαδικών ψηφίων. Θεωρούσε αυτούς τους αλγορίθμους ως μια ασαφή μορφή της φυσικής εξέλιξης. Οι ΓΑ του Holland μπορούν να αναπαρασταθούν από μια σειρά βημάτων που οδηγούν από τον ένα πληθυσμό τεχνητών χρωμοσωμάτων σε ένα νέο. Χρησιμοποιεί τη μέθοδο της φυσικής επιλογής και τεχνικές εμπνευσμένες από τη γενετική, γνωστές ως διασταύρωση και μετάλλαξη. Κάθε χρωμόσωμα αποτελείται από έναν αριθμό γονιδίων, και κάθε γονίδιο αναπαρίσταται με 0 ή 1.

Η φύση έχει την ικανότητα να προσαρμόζεται και να μαθαίνει χωρίς να της γίνονται υποδείξεις. Με άλλα λόγια, η φύση βρίσκει τα καλύτερα χρωμοσώματα τυφλά. Οι ΓΑ κάνουν ακριβώς το ίδιο. Δύο μηχανισμοί συνδέουν ένα ΓΑ με το πρόβλημα που καλείται να λύσει: **η κωδικοποίηση και η αξιολόγηση**.

Στην εργασία του Holland, η κωδικοποίηση επιτελείται με την αναπαράσταση των χρωμοσωμάτων ως αλληλουχίες 0 και 1. Παρόλο που έχουν προταθεί και αρκετές άλλες τεχνικές κωδικοποίησης, καμία δε δουλεύει τέλεια σε όλες τις περιπτώσεις. Γι’ αυτό θα χρησιμοποιήσουμε τις σειρές ψηφίων, καθώς είναι η δημοφιλέστερη.

Μια συνάρτηση αξιολόγησης J χρησιμοποιείται για τη μέτρηση της απόδοσης του χρωμοσώματος, ή της καταλληλότητας για το συγκεκριμένο πρόβλημα που καλείται να λύσει (μια συνάρτηση αξιολόγησης σε ένα ΓΑ παίζει τον ίδιο ρόλο που παίζει το περιβάλλον στη φυσική εξέλιξη). Ο ΓΑ χρησιμοποιεί μια μέτρηση της καταλληλότητας των επιμέρους χρωμοσωμάτων που θα επιτελέσουν την αναπαραγωγή. Καθώς η αναπαραγωγή λαμβάνει χώρα, ο τελεστής διασταύρωσης ανταλλάσει τμήματα δύο απλών χρωμοσωμάτων ώστε να δημιουργήσει τον απόγονο, και ο τελεστής μετάλλαξης μεταβάλλει την τιμή ενός τυχαίου γονιδίου του νέου χρωμοσώματος. Σαν αποτέλεσμα, ύστερα από ένα αριθμό επιτυχημένων αναπαραγωγών, τα λιγότερο κατάλληλα χρωμοσώματα εξαφανίζονται, ενώ αυτά που είναι ικανά να επιβιώσουν σταδιακά κυριαρχούν στον πληθυσμό. Μπορεί να είναι μια απλή προσέγγιση, ωστόσο ακόμα και οι λιγότερο περίτεχνοι μηχανισμοί αναπαραγωγής έχουν την ικανότητα να ανταπεξέρχονται σε απαιτητικά προβλήματα.

Ας δούμε τους αλγορίθμους τώρα με μεγαλύτερη λεπτομέρεια.

4.1 Γενετικοί αλγόριθμοι

Ξεκινάμε με έναν ορισμό: οι γενετικοί αλγόριθμοι είναι μια κατηγορία στοχαστικών ευρετικών αλγορίθμων βασισμένων στη βιολογική εξέλιξη. Δεδομένου ενός σαφώς καθορισμένου προβλήματος προς λύση και μιας δυαδικής σειριακής αναπαράστασης για τις υποψήφιες λύσεις (κωδικοποίηση), ένας βασικός ΓΑ μπορεί να αναπαρασταθεί όπως στο σχήμα 4.1. Ένας ΓΑ ακολουθεί τα εξής βασικά βήματα:

1. Αναπαράστησε την περιοχή μεταβλητών του προβλήματος ως χρωμόσωμα καθορισμένου μεγέθους, επέλεξε το μέγεθος του πληθυσμού των χρωμοσωμάτων N, την πιθανότητα διασταύρωσης p_c και την πιθανότητα μετάλλαξης p_m.

2. Όρισε μια συνάρτηση καταλληλότητας για τη μέτρηση της απόδοσης, του κάθε επιμέρους χρωμοσώματος του χώρου του προβλήματος. Η συνάρτηση καταλληλότητας θεμελιώνει τη βάση πάνω στην οποία θα γίνει η επιλογή των χρωμοσωμάτων που θα ταιριάξουν κατά την αναπαραγωγική διαδικασία.

3. Δημιούργησε με τυχαίο τρόπο έναν αρχικό αριθμό χρωμοσωμάτων μεγέθους N:

   $x_1, x_2, \ldots, x_N$

4. Υπολόγισε την καταλληλότητα κάθε επιμέρους χρωμοσώματος:

   $f(x_1), f(x_2), \ldots, f(x_N)$

5. Διάλεξε ένα ζευγάρι από χρωμοσώματα για να ζευγαρώσουν από τον υπάρχοντα πληθυσμό. Τα γονεϊκά χρωμοσώματα επιλέγονται με βάση μια πιθανότητα σχετική με την καταλληλότητα τους. Τα πολύ κατάλληλα χρωμοσώματα έχουν μεγαλύτερη πιθανότητα να επιλεγούν, σε σχέση με λιγότερο κατάλληλα.

6. Δημιούργησε ένα ζευγάρι από χρωμοσώματα απογόνους με την εφαρμογή γενετικών τελεστών διασταύρωσης και μετάλλαξης.

7. Τοποθέτησε τα χρωμοσώματα απογόνους στο νέο πληθυσμό.

8. Επανέλαβε το βήμα 5 μέχρι το μέγεθος του νέου πληθυσμού να γίνει ίσο με το μέγεθος του αρχικού πληθυσμού N.

9. Αντικατέστησε το αρχικό (γονεϊκό) χρωμοσωματικό πληθυσμό με το νέο (απόγονο) πληθυσμό.

10. Πήγαινε στο βήμα 4, και επανέλαβε τη διαδικασία μέχρι να ικανοποιηθεί το κριτήριο τερματισμού.

Όπως βλέπουμε, ένας ΓΑ αναπαριστά, εκτός των άλλων, και μια επαναληπτική διαδικασία. Κάθε επανάληψη ονομάζεται **γενεά**. Ένας τυπικός αριθμός γενεών για ένα απλό ΓΑ μπορεί να κυμαίνεται από 50 έως πάνω από 500. Το σύνολο των γενεών ονομάζεται **τρέξιμο**. Στο τέλος του αναμένεται να βρούμε ένα ή περισσότερα κατάλληλα χρωμοσώματα, ανάλογα με τις επιθυμίες μας.

Υπάρχουν τυπικά κριτήρια τερματισμού που χρησιμοποιούνται στους ΓΑ;

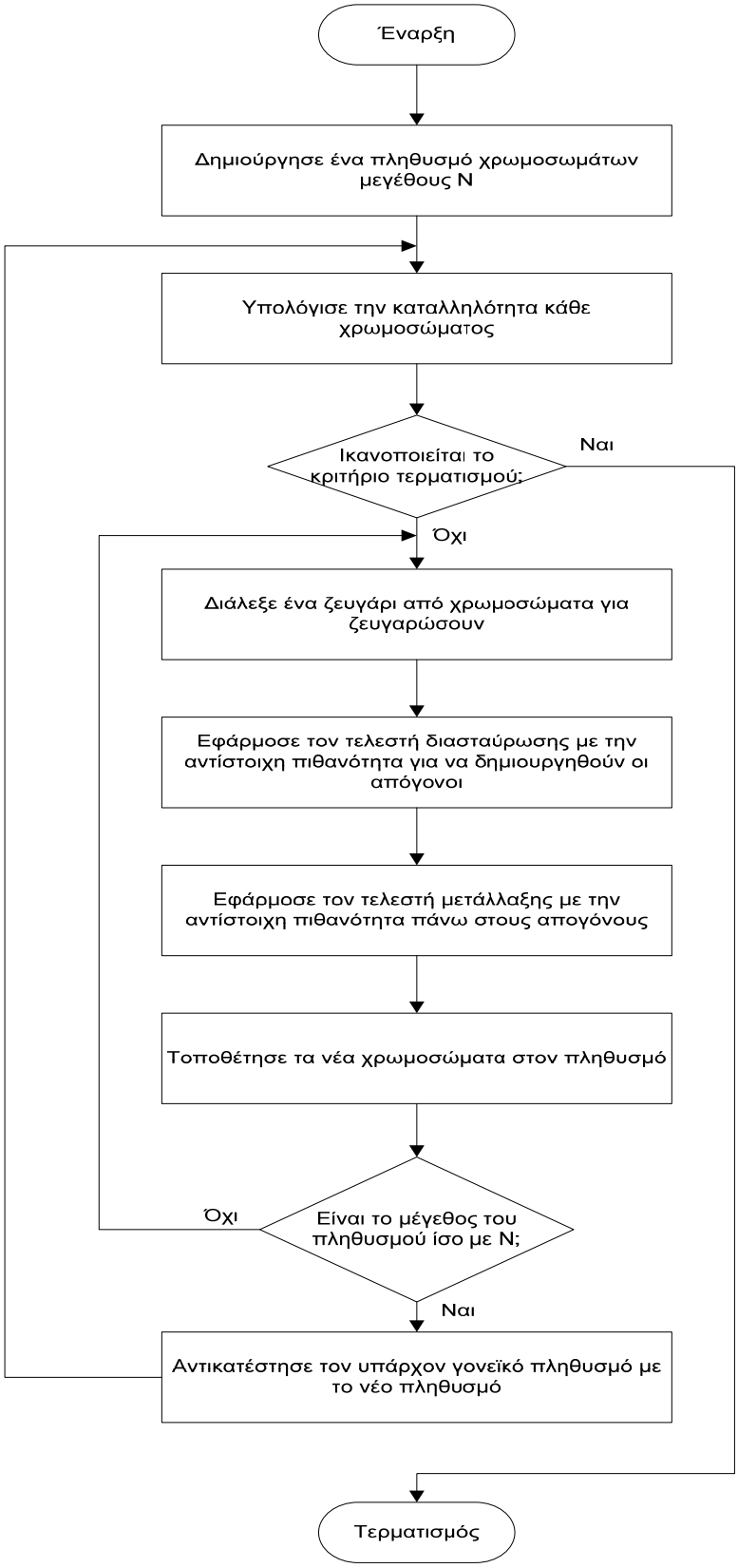
Επειδή οι ΓΑ χρησιμοποιούν μια στοχαστική μέθοδο έρευνας, η καταλληλότητα ενός πληθυσμού μπορεί να παραμείνει σταθερή για αρκετές γενιές προτού εμφανιστεί ένα ανώτερο χρωμόσωμα. Αυτό καθιστά τα τυπικά κριτήρια τερματισμού προβληματικά. Μια κοινή πρακτική είναι να τερματίζεται ο ΓΑ ύστερα από ένα προκαθορισμένο αριθμό γενεών και τότε να ελέγχονται τα καλύτερα χρωμοσώματα του πληθυσμού. Εάν καμία ικανοποιητική λύση δεν έχει βρεθεί, ο ΓΑ επανεκινά.

Ένα απλό παράδειγμα θα μας βοηθήσει να καταλάβουμε πως λειτουργεί ένας ΓΑ. Έστω ότι χρειάζεται να υπολογίσουμε τη μέγιστη τιμή της συνάρτησης $(15x - x^2)$, όπου η μεταβλητή x κυμαίνεται μεταξύ 0 και 15. Χάριν απλότητος, μπορούμε να υποθέσουμε ότι το x λαμβάνει μόνο ακέραιες τιμές. Συνεπώς, τα χρωμοσώματα που θα χρησιμοποιηθούν στο πρόβλημα συνίστανται από 4 γονίδια.

Έστω ότι το μέγεθος του πληθυσμού είναι N=6 χρωμοσώματα, η πιθανότητα διασταύρωσης είναι 0,7 και η πιθανότητα μετάλλαξης 0,001 (οι οποίες αποτελούν και τυπικές τιμές για τους ΓΑ). Η συνάρτηση αξιολόγησης στο παράδειγμά μας ορίζεται ως:

$$f(x) = 15x - x^2$$

η συνάρτηση δηλαδή που θέλουμε να μεγιστοποιήσουμε.



Σχήμα 4.1: Διάγραμμα ροής απλού Γενετικού Αλγορίθμου

Ο ΓΑ δημιουργεί έναν αρχικό πληθυσμό χρωμοσωμάτων με το να γεμίζει τον αρχικό πληθυσμό των 6 τετραψήφιων σειρών με τυχαίους άσσους και μηδενικά. Ένα τυπικό πρακτικό πρόβλημα θα είχε ένα πληθυσμό χιλιάδων χρωμοσωμάτων.

Το επόμενο βήμα είναι ο υπολογισμός της καταλληλότητος κάθε επιμέρους χρωμοσώματος. Μπορούμε να αυξήσουμε τη συνολική καταλληλότητα του πληθυσμού με το να ασκήσουμε πάνω τους 3 βασικούς γενετικούς τελεστές, την επιλογή, τη διασταύρωση και τη μετάλλαξη.

Στη φύση, μόνο τα δυνατότερα είδη μπορούν να επιβιώσουν, να αναπαραχθούν και κατ' επέκταση να περάσουν τα γονίδιά τους στην επόμενη γενιά. Οι ΓΑ χρησιμοποιούν μια παρόμοια προσέγγιση με τη μόνη διαφορά πως ο αριθμός των χρωμοσωμάτων του πληθυσμού (το μέγεθος του πληθυσμού), παραμένει αμετάβλητο από γενιά σε γενιά.

4.1.1 Βασικοί Τελεστές

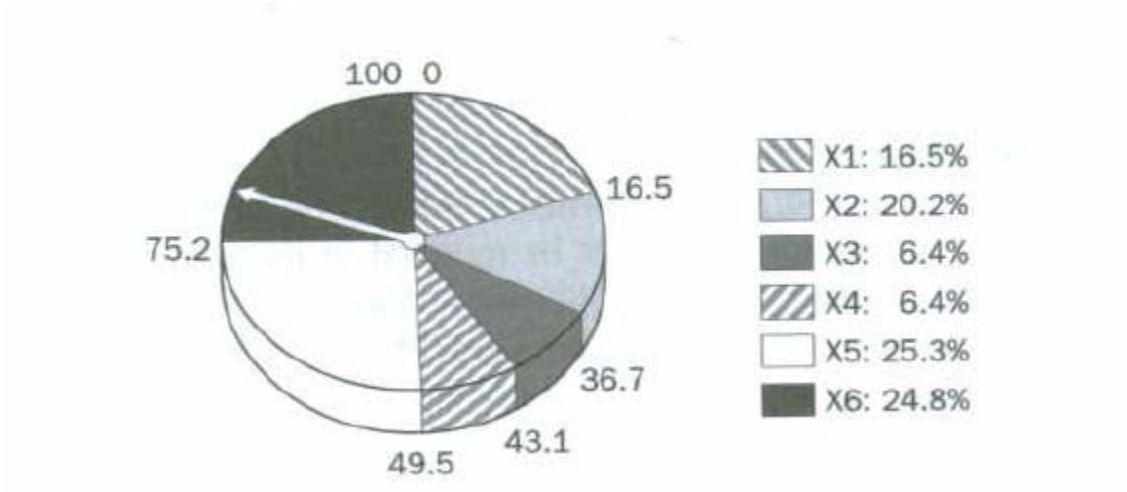
Επιλογή

Το βασικό μέλημα του τμήματος επιλογής είναι σε πρώτη φάση να κάνει τις σωστές επιλογές χρωμοσωμάτων ώστε με την πάροδο τον γενεών η μέση καταλληλότητα του πληθυσμού να αυξάνεται. Πέρα από αυτό όμως, θα πρέπει να φροντίζει το μέγεθος του πληθυσμού να διατηρείται σταθερό στην προκαθορισμένη τιμή Ν.

Σε κάθε χρωμόσωμα αντιστοιχεί μια τιμή καταλληλότητας, με βάση την ποιότητά του. Η τιμή αυτή διαιρεμένη με το συνολικό άθροισμα των τιμών καταλληλότητας όλου του πληθυσμού, μας δίνει το ποσοστό καταλληλότητας του κάθε χρωμοσώματος. Το ποσοστό αυτό, ταυτίζεται με την πιθανότητα επιλογής του χρωμοσώματος για να ζευγαρώσει. Προφανώς, τα πιο κατάλληλα χρωμοσώματα είναι πιο πιθανό να επιλεγούν, καθώς σε αυτά αντιστοιχεί μεγαλύτερο ποσοστό. Με αυτό τον τρόπο διασφαλίζουμε πως από γενιά σε γενιά θα υπάρχει βελτίωση του πληθυσμού, καθώς όταν τα καλύτερα χρωμοσώματα ζευγαρώνουν είναι πιθανότερο να δώσουν ποιοτικότερους απογόνους.

Πάνω σε αυτή τη λογική στηρίζεται μία από τις πιο συνηθισμένες τεχνικές επιλογής χρωμοσωμάτων η επιλογή ρουλέτας. Το σχήμα που ακολουθεί παρουσιάζει τη ρουλέτα του παραδείγματος που θα δώσουμε. Είναι προφανές, πως σε κάθε χρωμόσωμα αντιστοιχεί ένα κομμάτι της κυκλικής ρουλέτας. Η περιοχή του κομματιού μέσα στον κύκλο είναι ίση με την καταλληλότητα του χρωμοσώματος. Για παράδειγμα, τα χρωμοσώματα Χ5 και Χ6 (τα πιο κατάλληλα χρωμοσώματα) καταλαμβάνουν μεγαλύτερες περιοχές, ενώ τα χρωμοσώματα Χ3 και Χ4 (τα λιγότερο κατάλληλα) μικρότερες. Για την επιλογή ενός χρωμοσώματος για ζευγάρωμα, ένας τυχαίος αριθμός δημιουργείται στο διάστημα μεταξύ [0,100], και το χρωμόσωμα του οποίου το κομμάτι αντιστοιχεί στον τυχαίο αριθμό επιλέγεται. Είναι σαν μια ρουλέτα όπου κάθε χρωμόσωμα καταλαμβάνει ένα τμήμα στον τροχό ανάλογο της

καταλληλότητας του. Η ρουλέτα γυρνά, και όταν το βέλος ηρεμεί σε κάποια περιοχή, το αντίστοιχο (της περιοχής) χρωμόσωμα επιλέγεται.



Σχήμα 4.2: Επιλογή τύπου ρουλέτας. Παρατηρούμε ότι στα καλύτερα χρωμοσώματα αντιστοιχούν μεγαλύτερες περιοχές και συνεπώς μεγαλύτερες πιθανότητες επιλογής.

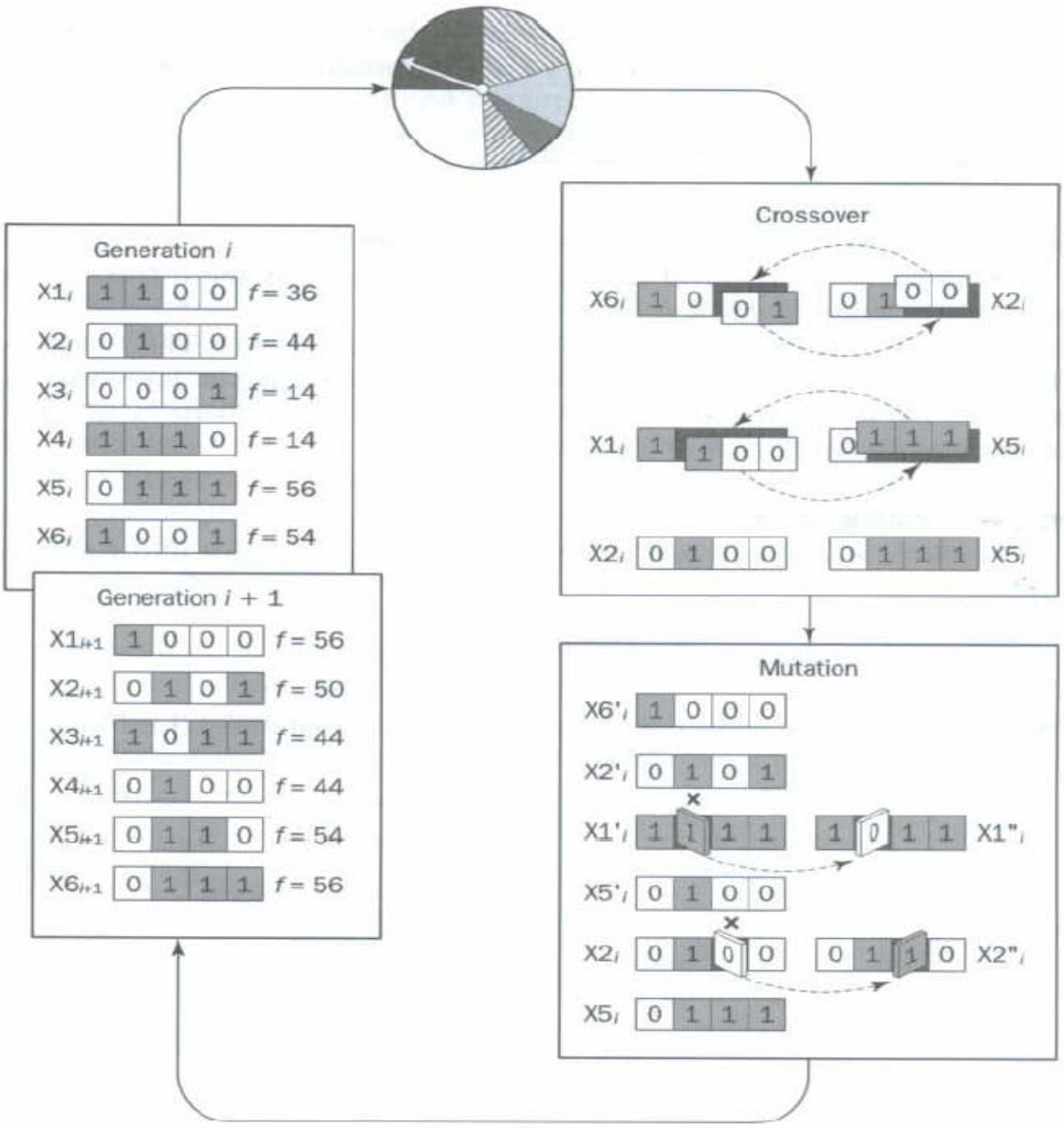
Στο παράδειγμα μας, έχουμε έναν αρχικό πληθυσμό 6 χρωμοσωμάτων. Έτσι, για να διατηρήσουμε τον ίδιο πληθυσμό και στην επόμενη γενεά, η ρουλέτα θα πρέπει να γυρίσει 6 φορές. Οι πρώτες δύο περιστροφές μπορεί να επιλέξουν τα χρωμοσώματα Χ6 και Χ2 για να γίνουν γονείς, το δεύτερο ζευγάρι περιστροφών μπορεί να επιλέξει τα χρωμοσώματα Χ1 και Χ5, και οι δύο τελευταίες περιστροφές τα χρωμοσώματα Χ2 και Χ5. Όταν ένα ζευγάρι από χρωμοσώματα επιλεγεί, ο τελεστής διασταύρωσης εφαρμόζεται σε αυτό για να παράξει τους απογόνους.

Διασταύρωση

Ο τελεστής διασταύρωσης λειτουργεί ως εξής· επιλέγει τυχαία ένα σημείο διασταύρωσης όπου τα δύο γονεϊκά χρωμοσώματα «σπάνε», και τότε ανταλλάσει τα κομμάτια των χρωμοσωμάτων από εκείνο το σημείο και έπειτα, με αποτέλεσμα τη δημιουργία δύο νέων χρωμοσωμάτων. Για παράδειγμα, τα χρωμοσώματα Χ6 και Χ2 μπορεί να διασταυρωθούν μετά το δεύτερο γονίδιο ώστε να παράγουν τους δύο απογόνους. Το σημείο διασταύρωσης δεν είναι απαραίτητο να είναι μοναδικό, αλλά μπορούν τα χρωμοσώματα να ανταλλάξουν περισσότερα κομμάτια τους στην προσπάθεια τους να δημιουργήσουν ποιοτικούς απογόνους.

Αν η τυχαία διαδικασία επιλογής του γονιδίου διασταύρωσης δεν επιλέξει κάποιο πραγματικό σημείο του χρωμοσώματος, τότε δεν πραγματοποιείται διασταύρωση αλλά λαμβάνει χώρα η κλωνοποίηση των χρωμοσωμάτων, και οι απόγονοι είναι ακριβή αντίγραφα κάθε γονέα. Για παράδειγμα, τα γονεϊκά χρωμοσώματα Χ2 και Χ5 μπορεί να μη διασταυρωθούν, οπότε οι

απόγονοι τους είναι ακριβή αντίγραφα τους. Μια τιμή γύρω στο 0,7 για την πιθανότητα διασταύρωσης συνήθως παράγει καλά αποτελέσματα. Οι διάφορες περιπτώσεις που μπορεί να εμφανιστούν κατά τη διασταύρωση φαίνονται στο παρακάτω σχήμα:



Σχήμα 4.3: Εφαρμογή των γενετικών τελεστών σε τυχαία χρωμοσώματα

Μετάλλαξη

Η μετάλλαξη, η οποία είναι σπάνια στη φύση, αναπαριστά μια αναίτια αλλαγή στα γονίδια (δηλαδή μια αλλαγή που γίνεται χωρίς κάποιο σκοπό, απλά συμβαίνει). Μπορεί να οδηγήσει σε μια σημαντική βελτίωση της καταλληλότητας, αλλά πιο συχνά έχει μάλλον καταστροφικά αποτελέσματα. Άρα γιατί να χρησιμοποιήσουμε τη μετάλλαξη ούτως ή άλλως;

Ο Holland παρουσίασε τη μετάλλαξη σαν ένα παρασκηνιακό τελεστή. Ο ρόλος του είναι να παρέχει μια εγγύηση πως ο αλγόριθμος έρευνας δε θα κολλήσει σε ένα τοπικό ακρότατο. Η σειρά των λειτουργιών της επιλογής και της διασταύρωσης μπορεί να παραμένει αμετακίνητη σε ένα ομογενές σύνολο λύσεων. Κάτω από αυτές τις συνθήκες, όλα τα χρωμοσώματα είναι ίδια και συνεπώς η μέση καταλληλότητα του πληθυσμού δε μπορεί να βελτιωθεί. Παρόλα αυτά, η προτεινόμενη λύση μπορεί να εμφανίζεται ως βέλτιστη, αλλά στην πραγματικότητα είναι τοπικά βέλτιστη, επειδή απλά ο αλγόριθμος έρευνας αδυνατεί να συνεχίσει παρακάτω. Η μετάλλαξη είναι το αντίστοιχο της τυχαίας έρευνας, και βοηθά στην προώθηση της γενετικής ποικιλομορφίας του πληθυσμού.

Ο τρόπος λειτουργίας του τελεστή μετάλλαξης είναι ο να αντιστρέφει ένα τυχαία επιλεγμένο γονίδιο σε ένα χρωμόσωμα. Για παράδειγμα ένα χρωμόσωμα Χ1 μπορεί να μεταλλαχθεί στο δεύτερο γονίδιο του, ενώ ένα χρωμόσωμα Χ2 στο τρίτο γονίδιο του. Προφανώς, είναι πιθανό κάποιο χρωμόσωμα να μην υποστεί καμία μετάλλαξη. Γι' αυτό μπορούμε να πούμε πως, η μετάλλαξη προκύπτει σε οποιοδήποτε χρωμόσωμα με κάποια πιθανότητα. Η πιθανότητα μετάλλαξης είναι πολύ μικρή στη φύση, και διατηρείται πολύ μικρή και για τους ΓΑ, συνήθως μεταξύ των τιμών 0.001 και 0.01.

Αναπαράσταση Πολλών Μεταβλητών

Οι γενετικοί αλγόριθμοι εγγυώνται τη συνεχή βελτίωση της μέσης καταλληλότητας του πληθυσμού, και μετά από έναν αριθμό γενεών (αρκετών εκατοντάδων συνήθως) ο πληθυσμός εξελίσσεται σε μια **υπό-βέλτιστη** λύση. Στο παράδειγμα μας, ο τελικός πληθυσμός θα αποτελείται μόνο από τα χρωμοσώματα 0111 και 1000. Οι τελικές θέσεις των χρωμοσωμάτων στη συνάρτηση καταλληλότητας φαίνονται στο σχήμα 7.3

Σε αυτό το παράδειγμα, υπάρχει μόνο μια μεταβλητή η οποία είναι εύκολο να αναπαρασταθεί. Αλλά ας υποθέσουμε πως είναι επιθυμητό να βρεθεί το μέγιστο μιας συνάρτησης δύο μεταβλητών:

$$f(x,y) = (1-x)^2 e^{-x^2-(y+1)^2} - (x - x^3 - y^3)e^{-x^2-y^2},$$

όπου οι μεταβλητές *x* και *y* κυμαίνονται μεταξύ του -3 και του 3.

Το πρώτο βήμα είναι να αναπαραστήσουμε τις μεταβλητές του προβλήματος σαν ένα χρωμόσωμα. Με άλλα λόγια, αναπαριστούμε τις μεταβλητές *x* και *y* σαν την ένωση των δύο

σειριακών δυαδικών επιμέρους αναπαραστάσεων (δηλαδή ενώνουμε την δυαδική αναπαράσταση του *x* με αυτή του *y*), όπου κάθε μεταβλητή αποτελείται από 8 δυαδικά ψηφία. Στη συνέχεια εκτελούμε όλες τις τυπικές διαδικασίες των ΓΑ, όπως περιγράφησαν πρωτύτερα.

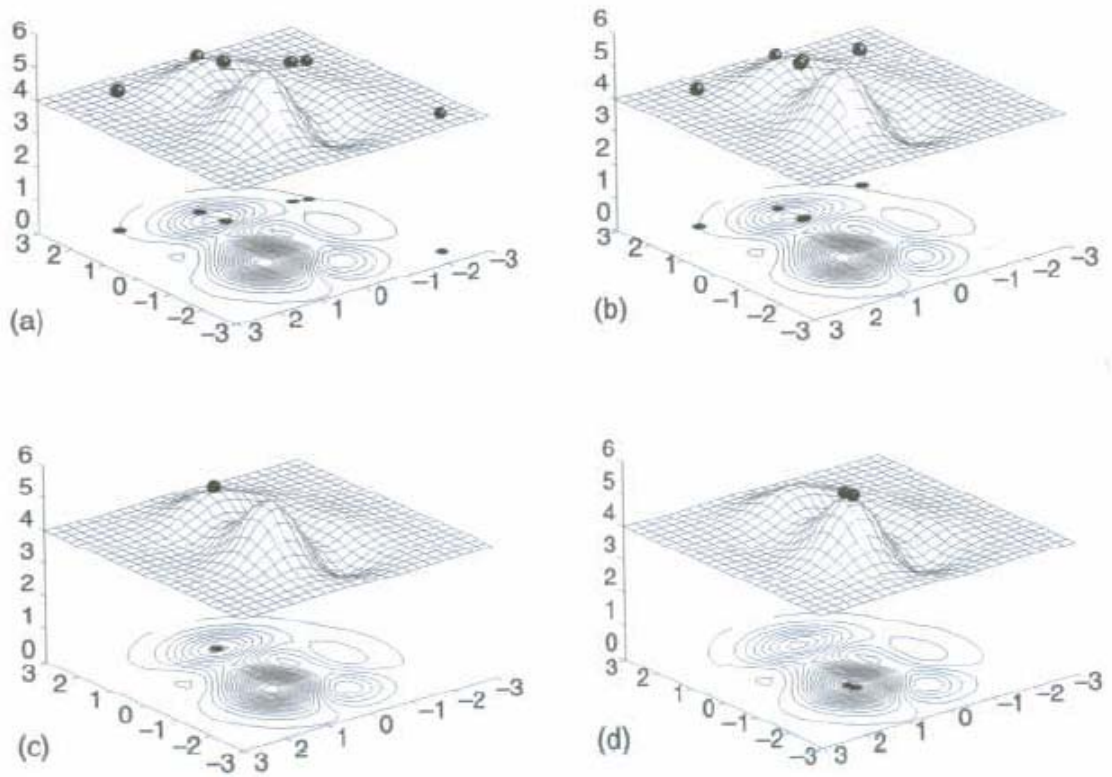
Αρχικά, επιλέγουμε το μέγεθος του πληθυσμού των χρωμοσωμάτων *N*, έστω 6, και με τυχαίο τρόπο δημιουργούμε έναν αρχικό πληθυσμό. Ύστερα υπολογίζουμε την καταλληλότητα κάθε χρωμοσώματος. Αυτό γίνεται σε δύο στάδια και εδώ είναι το σημείο που υπάρχει η διαφοροποίηση. Πρώτα το χρωμόσωμα αποκωδικοποιείται σε δύο πραγματικούς αριθμούς, εκ των οποίων ο ένας αντιστοιχεί στην τιμή του *x* και ο άλλος του *y*, στο διάστημα -3 και 3. Τέλος, οι αποκωδικοποιημένες τιμές για τα *x* και *y* αντικαθίστανται στη συνάρτηση. Όμως πως ακριβώς γίνεται η αποκωδικοποίηση;

Πρώτα, ένα χρωμόσωμα, δηλαδή στο συγκεκριμένο παράδειγμα μια σειρά από 16 ψηφία, χωρίζεται σε δύο σειρές των 8 ψηφίων. Αφού γίνει αυτό, οι δύο σειρές μετατρέπονται από τη δυαδική βάση στη δεκαδική και το διάστημα των επιτρεπόμενων τιμών των *x* και *y* διαιρείται με το 2^8-1, ώστε να δημιουργηθεί το βήμα στο σύνολο των διακριτών τιμών κατά το οποίο μπορούν να μεταβάλλονται οι μεταβλητές. Στη συνέχεια πολλαπλασιάζεται η δεκαδική βάση κάθε χρωμοσώματος με το βήμα και αφαιρείται το τμήμα του πεδίου τιμών που είναι μικρότερο του μηδενός, στη συγκεκριμένη περίπτωση που έχουμε [-3,3] αυτό είναι το -3. Έτσι καταφέραμε να βρούμε τις πραγματικές τιμές των *x* και *y.*

Αν κριθεί απαραίτητο, μπορούμε να εφαρμόσουμε άλλες τεχνικές κωδικοποίησης, όπως ο κώδικας Gray.

Χρησιμοποιώντας τις αποκωδικοποιημένες τιμές του *x* και *y* ως εισόδους στη μαθηματική συνάρτηση, ο ΓΑ υπολογίζει τελικά την καταλληλότητα κάθε χρωμοσώματος.

Για την εύρεση του μεγίστου της συνάρτησης, χρησιμοποιείται η διασταύρωση με πιθανότητα ίση με 0,7 και η μετάλλαξη με πιθανότητα 0,001. Όπως προαναφέρθηκε, μια κοινή πρακτική στους ΓΑ είναι να συγκεκριμενοποιήσουμε τον αριθμό των γενεών. Έστω ότι ο αριθμός των γενεών είναι 100, αυτό σημαίνει πως ο ΓΑ θα δημιουργήσει 100 γενεές 6 χρωμοσωμάτων προτού σταματήσει.



Σχήμα 4.4: Επιφάνεια έρευνας και κίνηση χρωμοσωμάτων σε αυτή.(α) Αρχικές τυχαίες θέσεις χρωμοσωμάτων. (b) Ενδιάμεσες θέσεις στην προσπάθεια προσέγγισης του μεγίστου. (c) Σύγκλιση σε τοπικό μέγιστο. (d) Σύγκλιση στο ολικό μέγιστο ύστερα από μεταβολή του συντελεστή μετάλλαξης σε καλύτερη τιμή

Στο παραπάνω σχήμα φαίνεται η πρόοδος της καταλληλότητας των χρωμοσωμάτων. Στο (a) παρουσιάζονται οι αρχικές θέσεις των χρωμοσωμάτων στην επιφάνεια. Κάθε χρωμόσωμα αναπαρίσταται από μια σφαίρα. Ο αρχικός πληθυσμός αποτελείται από τυχαία δημιουργημένα επιμέρους χρωμοσώματα τα οποία είναι διαφορετικά ή ανομοιογενή. Ωστόσο, ξεκινώντας από τη δεύτερη γενιά, η διασταύρωση αρχίζει να συνδυάζει τα χαρακτηριστικά των καλύτερων χρωμοσωμάτων, και ο πληθυσμός ξεκινά να συγκλίνει στην κορυφή που περιέχει το μέγιστο, όπως φαίνεται στο (b). Από τότε έως την τελική γενιά, ο ΓΑ ερευνά γύρω από αυτήν την κορυφή με τη βοήθεια της μετάλλαξης, οδηγώντας σε ποικιλομορφία. Το (c) δείχνει την τελική γενιά των χρωμοσωμάτων. Όμως, ο πληθυσμός έχει συγκλίνει σε ένα χρωμόσωμα που βρίσκεται σε ένα τοπικό μέγιστο της συνάρτησης.

Ο σκοπός είναι η εύρεση του ολικού μέγιστου, οπότε πώς μπορούμε να είμαστε σίγουροι ότι η έρευνα γίνεται για τη βέλτιστη λύση; Το πιο σοβαρό πρόβλημα στη χρήση των ΓΑ σχετίζεται με

την ποιότητα των αποτελεσμάτων, και συγκεκριμένα το κατά πόσο η πραγματικά βέλτιστη λύση μπορεί να βρεθεί. Ένας τρόπος να παρασχεθεί ένας βαθμός ασφάλειας είναι να συγκριθούν τα αποτελέσματα που προκύπτουν ύστερα από χρήση διαφορετικών βαθμών μετάλλαξης. Αν για παράδειγμα αυξηθεί η πιθανότητα μετάλλαξης στο 0.01 στο παράδειγμά μας, τότε τα χρωμοσώματα συγκλίνουν στο (d). Πάντως, για να έχουμε σταθερά αποτελέσματα θα πρέπει να αυξήσουμε το μέγεθος του πληθυσμού *N*, όσο αυξάνεται ο συντελεστής μετάλλαξης.

Γενικά οι συναρτήσεις καταλληλότητας των πραγματικών προβλημάτων δεν μπορούν να αναπαρασταθούν γραφικά εύκολα. Για το λόγο αυτό χρησιμοποιούμε τα γραφήματα απόδοσης.

Γράφημα Απόδοσης

Επειδή οι γενετικοί αλγόριθμοι είναι στοχαστικής φύσης, η απόδοσή τους συνήθως κυμαίνεται από γενιά σε γενιά. Έτσι ο πιο χρήσιμος τρόπος για να εξετάσουμε τη συμπεριφορά του ΓΑ ύστερα από ένα επιλεγμένο αριθμό γενεών, είναι η καμπύλη που δείχνει την απόδοση του καλύτερου επιμέρους χρωμοσώματος του πληθυσμού. Πιο συγκεκριμένα, δημιουργούμε μια γραφική παράσταση στην οποία ο άξονας x αντιστοιχεί στον αριθμό των γενεών και ο άξονας y την τιμή της συνάρτησης καταλληλότητας. Μαζί με την καταλληλότητα του καλύτερου χρωμοσώματος του πληθυσμού σε κάθε γενιά, μπορούμε να απεικονίζουμε και τη μέση καταλληλότητα του πληθυσμού.

Γενικά η συμπεριφορά των καμπυλών μέσης απόδοσης είναι αναξιόπιστη και αυτό οφείλεται στη μετάλλαξη. Ο τελεστής μετάλλαξης επιτρέπει σε έναν ΓΑ να εξερευνά το τοπίο με τυχαίο τρόπο. Μπορεί να οδηγήσει σε σημαντική βελτίωση της καταλληλότητας του πληθυσμού, αλλά συνήθως τη μειώνει. Για να εξασφαλίσουμε τη γενετική ποικιλομορφία και την ίδια στιγμή να μειώσουμε τις επικίνδυνες συνέπειες της μετάλλαξης, μπορούμε να αυξήσουμε το μέγεθος του πληθυσμού των χρωμοσωμάτων. Είναι εύλογο πως η μέση καμπύλη αυξάνεται ραγδαία στην αρχή του τρεξίματος, αλλά καθώς ο πληθυσμός αρχίζει να συγκλίνει σε μια σχεδόν βέλτιστη λύση, αυξάνεται πιο αργά, ώσπου τελικά σχεδόν οριζοντιώνεται. Αυτή είναι μια τυπική συμπεριφορά των καμπυλών απόδοσης των ΓΑ, την οποία θα απεικονίσουμε και με τα δεδομένα της εργασίας.

4.1.2 Θεωρητικό Υπόβαθρο των ΓΑ

Οι τεχνικές των ΓΑ έχουν πολύ ισχυρό θεωρητικό υπόβαθρο το οποίο στηρίζεται στο θεώρημα του σχήματος. Ο John Holland παρουσίασε την έννοια του σχήματος η οποία προέρχεται από την ελληνική λέξη που είναι συνώνυμη της μορφής. Το σχήμα είναι ένα σύνολο από σειρές ψηφίων με πιθανές τιμές τα 0,1 και τους αστερίσκους, όπου κάθε αστερίσκος σημαίνει είτε 0 είτε 1. Με άλλα λόγια, τα 0 και 1 συμβολίζουν σταθερές θέσεις σε ένα σχήμα, ενώ οι αστερίσκοι τους ¨μπαλαντέρ¨ ή διαφορετικά τα σημεία εκείνα που δε μας ενδιαφέρει η τιμή

τους. Για παράδειγμα, το σχήμα 1**0 αποτελεί ένα σύνολο με σειρές 4 ψηφίων. Κάθε σειρά σε αυτό το σύνολο ξεκινά με 1 και τελειώνει με 0. Αυτές οι σειρές ονομάζονται **στιγμιότυπα** του σχήματος. Με άλλα λόγια στιγμιότυπο ενός σχήματος είναι μια σειρά όπου οι αστερίσκοι έχουν αντικατασταθεί με τιμές.

Περνώντας τώρα στους ΓΑ, η σχέση μεταξύ ενός σχήματος και ενός χρωμοσώματος είναι πολύ απλή. Ένα χρωμόσωμα ταιριάζει με ένα σχήμα όταν οι σταθερές θέσεις του σχήματος ταιριάζουν με τις αντίστοιχες του χρωμοσώματος. Για παράδειγμα, το σχήμα H={1**0} ταιριάζει με τα εξής χρωμοσώματα:

1110
1100
1010
1000

Σε αυτήν την περίπτωση κάθε χρωμόσωμα ξεκινά με 1 και τελειώνει με 0. Έτσι μπορούμε να πούμε πως τα χρωμοσώματα αποτελούν στιγμιότυπα του σχήματος H, και συνεπώς αποτελούν υποπεριπτώσεις του. Αυτή είναι και η σχέση τους.

Ο αριθμός των καθορισμένων ψηφίων (όχι οι αστερίσκοι) σε ένα σχήμα καλείται **τάξη**. Στο προηγούμενο σχήμα, για παράδειγμα, υπάρχουν 2 καθορισμένα ψηφία, και άρα η τάξη του είναι 2.

Η απόσταση μεταξύ των δύο ακραίων καθορισμένων ψηφίων ενός σχήματος καλείται **χαρακτηριστικό μήκος**. Για παράδειγμα, το χαρακτηριστικό μήκος του ****1011 είναι 3, του *0*1*10* είναι 5.

Κοντολογίς, οι γενετικοί αλγόριθμοι εκμεταλλεύονται τα σχήματα όταν τρέχουν. Εάν οι ΓΑ χρησιμοποιούν μια τεχνική που κάνει την πιθανότητα αναπαραγωγής ανάλογη της καταλληλότητας του χρωμοσώματος, τότε σύμφωνα με το θεώρημα του σχήματος, μπορούμε να προβλέψουμε την παρουσία ενός δοσμένου σχήματος στη επόμενη γενιά χρωμοσωμάτων. Με άλλα λόγια, μπορούμε να περιγράψουμε τη συμπεριφορά ενός ΓΑ με όρους αύξησης ή μείωσης του αριθμού των στιγμιότυπων ενός δεδομένου σχήματος.

Ας υποθέσουμε ότι τουλάχιστον ένα στιγμιότυπο του σχήματος H είναι παρόν στην αρχική γενιά των χρωμοσωμάτων i. Έστω τώρα $m_H(i)$ να είναι ο αριθμός των στιγμιότυπων του σχήματος H στη γενεά I, και $f_H(i)$ να είναι η μέση καταλληλότητα αυτών των στιγμιότυπων. Θέλουμε να υπολογίσουμε των αριθμό των στιγμιότυπων της επόμενης γενεάς $m_H(i+1)$. Όσο η πιθανότητα αναπαραγωγής είναι ανάλογη της καταλληλότητας του χρωμοσώματος, μπορούμε

εύκολα να υπολογίσουμε τον αναμενόμενο αριθμό απογόνων ενός χρωμοσώματος Χ στην επόμενη γενεά:

$$m_X(i+1) = \frac{f_X(i)}{\hat{f}(i)},$$

όπου $f_X(i)$ είναι η καταλληλότητα ενός χρωμοσώματος Χ, και $\hat{f}(i)$ η μέση καταλληλότητα της αρχικής γενεάς του χρωμοσώματος i.

Ύστερα, υποθέτοντας ότι το χρωμόσωμα Χ είναι ένα στιγμιότυπο του σχήματος Η, έχουμε:

$$m_H(i+1) = \frac{\sum_{X=1}^{X=m_H(i)} f_X(i)}{\hat{f}(i)}, \qquad X \in H$$

και αφού εξ' ορισμού:

$$\hat{f}_H(i) = \frac{\sum_{X=1}^{X=m_H(i)} f_X(i)}{m_H(i)},$$

προκύπτει:

$$m_H(i+1) = \frac{\hat{f}_H(i)}{\hat{f}(i)} m_H(i)$$

Άρα αποδεδειγμένα ένα σχήμα με καταλληλότητα άνω του μέσου όρου, θα τείνει όντως να εμφανίζεται συχνότερα στην επόμενη γενεά χρωμοσωμάτων και ένα σχήμα με καταλληλότητα κάτω του μέσου όρου θα τείνει να εμφανίζεται λιγότερο συχνά.

Και τι συμβαίνει με τις επιδράσεις της διασταύρωσης και της μετάλλαξης;

Η διασταύρωση και η μετάλλαξη μπορούν και οι δύο να δημιουργήσουν και να καταστρέψουν στιγμιότυπα του σχήματος. Εδώ θα λάβουμε υπ' όψιν μας μόνο τις καταστροφικές επιδράσεις, δηλαδή τις επιδράσεις που μειώνουν τον αριθμό των στιγμιότυπων του σχήματος Η. Πρώτα όμως θα πρέπει να ποσοτικοποιήσουμε την καταστροφή που προκαλείται από τον τελεστή διασταύρωσης. Το σχήμα θα επιβιώσει μετά τη διασταύρωση, εάν τουλάχιστον ένας από τους απογόνους είναι επίσης και στιγμιότυπό του. Αυτή είναι η περίπτωση κατά την οποία η διασταύρωση δεν προκύπτει μέσα στο χαρακτηριστικό μήκος του σχήματος.

Εάν η διασταύρωση λαμβάνει χώρα μεταξύ του χαρακτηριστικού μήκους, το σχήμα Η μπορεί να καταστραφεί και απόγονοι που δεν είναι στιγμιότυπα του Η μπορούν να δημιουργηθούν

(ωστόσο το σχήμα Η δε είναι σίγουρο ότι θα καταστραφεί εάν δύο ίδια χρωμοσώματα διασταυρωθούν, ακόμα και όταν η διασταύρωση γίνεται μέσα στο χαρακτηριστικό μήκος).

Άρα, η πιθανότητα το σχήμα Η να επιβιώσει μετά τη διασταύρωση μπορεί να οριστεί ως:

$$p_H^{(c)} = 1 - p_c\left(\frac{l_d}{l-1}\right),$$

όπου p_c είναι η πιθανότητα διασταύρωσης, και l και l_d είναι αντίστοιχα, το μήκος και το χαρακτηριστικό μήκος του σχήματος Η.

Είναι προφανές, ότι η πιθανότητα επιβίωσης υπό συνθήκες διασταύρωσης είναι μεγαλύτερη για μικρά σχήματα απ' ότι για μεγάλα.

Ας αναλογιστούμε τώρα τις αρνητικές επιπτώσεις της μετάλλαξης. Έστω p_m η πιθανότητα μετάλλαξης για οποιοδήποτε ψηφίο του σχήματος Η, και *n* η τάξη του σχήματος Η. Τότε η παράσταση (1-p_m) αναπαριστά την πιθανότητα το ψηφίο να μη μεταλλαχθεί και συνεπώς η πιθανότητα το σχήμα Η να επιβιώσει και μετά τη διαδικασία μετάλλαξης είναι:

$$p_H^{(m)} = (1 - p_m)^n$$

Είναι επίσης προφανές, ότι η πιθανότητα επιβίωσης υπό τη διαδικασία της μετάλλαξης είναι μεγαλύτερη για σχήματα μικρότερης τάξης απ' ότι για σχήματα μεγαλύτερης. Μπορούμε τώρα να τροποποιήσουμε την εξ.(7.3) ώστε να λαμβάνει υπ' όψιν τις καταστροφικές επιπτώσεις της διασταύρωσης και της μετάλλαξης:

$$m_H(i+1) = \frac{\hat{f}_H(i)}{\hat{f}(i)} m_H(i)\left[1 - p_c\left(\frac{l_d}{l-1}\right)\right](1 - p_m)^n$$

Αυτή η συνάρτηση περιγράφει την ανάπτυξη του σχήματος από τη μια γενιά στην επόμενη και είναι γνωστή ως **το Θεώρημα του Σχήματος**. Επειδή η εξ. 7.6 λαμβάνει υπ' όψιν μόνο τις αρνητικές συνέπειες της διασταύρωσης και της μετάλλαξης, μας δίνει ένα κατώτατο όριο του αριθμού των στιγμιότυπων του σχήματος Η στην επόμενη γενεά.

Πέρα από το γεγονός ότι η διασταύρωση αποτελεί ένα από τα μεγαλύτερα πλεονεκτήματα των ΓΑ, δεν υπάρχει μέχρι στιγμής κάποια θεωρητική βάση ώστε να υποστηριχθεί η άποψη πως οι ΓΑ μπορούν να υπερσκελίσουν από πλευράς απόδοσης τις υπόλοιπες τεχνικές έρευνας και βελτιστοποίησης απλά και μόνο επειδή η διασταύρωση επιτρέπει το συνδυασμό μερικών λύσεων.

Οι γενετικοί αλγόριθμοι είναι ένα ισχυρό εργαλείο, αλλά πρέπει να χρησιμοποιούνται με εξυπνάδα. Για παράδειγμα, η κωδικοποίηση του προβλήματος ως σειρά ψηφίων μπορεί να αλλάξει τη φύση του προβλήματος προς εξερεύνηση. Με άλλα λόγια, υπάρχει κίνδυνος η κωδικοποίηση ως διαδικασία να μεταβάλλει το πρόβλημα σε ένα άλλο διαφορετικό από αυτό που αρχικά θέλαμε να λύσουμε.

4.2. Στρατηγικές Εξέλιξης

Μια άλλη προσέγγιση της εξομοίωσης της φυσικής εξέλιξης προτάθηκε στη Γερμανία στις αρχές του 1960. Σε αντίθεση με τους γενετικούς αλγορίθμους, αυτή η προσέγγιση, γνωστή ως στρατηγική εξέλιξης, σχεδιάστηκε για να επιλύει προβλήματα τεχνικής βελτιστοποίησης.

Το 1963 δύο μαθητές του Πολυτεχνείο του Βερολίνου, ο Ingo Rechenberg και ο Hans-Paul Schwefer, εργαζόντουσαν πάνω στην έρευνα για βέλτιστα σχήματα σωμάτων σε ροή. Κατά την εργασία τους, χρησιμοποίησαν το τούνελ ανέμου του Ινστιτούτου Μηχανικής Ροής. Επειδή η όλη διαδικασία είχε καθαρά πειραματική διάθεση, αποφάσισαν να μεταβάλλουν τυχαία τις διάφορες παραμέτρους που καθορίζουν το σχήμα του σώματος, ακολουθώντας το παράδειγμα της φυσικής μετάλλαξης. Αποτέλεσμα αυτού, η γέννηση των στρατηγικών εξέλιξης.

Οι στρατηγικές εξέλιξης αναπτύχθηκαν ως υποκατάστατο της διαίσθησης του μηχανικού. Μέχρι προσφάτως, οι στρατηγικές εξέλιξης χρησιμοποιούνταν σε προβλήματα τεχνικής βελτιστοποίησης όπου κάποια αναλυτική αντικειμενική συνάρτηση δεν ήταν διαθέσιμη, και καμία συμβατική μέθοδος βελτιστοποίησης δεν υπήρχε, με αποτέλεσμα οι μηχανικοί να στηρίζονται μόνο στη διαίσθησή τους. Σε αντίθεση με τους ΓΑ, οι στρατηγικές εξέλιξης χρησιμοποιούν μόνο τον τελεστή μετάλλαξης.

Πως υλοποιούμε μια στρατηγική εξέλιξης;

Στην απλούστερη μορφή της, γνωστή ως (1+1)-στρατηγική εξέλιξης, ένας γονέας παράγει έναν απόγονο ανά γενεά με την εφαρμογή κανονικώς κατανεμημένης μετάλλαξης. Η (1+1)-στρατηγική εξέλιξης μπορεί να υλοποιηθεί ως εξής:

1. Διάλεξε τον αριθμό των παραμέτρων *N* που αναπαριστούν το πρόβλημα, και στη συνέχεια καθόρισε ένα αληθοφανές εύρος τιμών για κάθε παράμετρο. Καθόρισε την τυπική απόκλιση για κάθε παράμετρο και τη συνάρτηση προς βελτιστοποίηση:

$$\{x_{1min}, x_{1max}\}, \{x_{2min}, x_{2max}\}, \dots, \{x_{Nmin}, x_{Nmax}\},$$

2. Επέλεξε τυχαία μια αρχική τιμή για κάθε παράμετρο από το αντίστοιχο προκαθορισμένο εύρος. Το σύνολο αυτών των παραμέτρων θα αποτελεί τον αρχικό πληθυσμό των γονεϊκών παραμέτρων.

$$x_1, x_2, \dots, x_N$$

3. Υπολόγισε τη λύση που σχετίζεται με τις γονεϊκές παραμέτρους:

$$X = f(x_1, x_2, \dots, x_N)$$

4. Δημιούργησε μια νέα παράμετρο (απόγονος) μέσω της προσθήκης μιας κανονικά κατανεμημένης τυχαίας μεταβλητής α με μηδενικό μέσο όρο και προεπιλεγμένη απόκλιση δ σε κάθε γονεϊκή παράμετρο:

$$x_i' = x_i + \alpha(0, \delta), \qquad i = 1,2, \dots, N$$

   Οι κανονικά κατανεμημένες μεταλλάξεις με μηδενικό μέσο όρο αντικατοπτρίζουν τη φυσική διαδικασία εξέλιξης όπου μικρότερες αλλαγές προκύπτουν συχνότερα από τις μεγαλύτερες.

5. Υπολόγισε τη λύση που σχετίζεται με παραμέτρους-παιδιά

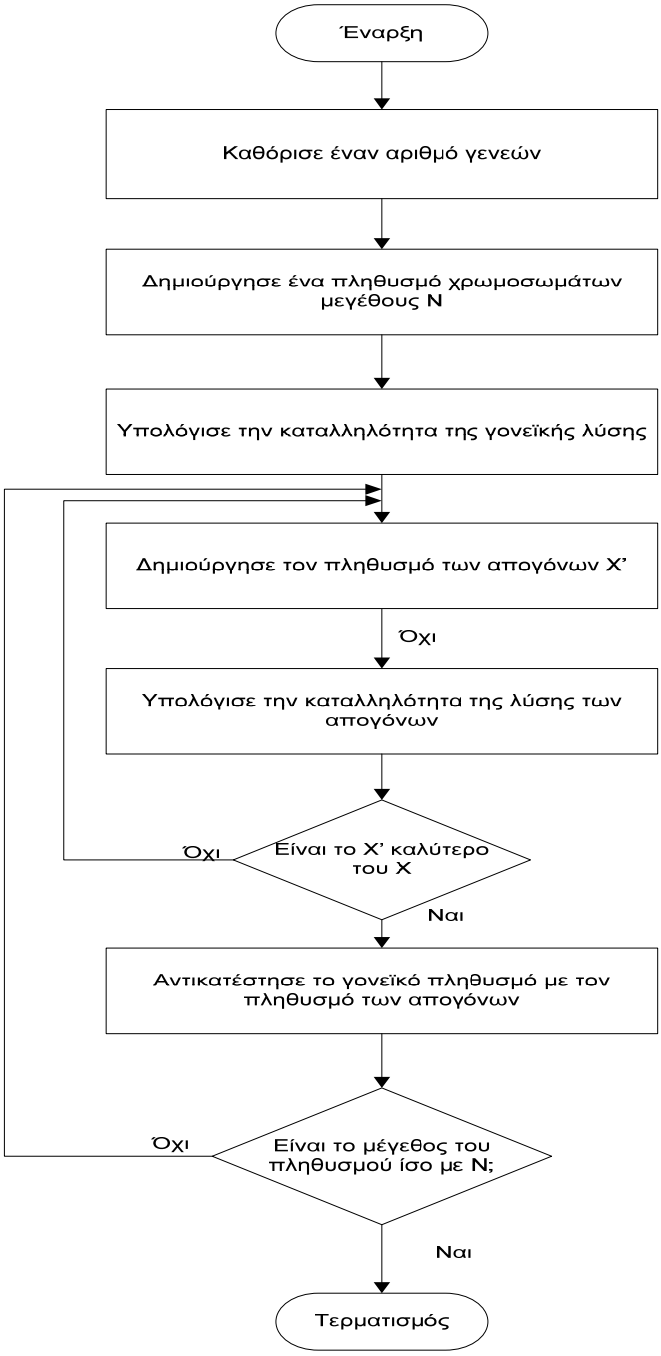
$$X' = f(x'_1, x'_2, \dots, x'_N))$$

6. Σύγκρινε τη λύση που προκύπτει από τις παραμέτρους-παιδιά με αυτή που προκύπτει από τους γονείς. Εάν η λύση των παιδιών είναι καλύτερη από αυτή των γονέων, αντικατέστησε τον γονεϊκό πληθυσμό με το πληθυσμό των απογόνων.
7. Πήγαινε στο βήμα 4, και επανέλαβε τη διαδικασία μέχρι να φθάσουμε σε μια ικανοποιητική λύση, ή σε μέχρι να παρέλθει ένας συγκεκριμένος αριθμός γενεών. Η (1+1)-στρατηγική εξέλιξης μπορεί να παρασταθεί και με το διάγραμμα του σχήματος 4.5

Μια στρατηγική εξέλιξης αντικατοπτρίζει τη φύση του χρωμοσώματος. Στην πραγματικότητα, ένα απλό γονίδιο μπορεί ταυτόχρονα να επηρεάζει διάφορα χαρακτηριστικά του ζωντανού οργανισμού. Από την άλλη πλευρά, ένα απλό χαρακτηριστικό ενός υποκειμένου μπορεί να καθορίζεται από την ταυτόχρονη αλληλεπίδραση διαφόρων γονιδίων. Η φυσική επιλογή λειτουργεί σε μια συλλογή από γονίδια, και όχι σε ένα γονίδιο απομονωμένα. Για το λόγο αυτό, μεταβάλλονται ταυτόχρονα όλες οι παράμετροι όταν παράγουμε μια νέα λύση.

Οι στρατηγικές εξέλιξης μπορούν να επιλύσουν ένα μεγάλο εύρος από περιορισμένα ή μη, μη-γραμμικά προβλήματα βελτιστοποίησης και παράγουν καλύτερα αποτελέσματα από πολλές συμβατικές, πολυσύνθετες, μη-γραμμικές τεχνικές. Πειράματα επίσης υποδεικνύουν ότι η απλή έκδοση των στρατηγικών εξέλιξης που χρησιμοποιεί το στυλ έρευνας ενός γονέα-ενός απογόνου δουλεύει καλύτερα.

Η κύρια διαφορά μεταξύ ενός ΓΑ και μιας στρατηγικής εξέλιξης είναι πως η πρώτη χρησιμοποιεί ταυτόχρονα και τη διασταύρωση και τη μετάλλαξη, ενώ η τελευταία μόνο τη μετάλλαξη. Επιπλέον, όταν χρησιμοποιούμε μια στρατηγική εξέλιξης δε χρειάζεται να κωδικοποιήσουμε το πρόβλημα. Μια εξελικτική στρατηγική χρησιμοποιεί μια καθαρά αριθμητική διαδικασία βελτιστοποίησης, παρόμοια με μια συγκεντρωμένη έρευνα Monte Carlo. Οι ΓΑ είναι εφαρμόσιμοι σε πιο γενικές κατηγορίες προβλημάτων, αλλά το δυσκολότερο κομμάτι τους είναι η κωδικοποίηση. Γενικά, για να απαντήσουμε στο ερώτημα ποια μέθοδος είναι καλύτερη, πρέπει να πειραματιστούμε για να το βρούμε. Εξαρτάται λοιπόν από την εκάστοτε εφαρμογή.



Σχήμα 4.5: Διάγραμμα ροής στρατηγικών εξέλιξης

5. ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΚΑΙ ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Εισαγωγή

Οι γενετικοί αλγόριθμοι λοιπόν, αποτελούν τους πιο δημοφιλείς εκπροσώπους του εξελικτικού υπολογισμού. Στο προηγούμενο κεφάλαιο έγινε η εκτενής γενική θεωρητική περιγραφή τους. Στο συγκεκριμένο κεφάλαιο θα δούμε πως μπορούμε να τους εφαρμόσουμε στο κομμάτι που μας ενδιαφέρει, δηλαδή για να πραγματοποιήσουμε την επιλογή χαρακτηριστικών. Επιλέχθηκε η χρήση τους καθότι μας προσφέρουν μια εντελώς διαφορετική φιλοσοφία προσέγγισης του ζητήματος επιλογής χαρακτηριστικών απ' ότι ο SFFS (ο έτερος αλγόριθμος που χρησιμοποιήθηκε) και κατ' επέκταση μας δίνουν τη δυνατότητα να εξερευνήσουμε με εντελώς διαφορετικό τρόπο των χώρο των λύσεων. Έτσι είναι πιθανό να καταλήξουμε σε πιο αποτελεσματικές λύσεις και γενικότερα να ξεπεράσουμε τις τυχόν αδυναμίες που είχε ο SFFS. Αυτό δε σημαίνει ότι οι γενετικοί αλγόριθμοι είναι τέλειοι, απλά εξερευνούμε τις πιθανότητες να προσφέρουν καλύτερες ποιοτικά λύσεις στη συγκεκριμένη κατηγορία προβλημάτων.

Στη συνέχεια του κεφαλαίου ακολουθούν όλες οι λεπτομέρειες υλοποίησης των επιμέρους χαρακτηριστικών τμημάτων του αλγορίθμου που χρησιμοποιήθηκε στην εργασία.

5.1 Απλός Γενετικός Αλγόριθμος για την επιλογή χαρακτηριστικών

Ένα σημαντικό ζήτημα στο σχεδιασμό ενός ΓΑ είναι η διαδικασία που ακολουθείται για τον έλεγχο των γενετικών διαδικασιών. Δύο εναλλακτικές επιλογές υπάρχουν η σταθερής κατάστασης και οι διαδικασίες γέννησης. Αυτές παρουσιάζονται παρακάτω με τη μορφή ψευδοκώδικα:

```
ΓΑ_σταθερής_κατάστασης()
{
        αρχικοποίησε τον πληθυσμό P;
        επανέλαβε
        {
                διάλεξε 2 γονείς ρ1 και ρ2 από τον P;
                απόγονος=διασταύρωση(ρ1,ρ2);
                μετάλλαξη(απόγονος);
                αντικατέστησε(P,απόγονος);
        } μέχρι (συνθήκη τερματισμού);
```

```
}
```

```
ΓΑ_γέννησης()
{
        αρχικοποίησε τον πληθυσμό P;
        επανέλαβε
        {
                από (i=1 έως |P|)
                {
                        διάλεξε 2 γονείς ρ1 και ρ2 από τον P;
                        απόγονος(i)=διασταύρωση(ρ1,ρ2);
                        μετάλλαξη(απόγονος(i));
                }
                αντικατέστησε P με απόγονος(1),...,απόγονος(|P|);
        } μέχρι (συνθήκη τερματισμού);
}
```

Η πιο σημαντική και ουσιαστική διαφορά ανάμεσά τους είναι στη μέθοδο που χρησιμοποιείται για την ενημέρωση του πληθυσμού. Η διαδικασία σταθερής κατάστασης γενικά ανανεώνει ένα χρωμόσωμα σε κάθε γενεά, ενώ η διαδικασία γέννησης ανανεώνει το σύνολο ή τουλάχιστον ένα μεγάλο μέρος του πληθυσμού. Αυτή η διαφορά έχει σημαντική επίδραση στη συμπεριφορά και την δυνατότητα του ΓΑ να βρει μια λύση σε ένα μεγάλο χώρο έρευνας. Ακολουθεί η λεπτομερής περιγραφή των υπολοίπων βασικών διαδικασιών που χρησιμοποιούνται στους ΓΑ και είναι κοινές και στους 2 τύπους.

Κωδικοποίηση Χρωμοσωμάτων

Για την κωδικοποίηση των χρωμοσωμάτων στο πρόβλημα επιλογής χαρακτηριστικών, χρησιμοποιείται μια ακολουθία με *D* δυαδικά ψηφία. Κάθε δυαδικό ψηφίο αναπαριστά ένα χαρακτηριστικό, με τις τιμές 1 και 0 να αντιστοιχούν στην επιλογή του ή όχι αντίστοιχα. Για παράδειγμα ένα χρωμόσωμα {00101000} σημαίνει πως το τρίτο και το πέμπτο χαρακτηριστικό έχουν επιλεγεί. Αυτό σημαίνει, ότι το χρωμόσωμα αναπαριστά το υποσύνολο X={3,5} ή εναλλακτικά το Y={1,2,4,6,7,8}, το οποίο προφανώς είναι το συμπληρωματικό του X.

Αρχικός πληθυσμός

Η γενιά του αρχικού πληθυσμού είναι άμεση, όπως φαίνεται και παρακάτω. Η συνάρτηση random_uniform() δημιουργεί ένα τυχαίο δεκαδικό αριθμό μεταξύ [0,1]. Με τη βοήθεια του προτεινόμενου αλγορίθμου, ο αναμενόμενος αριθμός επιλεγμένων χαρακτηριστικών (τα

γονίδια που έχουν την τιμή 1) σε μια αυθαίρετη αρχική λύση είναι *d*, με πάρα πολύ καλή πιθανότητα. Ακόμα και αν διαφέρει πάντως θα διορθωθεί σε επόμενο στάδιο. Ακολουθεί ο ψευδοκώδικας που περιγράφει τη δημιουργία του αρχικού πληθυσμού:

Αρχικός πληθυσμός:
για (*i*=1 έως |P|**)**
για (κάθε γονίδιο *g* στο *i*-στο χρωμόσωμα**)**
αν (random_uniform()<*d/D*) *g*=1**; αλλιώς** *g*=0**;**

Εκτίμηση καταλληλότητος, επιλογή, αντικατάσταση και λήξη

Η εκτίμηση είναι άμεση καθώς κάθε χρωμόσωμα αναπαριστά ένα σύνολο επιλεγμένων χαρακτηριστικών, X, και η συνάρτηση αξιολόγησης είναι συγκεκριμένη και βασισμένη στον αλγόριθμο FCM. Για να αναγκάσουμε ένα υποσύνολο χαρακτηριστικών που επεξεργάζεται ο αλγόριθμος να ικανοποιεί τις προαπαιτούμενες απαιτήσεις μεγέθους (των ενεργών χαρακτηριστικών), η τιμή του *d*, λαμβάνεται ως *περιορισμός* και επιβάλλεται ποινή στα χρωμοσώματα που τον παραβιάζουν. Η καταλληλότητα ενός χρωμοσώματος C καθορίζεται ως:

$$fitness(C) = J(X_c) - penalty(X_c)$$

όπου το X_c είναι το απαντούμενο σύνολο χαρακτηριστικών του C, και η ποινή

$$penalty(X_c) = w^* \big| |X_c| - d \big|$$

με ένα συντελεστή ποινής, *w*.

Η επιλογή χρωμοσώματος για την επόμενη γενεά γίνεται με βάση την καταλληλότητα. Ο μηχανισμός επιλογής πρέπει να διασφαλίζει ότι τα καταλληλότερα χρωμοσώματα έχουν μεγαλύτερη πιθανότητα επιβίωσης. Ο σχεδιασμός μας υιοθετεί το βασισμένο στην κατάταξη, σύστημα επιλογής της ρουλέτας. Τα χρωμοσώματα του πληθυσμού ταξινομούνται κατά φθίνουσα σειρά με βάση την καταλληλότητα τους, και το *i-στο* χρωμόσωμα ανατίθεται η πιθανότητα επιλογής με βάση την εξής μη-γραμμική συνάρτηση:

$$P(i) = q(1-q)^{i-1}$$

Η παράμετρος q ονομάζεται πίεση επιλογής, και όσο μεγαλύτερη η τιμή του q τόσο πιθανότερο είναι να επιλέγονται τα καλύτερα χρωμοσώματα για ζευγάρωμα και να αγνοούνται εντελώς τα χειρότερα. Με άλλα λόγια δίνει μεγαλύτερη αξία στην καταλληλότητα, δίνοντας μεγαλύτερο κομμάτι της ρουλέτα στα καταλληλότερα χρωμοσώματα. Η διαδικασία της ρουλέτας περιγράφεται παρακάτω:

1. Υπολόγισε τις συσωρευτικές πιθανότητας για το *i-στο* χρωμόσωμα μέσω της σχέσης

$$p_i = \sum_{j=1}^{i} P(j)\,, \qquad \gamma\iota\alpha\ i = 1, \dots, n\ \kappa\alpha\iota\ p_0 = 0$$

2. Δημιούργησε ένα τυχαίο αριθμό r μεταξύ $[0,p_n]$
3. Διάλεξε το *i-στο* χρωμόσωμα έτσι ώστε $p_{i-1}<r<p_i$,

όπου *n* ο αριθμός των χρωμοσωμάτων του πληθυσμού (κυρίως συμβολίζεται με P, απλά εδώ συμπίπτει με το σύμβολο της πιθανότητας που έχει γενικότερη ισχύ)

Για τα ΓΑ σταθερής κατάστασης, επιλέγουμε δύο γονεϊκά χρωμοσώματα χρησιμοποιώντας την παραπάνω μέθοδο. Η διαδικασία διασταύρωσης γεννά ένα νέο χρωμόσωμα (απόγονος) από τους δύο γονείς, και η διαδικασία μετάλλαξης το μεταβάλει ελαφρά.

Εάν το μεταλλαγμένο χρωμόσωμα είναι καλύτερο και από τους δύο γονείς αντικαθιστά τον καλύτερο γονέα, εάν βρίσκεται ανάμεσα τους αντικαθιστά τον χειρότερο γονέα, αλλιώς αντικαθιστά το χειρότερο χρωμόσωμα του πληθυσμού. Ο ΓΑ σταματά όταν ο αριθμός των γενεών φθάνει το μέγιστο αριθμό γενεών *T* , ο οποίος είναι προκαθορισμένος.

Διασταύρωση και Μετάλλαξη

Χρησιμοποιούμε τους συνηθισμένους τελεστές διασταύρωσης και μετάλλαξης με μια μικρή μετατροπή. Ένας τελεστής διασταύρωσης *m*-σημείων χρησιμοποιείται ο οποίος επιλέγει *m* σημεία κοπής στην τύχη και εναλλάξ αντιγράφει κάθε κομμάτι από τον ένα γονέα στον άλλο. Οι τελεστές απεικονίζονται στο σχήμα 1. Η διασταύρωση μπορεί να δώσει έναν απόγονο που ξεπερνά τις ανάγκες μεγέθους του υποσυνόλου, επειδή τα αντιλλαγμένα τους γονιδιακά τμήματα μπορεί να έχουν διαφορετικές περιπτώσεις άσσων (δηλαδή να περιέχουν διαφορετικά ενεργά γονίδια). Αυτό είναι εν δυνάμει πρόβλημα, για το οποίο όμως έχει βρεθεί λύση και θα παρουσιαστεί παρακάτω.

Η μετάλλαξη ασκείται πάνω στον απόγονο. Όπως και στην περίπτωση της διασταύρωσης, η μετάλλαξη είναι πιθανόν να επηρεάσει τους περιορισμούς μεγέθους του υποσυνόλου, οπότε χρειάζεται να ελέγχουμε προσεκτικά των αριθμό των μετατροπών από 1 σε 0 και αντιθέτως. Ο ακόλουθος κώδικας αναγκάζει τους αριθμούς να είναι ίδιοι, όπου η παράμετρος p_m αναπαριστά τον ρυθμό μετάλλαξης (ή πιθανότητα μετάλλαξης).

Ελεγχόμενη μετάλλαξη:

1. Έστω n_0 και n_1 είναι οι πληθυσμοί των ψηφίων 0 και 1 αντίστοιχα στο χρωμόσωμα.
2. Θέτουμε $p_1=p_m$ και $p_0=p_m(n_1/n_0)$
3. Για κάθε γονίδιο g στο χρωμόσωμα κάνε τα εξής:
    a. Δημιούργησε ένα τυχαίο αριθμό r ανάμεσα στο [0,1]
    b. αν (g=1 και $r<p_1$) μετέτρεψε το g σε 0,
    αλλιώς αν (g=0 και $r<p_0$) μετέτρεψε το g σε 1

Παράμετροι

Καμία συστηματική διαδικασία βελτιστοποίησης των χαρακτηριστικών δεν έχει ακόμα δημιουργηθεί, αλλά το ακόλουθο σύνολο παραμέτρων χρησιμοποιήθηκε στα πειράματα μας. Η ακριβής ρύθμιση των παραμέτρων ανάλογα με το πείραμα είναι πιθανόν να αυξήσει την απόδοση.

Προτεινόμενες τιμές παραμέτρων

- Διαδικασία ελέγχου: σταθερής κατάστασης
- P(μέγεθος πληθυσμού)=20
- p_c(πιθανότητα διασταύρωσης)=1,0
- p_m(πιθανότητα μετάλλαξης)=0,1
- q(πίεση επιλογής)=0,25
- w(συντελεστής ποινής)=0,5
- T(αριθμός γενεών)= εξαρτάται από το σύνολο δεδομένων

5.2 ΥΒΡΙΔΙΚΟΣ ΓΑ

Οι γενετικοί αλγόριθμοι έχουν την ικανότητα να ξεφεύγουν από τα τοπικά ελάχιστα με τη βοήθεια των τελεστών διασταύρωσης και μετάλλαξης, και να εξερευνούν ένα μεγάλο εύρος της περιοχής έρευνας, όταν η παράμετρος της πίεση επιλογής ελέγχεται σωστά. Παρόλα αυτά, παρουσιάζουν αδυναμία στην ακριβή ρύθμιση των τοπικών ελαχίστων τους, και αυτό προκαλεί μεγάλους χρόνους εκτέλεσης. Για τη βελτίωση της ακρίβειας των απλών ΓΑ, έχουν αναπτυχθεί σε πολλές εφαρμογές υβριδικοί ΓΑ, Αυτές περιλαμβάνουν το πρόβλημα του κινούμενου πωλητή, το πρόβλημα διαχωρισμού των γράφων και τη συμπίεση εικόνων. Σε έναν υβριδικό ΓΑ, τα χρωμοσώματα βελτιώνονται με τη χρήση κατάλληλων τελεστών τοπικής έρευνας, για αυτό και αυτές οι διαδικασίες είναι γνωστές και ως διαδικασίες τοπικής βελτιστοποίησης. Παρακάτω ακολουθεί η περιγραφή ενός υβριδικού ΓΑ για το πρόβλημα επιλογής χαρακτηριστικών.

Η βασική ιδέα πίσω από το συγκεκριμένο αλγόριθμο είναι η ενσωμάτωση στον τυπικό ΓΑ των τελεστών τοπικής έρευνας. Η διαδικασία σταθερής κατάστασης που υλοποιεί αυτήν την ιδέα φαίνεται παρακάτω με τη μορφή ψευδοκώδικα:

Υβριδικός_ΓΑ ()
{
    αρχικοποίησε τον πληθυσμό P;
    επανέλαβε
    {
        διάλεξε 2 γονείς ρ1 και ρ2 από τον P;
        απόγονος=διασταύρωση(ρ1,ρ2);
        μετάλλαξη(απόγονος);

        τοπική βελτιστοποίηση(απόγονος);

        αντικατέστησε(P,απόγονος);
    } μέχρι (συνθήκη τερματισμού);
}
όπου,

Τοπική Βελτιστοποίηση(C) /*C ένα χρωμόσωμα */
{
    βάλε τα χαρακτηριστικά του C με τιμή 1 στο X;
    βάλε τα χαρακτηριστικά του C με τιμή 0 στο Y;

    άλλαξε
    {
    περίπτωση |X|=d: ripple_rem(r); ripple_add(r);
    περίπτωση |X|<d: **επανέλαβε** ripple_add(r) (**d-|X|) φορές**
    περίπτωση |X|>d: **επανέλαβε** ripple_rem(r) (|**X|-d) φορές**
    }

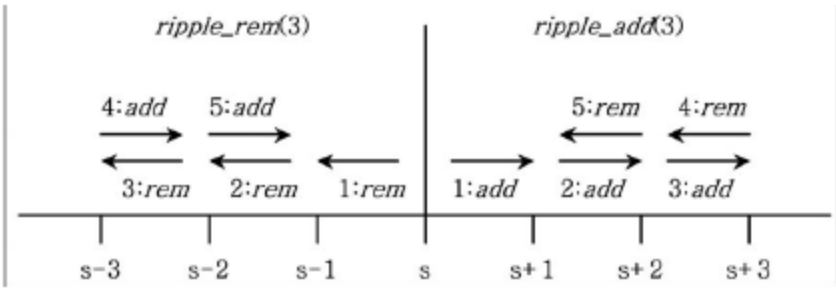
    θέσε την τιμή των στοιχείων που δείχνει ο X στον C ίση με 1;
    θέσε την τιμή των στοιχείων που δείχνει ο Y στον C ίση με 0;

}

Είναι πιθανόν ο απόγονος που γεννάται ύστερα από τη διασταύρωση και τη μετάλλαξη να κληρονομήσει τα καλά χαρακτηριστικά των γονιών και έτσι να είναι είτε ανώτερος και από τους 2 είτε ανάμεσα τους. Στην χειρότερη περίπτωση θα είναι κατώτερος και των δύο. Στο συγκεκριμένο υβριδικό ΓΑ, πριν την αντικατάσταση, ο απόγονος έχει την ευκαιρία να αυξήσει την καταλληλότητα με χρήση τελεστών τοπικής βελτιστοποίησης.

Ένα βασικό πρόβλημα στην πραγματοποίηση της παραπάνω ιδέας σχετίζεται με το πρόβλημα της επιλογής των κατάλληλων διαδικασιών για την τοπική βελτιστοποίηση. Οι διαδικασίες που χρησιμοποιούνται είναι απλές˙ η add βρίσκει το καλύτερο χαρακτηριστικό εκτός συνόλου και το προσθέτει σε αυτό. Αντίστοιχα η rem αφαιρεί το λιγότερο σημαντικό από το σύνολο. Με βάση αυτές δημιουργούνται η ADD(k) και η REM(k) που εκτελούν τις αντίστοιχες διαδικασίες k φορές. Οι διαδικασίες χρησιμοποιούν ένα συγκεκριμένο υποσύνολο χαρακτηριστικών X, το οποίο αναπαρίσταται από ένα χρωμόσωμα, και κάνουν τοπικές έρευνες γύρω από τη συγκεκριμένη λύση με την απομάκρυνση των λιγότερο σημαντικών χαρακτηριστικών ή την προσθήκη των περισσότερο σημαντικών, ανάλογα με το τι είναι απαραίτητο. Με την εφαρμογή τους σε συγκεκριμένη σειρά, μπορούμε να πετύχουμε την τοπική βελτίωση του χρωμοσώματος. Επιπλέον της τοπικής βελτιστοποίησης, με τη χρήση των παραπάνω διαδικασιών σε κατάλληλους συνδυασμούς καταφέρνουμε να ελέγξουμε πλήρως το μέγεθος του υποσυνόλου, αντιμετωπίζοντας έτσι, τις αρκετά πιθανές μεταβολές του μεγέθους του λόγω της εφαρμογής των τελεστών διασταύρωσης και μετάλλαξης. Με άλλα λόγια, η απομάκρυνση ή η προσθήκη χαρακτηριστικών μπορεί να θεωρηθούν ως *διορθωτικοί* τελεστές.

Η τοπική βελτίωση ενός χρωμοσώματος επιτυγχάνεται με τη χρήση του παραπάνω κώδικα. Πρώτα, μετατρέπει το χρωμόσωμα C σε δύο υποσύνολα χαρακτηριστικών, X και Y. Το X περιέχει τα ενεργά γονίδια, ενώ το Y είναι το συμπληρωματικό του. Στη συνέχεια, χρησιμοποιούμε έναν από τους δύο, ή και τους δύο, τελεστές τοπικής έρευνας, Η διαδικασία *ripple_rem(r)* πρώτα αφαιρεί τα λιγότερο σημαντικά χαρακτηριστικά r φορές και στη συνέχεια προσθέτει τα πιο σημαντικά χαρακτηριστικά r-1 φορές, οδηγώντας στην απομάκρυνση ενός χαρακτηριστικού τελικά. Αντιστοίχως, η *ripple_add(r)* επιτελεί την ακριβώς αντίθετη διαδικασία και τελικά αυξάνει το μέγεθος του X κατά 1. Ονομάζουμε την παράμετρο *r, παράγοντα κυματισμού.* Το ακόλουθο σχήμα δείχνει δύο παραδείγματα διαδικασιών τοπικής έρευνας.



Σχήμα 5.1: Εφαρμογή της τοπικής βελτιστοποίησης για r=3

Ακολουθούν πιο συνοπτικά τα συστατικά των δύο αλγορίθμων:

$$ripple_rem(r) \equiv \{REM(r); ADD(r-1);\}, r \geq 1$$
$$ripple_add(r) \equiv \{ADD(r); REM(r-1);\}, r \geq 1$$

Ο αλγόριθμος καλείται να αντιμετωπίσει τρείς διαφορετικές καταστάσεις , ανάλογα με τον αριθμό των χαρακτηριστικών στο Χ και τους περιορισμούς μεγέθους του υποσυνόλου *d*. Αυτές είναι:

1. Ο περιορισμός μεγέθους ικανοποιείται (δηλαδή το Χ περιέχει d χαρακτηριστικά):
   Πάνω στο Χ ασκούνται οι διαδικασίες *ripple_rem(r)* και *ripple_add(r)*.
2. Υπάρχουν λιγότερα χαρακτηριστικά στο Χ από το επιθυμητό:
   το μέγεθος του Χ αυξάνεται με την εφαρμογή της *ripple_add(r)* όσες φορές χρειάζεται.
3. Υπάρχουν περισσότερα χαρακτηριστικά στο Χ από το επιθυμητό:
   το μέγεθος του Χ μειώνεται με την εφαρμογή της *ripple_rem(r)*όσες φορές χρειάζεται.

Ο παράγοντας κυματισμού *r* χρησιμοποιείται για τον έλεγχο της δύναμης της τοπικής βελτίωσης. Παρόλο που η μείωση στον αριθμό των χαρακτηριστικών που προκαλείται από την *ripple_rem(r)* είναι 1, ανεξάρτητη δηλαδή του *r* , ο παράγοντας κυματισμού άμεσα επηρεάζει τον πραγματικό αριθμό των διαδικασιών *rem* και *add* που πρόκειται να εκτελεστούν. Όσο μεγαλύτερος είναι ο παράγοντας τόσο ισχυρότερη τοπική βελτίωση παρέχεται καθώς γίνεται έρευνα σε μεγαλύτερο βάθος. Για παράδειγμα, η *ripple_rem(2)* εκτελεί δύο *rem* και μία *add* και είναι προφανώς πιο αποτελεσματική από τη *ripple_rem(1)* η οποία εκτελεί μόνο μία *rem*.

Από αυτή τη διαδικασία έχουμε δύο ευεργετικά αποτελέσματα, που είναι πολύ χρήσιμα στην επίλυση των προβλημάτων επιλογής χαρακτηριστικών. Το πρώτο είναι η γενικότερη βελτίωση της απόδοσης που προέρχεται από την τοπική βελτίωση των χρωμοσωμάτων. Επειδή χρησιμοποιούμε τις βασικές διαδικασίες *rem* και *add* οι οποίες ψάχνουν τα λιγότερο σημαντικά χαρακτηριστικά και τα περισσότερο σημαντικά αντίστοιχα, ένα χρωμόσωμα τείνει να βελτιώνει την απόδοση του τοπικά. Το δεύτερο ευεργετικό αποτέλεσμα αφορά στον έλεγχο του μεγέθους του υποσυνόλου. Με τη χρήση μιας σειράς από διαδικασίες *add* και *rem*,

μπορούμε εύκολα να επισκευάσουμε ένα αποκλίνον υποσύνολο, ώστε να τηρεί τις προδιαγραφές μεγέθους *d*.

Οι διαδικασίες τοπικού ελέγχου απαιτούν αρκετό υπολογιστικό χρόνο. Ένας ευρύτερος παράγοντας κυματισμού αυξάνει πολύ τον υπολογιστικό χρόνο. Αυτό είναι και το μοναδικό ουσιαστικό μειονέκτημα, το οποίο όμως δυστυχώς είναι και πολύ έντονο. (Από χρόνους σε κάτω του ενός λεπτού για τον απλό ΓΑ, χρειαζόμαστε αρκετά λεπτά για τον υβριδικό ΓΑ, διατηρώντας όλες τις παραμέτρους ίδιες).

Από την άλλη πλευρά, ο ΓΑ συγκλίνει ταχύτερα λόγω της δυνατότητας ακριβής ρύθμισης των τοπικών ερευνών, συνεπώς θεωρητικά απαιτεί λιγότερες εποχές για να φθάσει σε μια καλή λύση. Ένα από τα σημαντικότερα ζητήματα σχεδίασης ενός αποδοτικού υβριδικού ΓΑ είναι το πως θα αυξήσουμε τις διαδικασίες τοπικής έρευνας, καθώς ελαχιστοποιούμε τον αριθμό των διαδικασιών προς εκτέλεση, ώστε τελικά να χρησιμοποιήσουμε τα πλεονεκτήματα της ιδέας εξαλείφοντας όσο γίνεται τα μειονεκτήματα της.

6. ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΕΣ

Εισαγωγή

Είδαμε λοιπόν στα προηγούμενα κεφάλαια τις διαδικασίες που είναι απαραίτητες για την προεπεξεργασία και την επιλογή χαρακτηριστικών. Τα αποτελέσματα τους τροφοδοτούνται σε στο υποσύστημα των κατηγοριοποιητών, το οποίο αποτελεί το τρίτο και τελευταίο κομμάτι του συστήματός μας. Το κομμάτι αυτό είναι το πραγματικά υπεύθυνο για την κατηγοριοποίηση των δεδομένων και την παρουσίαση των αποτελεσμάτων. Τα είδη των κατηγοριοποιητών ποικίλουν και βασικά χωρίζονται σε δύο κατηγορίες ανάλογα με τον τρόπο μάθησης που τους χαρακτηρίζει:

- επιβλεπόμενη μάθηση
- μη-επιβλεπόμενη μάθηση

Ως μάθηση ορίζεται η διαδικασία που ακολουθεί ο κατηγοριοποιητής για να ρυθμίσει σωστά τις κατασκευαστικές παραμέτρους του. Σωστά σημαίνει ότι η ρύθμιση τους του επιτρέπει να τοποθετεί κάθε πρότυπο που δέχεται ως είσοδο στην κατηγορία που πραγματικά ανήκει.

Στην επιβλεπόμενη μάθηση έχουμε ένα σύνολο εκπαίδευσης (το οποίο γενικά είναι υποσύνολο του αρχικού συνόλου δεδομένων που θέλουμε να κατηγοριοποιήσουμε), και ένας «δάσκαλος» παρέχει από πριν τη γνώση για την κατηγορία στην οποία ανήκει κάθε πρότυπο από το σύνολο αυτό. Έτσι ο κατηγοριοποιητής χρησιμοποιεί αυτήν την πληροφορία για να μάθει τον τρόπο με τον οποίο κατηγοριοποιείται αυτό το είδος δεδομένων, με απώτερο στόχο τη σωστή λειτουργία του (σωστή κατηγοριοποίηση) όταν έρθει σε επαφή με άγνωστα δεδομένα του ιδίου είδους.

Στη μη-επιβλεπόμενη μάθηση ή ομαδοποίηση, δεν υπάρχει η γνώση των κατηγοριών κάθε προτύπου εισόδου και ο κατηγοριοποιητής μαθαίνει με το να ομαδοποιεί τα γειτονικά στο χώρο χαρακτηριστικών πρότυπα με βάση ένα σύνολο κριτηρίων (παράμετροι). Προφανώς εδώ δεν υπάρχει σύνολο εκπαίδευσης καθώς όλα τα πρότυπα είναι άγνωστα για το σύστημα.

Στην εργασία υλοποιήσαμε μια πληθώρα κατηγοριοποιητών και από τις δύο μεγάλες αυτές κατηγορίες. Μάλιστα, επειδή το υποσύστημα επιλογής χαρακτηριστικών είναι τύπου φίλτρου (δηλαδή ανεξάρτητο του συστήματος κατηγοριοποίησης), μπορούσαμε ταυτόχρονα να τους χρησιμοποιούμε όλους και έτσι να βγάλουμε πιο εμπεριστατωμένα συμπεράσματα για την απόδοσή τους και τις μεταξύ τους διαφορές. Στη συνέχεια του κεφαλαίου ακολουθεί η θεωρητική περιγραφή όλων των σχημάτων κατηγοριοποίησης που χρησιμοποιήθηκαν.

6.1 Τεχνητά Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα ξεκίνησαν από την προσπάθεια μοντελοποίησης των βιολογικών νευρώνων από τους McCulloch και Pitts. Αυτοί πρώτοι έδωσαν μια μαθηματική περιγραφή σχετικά με τον τρόπο λειτουργίας ενός βιολογικού νευρώνα. Η γενική σχέση που περιγράφει το νευρώνα είναι:

$$y = f(u), \qquad \mu\varepsilon\ u = \sum_{i=1}^{P} w_i x_i$$

Η κατάσταση του νευρώνα περιγράφεται από ένα δυαδικό αριθμό y ο οποίος λαμβάνει δύο τιμές 1 και 0, ανάλογα με τον εάν είναι ενεργοποιημένος ή όχι. Οι συνάψεις του νευρώνα περιγράφονται από τα **συναπτικά βάρη** $\boldsymbol{w_i}$ τα οποία είναι πραγματικοί αριθμοί, θετικοί εάν έχουμε ενισχυτικές συνάψεις και αρνητικοί εάν έχουμε ανασταλτικές. Τέλος $x_1, x_2,\ldots,x_n$ είναι οι είσοδοι του νευρώνα. Αυτές οι είσοδοι δημιουργούν το διάνυσμα εισόδου του νευρώνα. Σε κάθε είσοδο αντιστοιχεί ένα συναπτικό βάρος w_i, το οποίο πολλαπλασιάζεται με την τιμή εισόδου. Αντίστοιχα τα βάρη δημιουργούν το διάνυσμα συναπτικών βαρών. Συνεπώς, το άθροισμα του φορτίου που δέχεται ο νευρώνας στην είσοδό του είναι:

$$u = \sum_{i=1}^{P} w_i x_i$$

γνωστό και ως **διέγερση**.

Αν το άθροισμα αυτό είναι μικρότερο από ένα **κατώφλι θ**, τότε ο νευρώνας γίνεται ανενεργός, ειδάλλως ενεργοποιείται (το y γίνεται 0 και 1 αντίστοιχα).

$$y = f(u - \theta)$$

όπου f(.) είναι η γνωστή βηματική συνάρτηση.

Το κατώφλι γενικά θεωρείται και ως συναπτικό βάρος w_0 το οποίο πολλαπλασιάζεται με μοναδιαία είσοδο και αφαιρείται. Το νέο διάνυσμα συναπτικών βαρών, που περιέχει και το w_0, ονομάζεται επαυξημένο διάνυσμα συναπτικών βαρών. Έτσι έχουμε μια πιο απλή μορφή για την είσοδο χρησιμοποιώντας μόνο τα γινόμενα των βαρών επί τις εισόδους, χωρίς πρόσθετα στοιχεία:

$$u = \sum_{i=1}^{P} w_i x_i - \theta = u = \sum_{i=0}^{P} w_i x_i$$

όπου $w_0=\vartheta$ και $x_0=-1$

Συνεπώς ύστερα και από αυτή τη διευκρίνηση μπορούμε να πούμε πως ο νευρώνας ενεργοποιείται όταν η είσοδος είναι μεγαλύτερη του 0, ενώ μένει ανενεργός όταν είναι μικρότερη ή ίση. Τέλος, η συνάρτηση f είναι αυτή που διαφοροποιεί το μοντέλο. Πέρα από τη βηματική που περιγράφηκε παραπάνω και θεωρείται η πιο απλή, χρησιμοποιούνται και άλλες όπως η σιγμοειδής, η υπερβολική εφαπτομένη, η συνάρτηση κατωφλίου, η συνάρτηση ράμπας κτλ.

Γενικά τα νευρωνικά δίκτυα μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες ανάλογα με το αν τα βάρη τους παραμένουν σταθερά ή μεταβάλλονται με βάση μια διαδικασία εκπαίδευσης. Στην εργασία αυτή χρησιμοποιούμε νευρωνικά με εκπαιδευόμενα βάρη με επίβλεψη. Αυτό σημαίνει πως η πληροφορία για κάθε πρότυπο εκπαίδευσης είναι ήδη γνωστή, οπότε σκοπός του νευρωνικού είναι να δώσει τέτοιες τιμές στις παραμέτρους του ώστε να κατηγοριοποιεί σωστά τα δοσμένα πρότυπα.

Το πιο απλό νευρωνικό δίκτυο που μπορεί να κατασκευασθεί είναι αυτό που περιέχει μόνο ένα νευρώνα και ονομάζεται perceptron. Η λειτουργία του είναι ακριβώς όπως περιγράφηκε παραπάνω. Σαν επιπλέον πληροφορία να παραθέσουμε πως η έξοδος y μπορεί να έχει δύο μορφές. Την κλασσική μορφή (1/0) και τη λεγόμενη διπολική μορφή (1/-1), αν και η επιλογή μεταξύ των δύο έχει λίγη σημασία. Η βασική παράμετρος που καθορίζει τη συμπεριφορά του νευρώνα είναι το επαυξημένο διάνυσμα των συναπτικών βαρών.

Η διέγερση u μπορεί να γραφεί λοιπόν και ως το εσωτερικό γινόμενο μεταξύ των διανυσμάτων **x** και **w**:

$$u=\mathbf{w}^T\mathbf{x}$$

Εκπαίδευση, είναι η διαδικασία κατά την οποία μεταβάλλεται το διάνυσμα βαρών **w** του νευρωνικού δικτύου με σκοπό να επιτύχει τη σύγκλιση των τιμών εξόδου του, όπως προκύπτουν από την είσοδο κάθε προτύπου $\mathbf{x}_i$, με τις προκαθορισμένες τιμές-στόχους που αντιστοιχούν στο πρότυπο αυτό $\mathbf{d}_i$. Η διαδικασία εκμάθησης είναι **αυτόματη** (δηλαδή μόνο του το σύστημα προσπαθεί να βρει τις σωστές τιμές) και **επιβλεπόμενη** (δηλαδή υπάρχει ένα είδος δασκάλου που παρέχει την πληροφορία για τους επιθυμητούς στόχους και έτσι βοηθά το δίκτυο να καταλάβει πότε ένα έκανε μια κίνηση προς την σωστή κατεύθυνση του στόχου και πότε μια προς την αντίθετη λανθασμένη)

Ο κλασσικός κανόνας εκπαίδευσης του Perceptron είναι γνωστός και ως κανόνας σταθερής αύξησης. Ο κανόνας είναι επαναληπτικός. Στο δίκτυο παρουσιάζονται τα πρότυπα με τη σειρά εισόδου και μόλις ολοκληρωθεί η παρουσίαση όλων, η διαδικασία επαναλαμβάνεται. Κάθε κύκλος εκπαίδευσης ονομάζεται **εποχή**. Το επαυξημένο διάνυσμα συναπτικών βαρών μεταβάλλεται μόνο αν υπάρχει διαφορά μεταξύ της εξόδου του νευρωνικού από την

επιθυμητή τιμή του στόχου, περίπτωση που ονομάζεται *σφάλμα ταξινόμησης.* Τότε ακολουθείται ο κανόνας διόρθωσης

$$\boldsymbol{w}(k) = \boldsymbol{w}(k-1) + \beta(d^{(p)} - y)\boldsymbol{x}^{(p)}$$

όπου k είναι ο αριθμός της επανάληψης στην οποία βρισκόμαστε και **w**(k) η κατάσταση του επαυξημένου διανύσματος κατά την επανάληψη αυτή. Η παράμετρος *β* ονομάζεται **ρυθμός εκπαίδευσης ή βήμα εκπαίδευσης** και καθορίζει το μέγεθος της διόρθωσης που θα λάβει χώρα με βάση τον παραπάνω κανόνα. Είναι ένας μικρός θετικός αριθμός.

Αποδεικνύεται πως η εκπαίδευση του w γίνεται με τέτοιο τρόπο ώστε το πρότυπο που ταξινομήθηκε τώρα εσφαλμένα, την επόμενη φορά είτε θα ταξινομηθεί σωστά (θα έχει φθάσει το στόχο), είτε θα πλησιάσει περισσότερο την ορθότερη ταξινόμηση (θα πλησιάζει το στόχο).

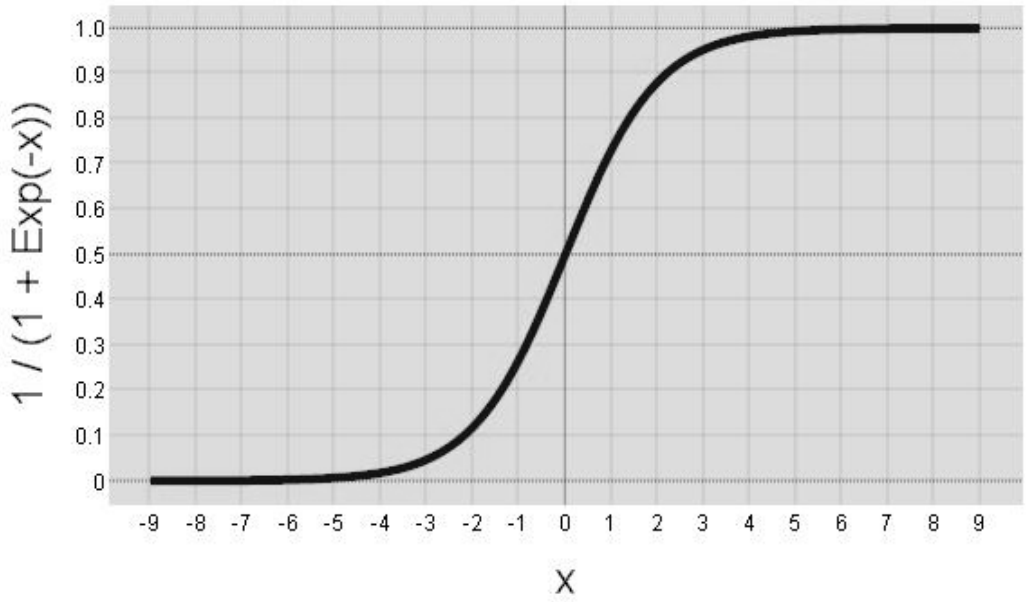
Η βασική ιδιότητα του δικτύου είναι πως μπορεί να λύσει οποιοδήποτε γραμμικά διαχωρίσιμο πρόβλημα, δηλαδή οποιοδήποτε πρόβλημα δύο διαστάσεων που μπορεί να λυθεί, σε πεπερασμένο αριθμό επαναλήψεων. Το δίκτυο perceptron σε αυτή την περίπτωση αντιστοιχεί στη διαχωριστική ευθεία των δύο κατηγοριών, στο επίπεδο xy. Σε όλες τις άλλες περιπτώσεις (μεγαλύτερων διαστάσεων δηλαδή) δε μπορεί να δώσει καμία λύση.

Multilayer Perceptrons

Όσα προηγήθηκαν αφορούσαν στο απλούστερο νευρωνικό δίκτυο, το δίκτυο Perceptron. Οι δυνατότητες αναπαράστασης διαχωριστικών επιφανειών είναι περιορισμένες, καθώς με ένα νευρώνα μπορούμε να αναπαραστήσουμε μόνο επίπεδες επιφάνειες. Όμως, μπορούμε να δημιουργήσουμε ένα δίκτυο που θα αποτελείται από περισσότερους του ενός νευρώνες. Αυτά τα δίκτυα ονομάζονται Multi-Layer Perceptrons (MLP), δηλαδή perceptron πολλαπλών στρωμάτων. Στα δίκτυα αυτά το πρώτο στάδιο παραμένει αυτό των εισόδων και το τελευταίο παραμένει ένας νευρώνας. Όμως στο ενδιάμεσο προστίθενται στρώματα νευρώνων τα οποία ονομάζονται **κρυφά επίπεδα**. Κάθε κρυφό επίπεδο περιέχει περισσότερους του ενός νευρώνες παράλληλα. Το βασικό χαρακτηριστικό των δικτύων αυτών είναι ότι οι νευρώνες ενός στρώματος *(l)* τροφοδοτούν αποκλειστικά τους νευρώνες του επομένου στρώματος (*l+1)* και τροφοδοτούνται αποκλειστικά από τους νευρώνες του προηγουμένου στρώματος *(l-1).* Επίσης κάθε νευρώνας υλοποιεί και μια διαφορετική διαχωριστική ευθεία. Οι ευθείες αυτές συνδυάζονται από το νευρώνα εξόδου για να μας δώσουν την περιοχή λύσεων. Αυτή η προσθήκη επιτρέπει στο δίκτυο να λύνει συνθετότερα προβλήματα, τόσο στο επίπεδο όσο και σε χώρους περισσοτέρων διαστάσεων.

Να παρατηρήσουμε εδώ με την ευκαιρία πως στα MLP η χρήση της βηματικής συνάρτησης ως συνάρτησης ενεργοποίησης δεν προτιμάται. Αυτό οφείλεται στο ότι οι περισσότεροι κανόνες εκπαίδευσης βασίζονται σε μεθόδους βελτιστοποίησης για να επιτύχουν και μεγαλύτερη απόδοση και μικρότερο χρόνο τρεξίματος, οι οποίες χρησιμοποιούν παραγώγους, όπως π.χ. η

μέθοδος της κατάβασης δυναμικού, ενώ η βηματική συνάρτηση δεν είναι παραγωγίσιμη. Η τεχνική αυτή δυσκολία αντιμετωπίζεται με τη χρήση της **σιγμοειδούς συναρτήσεως:**



Σχήμα 6.1: Σιγμοειδής συνάρτηση. Παρατηρούμε ότι αποτελεί μια "μαλακή" εκδοχή της μοναδιαίας συνάρτησης

Η συνάρτηση αυτή μοιάζει πολύ με τη βηματική όπως φαίνεται και από το σχήμα.

Μια άλλη συνάρτηση που μπορεί να χρησιμοποιηθεί εναλλακτικά είναι αυτή της **υπερβολικής εφαπτομένης**:

$$\frac{\quad}{\quad}$$

Η χρήση τέτοιων «μαλακών» συναρτήσεων κατωφλίωσης οδηγεί σε πιο ομαλές επιφάνειες χωρίς απότομες μεταβολές στην τιμή y της εξόδου του δικτύου.

Συνοψίζοντας τα MLP έχουν πολλές δυνατότητες αναπαράστασης συναρτήσεων και μπορούν να προσεγγίσουν οποιαδήποτε ομαλή συνάρτηση όσο κοντά επιθυμούμε, αρκεί να χρησιμοποιηθούν μόλις δύο στρώματα νευρώνων (χωρίς να λαμβάνουμε υπ' όψιν το στρώμα εισόδου). Για το λόγο αυτό τα δίκτυα ονομάζονται **«Καθολικοί Προσεγγιστές».**

Ο κανόνας εκπαίδευσης που χρησιμοποιείται ονομάζεται κανόνας εκπαίδευσης Back-Propagation. Χρησιμοποιεί διάνυσμα στόχων, οπότε η μάθηση είναι **επιβλεπόμενη**. Η λογική

του είναι να μεταδίδει το σφάλμα αντίθετα από την ορθή κατεύθυνση λειτουργίας του δικτύου, δηλαδή από την έξοδο προς την είσοδο. Έτσι το σφάλμα αρχικά υπολογίζεται στο στάδιο εξόδου και ύστερα με βάση αυτό διορθώνονται τα βάρη του προηγούμενου στρώματος, και ύστερα με βάση αυτά διορθώνονται αυτά του προηγούμενου και η διαδικασία αυτή επαναλαμβάνεται μέχρι να φθάσουμε στο στρώμα εισόδου, οπότε και ολοκληρώνεται η διαδικασία. Στην ιδανική περίπτωση θέλουμε οι έξοδοι του δικτύου για κάθε πρότυπο, να ταυτίζονται ακριβώς με τους αντίστοιχους στόχους. Όμως αυτό δύσκολα είναι εφικτό στις τυπικές περιπτώσεις με αποτέλεσμα να επιζητούμε τη βέλτιστη προσέγγιση της επιθυμητής κατάστασης χρησιμοποιώντας ένα κριτήριο κόστους. Το πιο συνηθισμένο, και συνεπώς αυτό που θα χρησιμοποιήσουμε είναι το μέσο τετραγωνικό σφάλμα:

$$J = \frac{1}{P}\sum_{p=1}^{P}\left\|d^{(p)} - y^{(p)}\right\|^2 = \frac{1}{P}\sum_{p=1}^{P}\sum_{i=1}^{m}\left[d^{(p)} - y^{(p)}\right]^2$$

Το κριτήριο αυτό έχει το πλεονέκτημα ότι η ελαχιστοποίηση του σημαίνει την ελαχιστοποίηση της τετραγωνικής απόστασης μεταξύ των διανυσμάτων $\mathbf{y}^{(i)}$,$\mathbf{d}^{(i)}$ και επιπλέον παραγωγίζεται εύκολα οπότε διευκολύνει την εφαρμογή κλασσικών μεθόδων βελτιστοποίησης.

Ο αλγόριθμος Back-Propagation αποτελεί στην ουσία μια προσέγγιση της κατάβασης δυναμικού η οποία περιγράφεται από μια δυναμική εξίσωση με παραγώγους.

Τα συναπτικά βάρη w_{ij} είναι οι παράμετροι που πρέπει να διορθωθούν ώστε να ελαχιστοποιηθεί το σφάλμα J. Η μεταβολή των παραμέτρων ως προς το χρόνο t γίνεται χρησιμοποιώντας την παράγωγο του J ως προς w_{ij}:

$$\frac{dw_{ij}}{dt} = -\frac{\partial J}{\partial w_{ij}}$$

η οποία ονομάζεται και κλίση του J ως προς w_{ij}. Προφανώς στους υπολογιστές χρησιμοποιούμε διακριτό χρόνο k. Η μορφή διακριτού χρόνου της παραπάνω συνάρτησης είναι:

$$w_{ij}(l, k+1) - w_{ij}(l, k) = -\beta\frac{\partial J}{\partial w_{ij}(l, k)}$$

Το δύσκολο κομμάτι είναι ο υπολογισμός της παραγώγου $\partial J/\partial w_{ij}(l, k)$. Για να διευκολυνθούμε θα ορίσουμε ένα νέο μέγεθος, την παράγωγο του κόστους ως προς τη δικτυακή διέγερση του νευρώνα *i*:

$$\delta_i^{(k)}(l) = -\frac{\partial J}{\partial u_i^{(k)}(l)}$$

Λόγω του γράμματος δ που χρησιμοποιήθηκε για το συμβολισμό αυτής της παράστασης, ο αλγόριθμος Back-Propagation ονομάζεται και **«Γενικευμένος Κανόνας Δέλτα».** Χρησιμοποιώντας την παραπάνω έκφραση μπορούμε να γράψουμε το εξής:

$$\frac{\partial J}{\partial w_{ij}(l,k)} = \frac{\partial J}{\partial u_i^{(k)}(l)} \frac{\partial u_i^{(k)}(l)}{\partial w_{ij}(l,k)} = -\delta_i^{(k)}(l) \frac{\partial u_i^{(k)}(l)}{\partial w_{ij}(l,k)}$$

όπου πλέον η δεύτερη παράγωγος υπολογίζεται εύκολα καθώς η σχέση είναι γραμμική και ισούται με $\alpha_j^{(k)}(l-1)$ για *j≠0* και με 1 για *j=0.*

Αντίστοιχα η τιμή της δ είναι:

- Για το στρώμα εξόδου L:

$$\delta_i^{(k)}(L) = \left(d_i^{(k)} - y_i^{(k)}\right) f'(u_i^{(k)}(L))$$

- Για οποιοδήποτε άλλο προηγούμενο στρώμα:

$$\delta_i^{(k)}(l) = \sum_{\mu=1}^{N(l+1)} \delta_\mu^{(k)}(l+1) w_{\mu i}^{(k)} f'(u_i^{(k)}(l))$$

Παρατηρούμε πως το σφάλμα σε οποιοδήποτε νευρώνα του στρώματος *l* είναι συνάρτηση του σφάλματος του επόμενου σταδίου *l+1*. Συνεπώς τα σφάλματα μεταδίδονται ανάποδα από την κανονική φορά διάδοσης του δικτύου. Από εκεί πήρε και το όνομα του ο αλγόριθμος. Τελικά η εξίσωση ενημέρωσης των βαρών έχει τη μορφή:

$$w_{ij}(l,k+1) = w_{ij}(l,k) + \beta \delta_i^{(k)} \alpha_j^{(k)}(l-1), \qquad \mu\varepsilon\ j = 0,1,\dots,N(l), \qquad l = 1,\dots,L$$

Η παράμετρος του βήματος εκπαίδευσης *β* λαμβάνει και σε αυτήν την περίπτωση μικρές θετικές τιμές. Όμως εδώ παρουσιάζεται το εξής πρόβλημα· εάν η τιμή του *β* είναι πολύ μικρή για το συγκεκριμένο πρόβλημα τότε ο αλγόριθμος εκπαίδευσης θα συγκλίνει αλλά με πολύ πιο αργό ρυθμό από τον εφικτό μιας βέλτιστης λύσης. Από την άλλη, εάν το *β* έχει μεγάλες τιμές για το πρόβλημα μας ο αλγόριθμος θα συγκλίνει πολύ γρήγορα, όμως είναι αρκετά πιθανό η σύγκλιση να είναι κακή ή και να μη συγκλίνει καθόλου. Γενικά έχουν προταθεί διάφορες μέθοδοι για τη βελτίωση της συμπεριφοράς του *β* σε σχέση με τους στόχους της εξόδου όπως ο κανόνας "delta-delta" ή η χρήση διαφορετικού *β* για κάθε νευρώνα. Περισσότερες πληροφορίες μπορούν να αναζητηθούν στη σχετική βιβλιογραφία καθώς η περιγραφή τους ξεφεύγει της παρούσης εργασίας.

Ο αλγόριθμος Back-Propagation τερματίζει όταν το σφάλμα εξόδου ανάμεσα σε δύο διαδοχικές εποχές είναι μικρότερο από ένα συγκεκριμένο και προκαθορισμένο κατώφλι. Εναλλακτικά μπορεί να χρησιμοποιηθεί η μεταβολή της τιμής των βαρών και να ζητήσουμε αυτή ή το τετράγωνο της να είναι μικρότερο από ένα κατώφλι. Στο δίκτυο της εργασίας ακολουθείται η πρώτη μέθοδος, η οποία είναι και η συνηθέστερη.

Τα βάρη καλό είναι να αρχικοποιούνται σε μικρές τιμές καθώς έτσι αποφεύγεται το φαινόμενο κορεσμού των βαρών, δηλαδή το φαινόμενο κατά το οποίο τα βάρη χάνουν τη δυνατότητα τους να μεταβάλλονται προτού το δίκτυο συγκλίνει σε μια καλή τιμή. Επίσης μια καλή αρχικοποίηση βοηθά τα δίκτυο να ξεφύγει από το πρόβλημα των τοπικών ελαχίστων, δηλαδή ένα άλλο φαινόμενο κατά το οποίο το δίκτυο συγκλίνει σε μια λύση η οποία είναι τοπικά βέλτιστη και όχι ολικά. Ονομάζεται πρόβλημα τοπικών ελαχίστων καθότι όπως προαναφέραμε ο αλγόριθμος αποτελεί μια προσέγγιση της κατάβασης δυναμικού, οπότε και μειώνεται με κάθε επανάληψη μέχρι να συναντήσει το ελάχιστο. Η αιτία της σύμφυτης αυτής αδυναμίας του αλγορίθμου δεν είναι σαφής. Πάντως, αν στην αρχή αποφευχθεί η σύγκλιση σε τέτοιο σημείο, τότε είναι σχεδόν σίγουρο πως ο αλγόριθμος θα δώσει ένα πολύ ικανοποιητικό αποτέλεσμα, ακόμα και αν έχει τελικά συγκλίνει σε τοπικό ελάχιστο. Ο μόνος τρόπος να βελτιώσουμε τις πιθανότητες σωστής εφαρμογής του αλγορίθμου είναι να δώσουμε καλές αρχικές τιμές στα συναπτικά βάρη, και αυτό όπως προαναφέραμε σημαίνει να τους δώσουμε μικρές τιμές.

Βελτιώσεις Αλγορίθμου Εκπαίδευσης

Ο αρχικός αλγόριθμος έχει δύο βασικές αδυναμίες οι οποίες είναι η ύπαρξη τοπικών ελαχίστων και κυρίως η αργή ταχύτητα σύγκλισης. Για να υπερσκελιστούν εισήχθησαν βελτιώσεις και νέες παραλλαγές. Η πρώτη στην πράξη είναι ιδιαίτερα δύσκολο να αντιμετωπιστεί, όμως για τη δεύτερη έχουν αναπτυχθεί διάφορες μέθοδοι όπως *η χρήση ορμής*, *η συζυγής κατάβαση δυναμικού* κ.α.

Χρήση ορμής

Όπως αναφέραμε και προηγουμένως σημαντικό ρόλο για τη συμπεριφορά του δικτύου κατά τη διάρκεια της εκπαίδευσης παίζει η επιλογή του κατάλληλου βήματος εκπαίδευσης *β*, καθώς για μικρές τιμές εμφανίζεται μικρή ταχύτητα σύγκλισης, ενώ για μεγάλες υψηλές ταλαντώσεις ή ακόμα και αδυναμία σύγκλισης. Η επιλογή μας αυτή μπορεί να έχει χαμηλότερες συνέπειες στο δίκτυο, αν χρησιμοποιήσουμε την κατάβαση με χρήση ορμής.

Η παραλλαγή αυτή βασίζεται στην ιδέα ότι τα βάρη είναι σαν αντικείμενα που κατεβαίνουν την πλευρά ενός λόφου με συγκεκριμένο βήμα ανά επανάληψη και συνεπώς διατηρούν, τουλάχιστον ως ένα βαθμό, την πορεία που είχαν κατά την προηγούμενη επανάληψη. Σε πρακτικό επίπεδο αυτό σημαίνει ότι για τη διαμόρφωση της τιμής της διόρθωσης λαμβάνουμε υπ’ όψιν τόσο το αποτέλεσμα της αρχικής μεθόδου, όσο και την τιμή της διόρθωσης κατά την

προηγούμενη επανάληψη, πολλαπλασιασμένο επί μια σταθερά $\boldsymbol{\mu}$ η οποία ονομάζεται **παράμετρος ορμής**. Έτσι αποφεύγουμε τις μεγάλες ταλαντώσεις καθώς επιβάλλεται στο σύστημα να ακολουθήσει μια σταθερότερη πορεία, αλλά και επιτυγχάνουμε επιτάχυνση της μεθόδου ίση με 1/(1-μ) σε περιοχές με σταθερή κλίση. Ο τελικός τύπος που συνοψίζει τα παραπάνω είναι:

$$\Delta w_{ij}(k) \approx -\frac{\beta}{1-\mu}\frac{\partial J}{\partial w_{ij}(k)}$$

Αναζήτηση Ελαχίστου σε ευθεία

Σκοπός αυτής της μεθόδου είναι η αναζήτηση και εύρεση της βέλτιστης τιμής της παραμέτρου εκπαίδευσης *β*. Η τιμή αυτή θα ελαχιστοποιεί τη συνάρτηση κόστους J, η οποία γράφεται στη μορφή:

$$J(\beta)=J(\mathbf{w}(k)+\beta\mathbf{d}_k)$$

Η μέθοδος υποθέτει πως η διεύθυνση $\mathbf{d}_k$ προς την οποία γίνεται η διόρθωση είναι επιλεγμένη και γνωστή και συνεπώς το πρόβλημα είναι πλέον τύπου ελαχιστοποίησης σε μια διάσταση η οποία είναι μια ευθεία που περνά από το σημείο **w**(k) και εκτείνεται προς την κατεύθυνση $\mathbf{d}_k$**.**

Η λογική λειτουργίας συνοψίζεται στις ακόλουθες δύο φάσεις:

- **Περιορισμός** του ελαχίστου μεταξύ δύο ακραίων τιμών
- **Διαίρεση** του διαστήματος σε ολοένα και μικρότερα υποδιαστήματα τα οποία συνεχίζουν να περιέχουν μέσα τους το ελάχιστο

Με αυτόν τον τρόπο ουσιαστικά περιορίζουμε την έρευνα σε μόνο στα σημεία που αυτή είναι απαραίτητο να χρησιμοποιηθεί μειώνοντας κατά πολύ το υπολογιστικό κόστος, χωρίς να θυσιάσουμε την ακρίβεια των αποτελεσμάτων.

Μέθοδος Συζυγούς Κατάβασης Δυναμικού

Κατά τη μέθοδο αυτή γίνεται χρήση συνάρτησης δευτέρου βαθμού για την προσέγγιση της κατάβασης δυναμικού. Έτσι η συνάρτηση κόστους ορίζεται να είναι τετραγωνικής μορφής, η οποία έχει την ιδιότητα να ελαχιστοποιείται εύκολα.

$$J(\boldsymbol{w}) = J(0) + \boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{w} + \frac{1}{2}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{H}\boldsymbol{w}$$

όπου **H** ο πίνακας του Hess του οποίου τα στοιχεία είναι οι διπλές παράγωγοι $H_{ij} = \frac{\partial J}{\partial w_i \partial_w j}$. Υποθέτουμε πως ο πίνακας αυτός είναι συμμετρικός, δηλαδή $\mathbf{H}=\mathbf{H}^{\mathbf{T}}$, θετικά ορισμένος και έχει διάσταση NxN. Μια τέτοια τετραγωνική μορφή λέγεται προβολή στις N διαστάσεις, όπου N ο

αριθμός της διάστασης του διανύσματος **w**. Πράγματι, στο σημείο όπου επιτυγχάνεται το ελάχιστο πρέπει να ισχύει η συνθήκη Kuhn-Tucker:

$$\frac{\partial J}{\partial \boldsymbol{w}} = 0$$

άρα,

$$\boldsymbol{b} + \boldsymbol{H}\boldsymbol{w}^* = 0$$

$$\boldsymbol{w}^* = -\boldsymbol{H}^{-1}\boldsymbol{b}$$

Τα πράγματα θα ήταν πολύ απλά αν η συνάρτηση κόστους του νευρωνικού ήταν η προαναφερθείσα και γνωρίζαμε τις τιμές **b,H.** Όμως η συνάρτηση κόστους του δικτύου μόνο τοπικά μπορεί να προσεγγισθεί από αυτή τη μορφή, και επίσης ο υπολογισμός του **H** είναι δύσκολος. Τα ίδιο δεν ισχύει για το b οποίο σχετίζεται με την κλίση:

$$\boldsymbol{g}(\boldsymbol{w}) = \frac{\partial J}{\partial \boldsymbol{w}} = \boldsymbol{b} + \boldsymbol{H}\boldsymbol{w}(k)$$

οπότε b=g(0). Ολόκληρος ο αλγόριθμος Back-Propagation ασχολείται με αυτό, οπότε ο υπολογισμός είναι ευκολότερος.

Βασικό πλεονέκτημα της μεθόδου συζυγούς κατάβασης δυναμικού είναι ότι παρακάμπτει χωρίς συνέπειες τον υπολογισμό του πίνακα Hess. Οπότε μένει η προσέγγιση της τετραγωνικής συνάρτησης κόστους. Για να το πετύχει αυτό χρησιμοποιεί ιδέες τόσο από την απλή κατάβαση δυναμικού, όσο και από την αναζήτηση ελαχίστου σε ευθεία.

Αποδεικνύεται ότι το βήμα εκπαίδευσης προκύπτει από τον τύπο:

$$\beta_k = -\frac{\boldsymbol{d}_k^T \boldsymbol{g}_k}{\boldsymbol{d}_k^T \boldsymbol{H} \boldsymbol{d}_k}$$

Η βασική διαφορά από τις άλλες μεθόδους είναι το ότι επιλέγουμε τόσο την κατεύθυνση όσο και το βήμα εκπαίδευσης ταυτόχρονα. Έτσι αφού οριστεί το βήμα εκπαίδευσης, επιλέγουμε μια κατεύθυνση $\mathbf{d_k}$ και κινούμαστε προς τα εκεί για να βρούμε το ελάχιστο. Αυτός ο τρόπος σκέψης είναι ίδιος με αυτόν της αναζήτησης ελαχίστου σε ευθεία. Όμως ο τρόπος επιλογής του d_k μοιάζει με τη μέθοδο της ορμής, καθώς το $\mathbf{d_{k+1}}$ είναι το άθροισμα της αρνητικής κλίσης $-\mathbf{g_{k+1}}$ στο σημείο **w**(k+1) συν την προηγούμενη κατεύθυνση διόρθωσης $\mathbf{d_k}$ πολλαπλασιασμένη επί μια παράμετρο ορμής μ_κ.

$$\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \boldsymbol{\mu}_k \boldsymbol{d}_k, \qquad \mu\varepsilon\ \boldsymbol{d}_1 = -\boldsymbol{g}_1$$

Αν η παράμετρος ορμής είναι μηδέν έχουμε την απλή κατάβαση δυναμικού. Αν είναι σταθερή και διάφορη του μηδενός έχουμε την απλή μέθοδο της ορμής. Ωστόσο η μέθοδος συζυγούς κατάβασης δυναμικού υπολογίζει σε κάθε επανάληψη τη βέλτιστη τιμή της παραμέτρου της ορμής η οποία δεν είναι σταθερή. Δηλαδή η μέθοδος αυτή αποτελεί την γενικότερη περίπτωση των άλλων δύο μεθόδων και είναι λογικά υπερκείμενη αυτών.

Συνοψίζοντας τα παραπάνω είναι προφανές πως για την υλοποίηση της μεθόδου είναι απαραίτητος ο υπολογισμός του βήματος εκπαίδευσης β και του όρου της ορμής μ_κ.

Υπάρχουν διάφοροι τύποι υπολογισμού της ορμής, οι οποίοι παρακάμπτουν τον υπολογισμό του πίνακα Hess.

Αυτοί είναι:

- **Hestenes-Stiefel**

$$\mu_k = \frac{\boldsymbol{g_{k+1}^T}(\boldsymbol{g_{k+1}} - \boldsymbol{g_k})}{\boldsymbol{d_k^T}(\boldsymbol{g_{k+1}} - \boldsymbol{g_k})}$$

- **Polack-Ribiere:**

$$\mu_k = \frac{\boldsymbol{g_{k+1}^T}(\boldsymbol{g_{k+1}} - \boldsymbol{g_k})}{\boldsymbol{g_k^T g_k}}$$

- **Fletcher-Reeves:**

$$\mu_k = \frac{\boldsymbol{g_{k+1}^T g_{k+1}}}{\boldsymbol{g_k^T g_k}}$$

Οι συναρτήσεις είναι ισοδύναμες σε περιπτώσεις όπου η συνάρτηση κόστους είναι τετραγωνικής μορφής. Όμως στα δίκτυα MLP όπου κάτι τέτοιο συμβαίνει μόνο τοπικά, ο τύπος των Polack-Ribiere δίνει ελαφρώς καλύτερα αποτελέσματα.

Τέλος το β_k υπολογίζεται ελαχιστοποιώντας την τετραγωνική συνάρτηση που δόθηκε στην αρχή της παραγράφου, απλά αντικαθιστώντας το **w=w(k)+βd$_k$**. Η συνάρτηση είναι πλέον μιας μεταβλητής J(β) και συνεπώς έχουμε το πρόβλημα αναζήτησης ελαχίστου σε ευθεία όπως αυτό παρουσιάστηκε στην προηγούμενη ενότητα. Έτσι αποφεύγουμε πλήρως τον υπολογισμό του πίνακα του Hess.

6.2 SUPPORT VECTOR MACHINES

Τα SVMs είναι ταξινομητές οι οποίοι συνδυάζουν μικρούς χρόνους εκπαίδευσης με ικανότητα λύσης προβλημάτων που υφίστανται σε μη γραμμικούς χώρους (πολλαπλών διαστάσεων). Άρα σε ότι αφορά τον τομέα της ταξινόμησης προτύπων, συνδυάζουν τα πλεονεκτήματα τόσο του Perceptron όσο και του MLP, δίχως τα μειονεκτήματα τους. Σκοπός τους είναι να βρουν τη διαχωριστική ευθεία η οποία έχει την ελάχιστη απόσταση από τα συνοριακά πρότυπα κάθε κατηγορίας. Η διαχωριστική αυτή ευθεία ονομάζεται βέλτιστη διαχωριστική ευθεία. Όμως τα SVM μπορούν να λύσουν και προβλήματα περισσοτέρων διαστάσεων, εφ' όσον αυτά έχουν λύση. Για το λόγο αυτό λέμε πως γενικά έχουν τη δυνατότητα να υπολογίζουν τη βέλτιστη διαχωριστική επιφάνεια (καθώς ο όρος ευθεία αφορά στις 2 διαστάσεις μόνο, ενώ ο όρος επιφάνεια είναι η γενικότερη έννοια). Το κριτήριο σύμφωνα με το οποίο επιλέγουν τη βέλτιστη λύση και μας την παρουσιάζουν είναι η μεγιστοποίηση του περιθωρίου ταξινόμησης γ, μεταξύ των 2 κατηγοριών. Το περιθώριο ταξινόμησης ορίζεται ως το άθροισμα

$$\gamma = \gamma_0 + \gamma_1,$$

μεταξύ των δύο παρακάτω περιθωρίων, όπου γ_0 και γ_1 ορίζονται ως:

$$\gamma_0 = \min_{\boldsymbol{x} \in C_0} \frac{|\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\mathbf{0}}|}{\|\boldsymbol{w}\|} = \min_{\boldsymbol{x} \in C_0} \frac{-(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\mathbf{0}})}{\|\boldsymbol{w}\|},$$

$$\gamma_1 = \min_{\boldsymbol{x} \in C_1} \frac{|\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\mathbf{0}}|}{\|\boldsymbol{w}\|} = \min_{\boldsymbol{x} \in C_1} \frac{(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\mathbf{0}})}{\|\boldsymbol{w}\|},$$

Τα μεγέθη αυτά ονομάζονται περιθώρια γιατί παρατηρούμε πως η τιμή $|\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\mathbf{0}}|/\|\boldsymbol{w}\|$ είναι ευθέως ανάλογη με την απόσταση του διανύσματος **x** από τη γραμμική διαχωριστική επιφάνεια

$$\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\mathbf{0}} = 0$$

Άρα οι αριθμοί γ_0, γ_1 είναι θετικοί και αντιστοιχούν στην απόσταση του πιο κοντινού προτύπου της κάθε κατηγορίας από τη διαχωριστική επιφάνεια. Όσο οι τιμές των αποστάσεων πλησιάζουν στο 0 τόσο πιο οριακή θεωρείται η ταξινόμηση τους, καθώς ένα μικρό σφάλμα από ύπαρξη θορύβου μπορεί να προκαλέσει τη μετακίνηση ενός προτύπου που βρίσκεται κοντά στο σύνορο, στην άλλη κατηγορία.

Τα πρότυπα από κάθε κατηγορία **x** και **x'** τα οποία βρίσκονται πιο κοντά στην ευθεία διαχωρισμού και συνεπώς καθορίζουν και την ελάχιστη απόσταση από αυτή, δηλαδή τα διανύσματα για τα οποία ισχύει $|\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\mathbf{0}}|/\|\boldsymbol{w}\| = \gamma_0$ και $|\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x}' + w_{\mathbf{0}}|/\|\boldsymbol{w}\| = \gamma_1$, ονομάζονται **διανύσματα υποστήριξης (support vectors)**. Με βάση τα παραπάνω, μπορούμε

να πούμε πως το συνολικό περιθώριο ταξινόμησης γ αποτελεί ένα μέτρο εγγύτητας των δύο κλάσεων.

Η εξίσωση του υπερεπιπέδου εμπεριέχει αρκετό πλεονασμό. Για παράδειγμα, τα ζεύγη ($\alpha w, \alpha w_0$) και (w, w_0) περιγράφουν ακριβώς το ίδιο υπερεπίπεδο, και αυτό ισχύει για κάθε σταθερά α. Επίσης η τιμή γ όπως ορίστηκε προηγουμένως είναι ανεξάρτητη της κλιμάκωσης α. Ακόμα, όπως είναι γνωστό από την άλγεβρα, το κατώφλι w_0 είναι υπεύθυνο για την παράλληλη μετατόπιση του διαχωριστικού υπερεπιπέδου, ενώ το διάνυσμα w για την κατεύθυνση (κλίση) του υπερεπιπέδου. Με βάση τα παραπάνω, ονομάζουμε κανονικό διαχωριστικό υπερεπίπεδο αυτό για το οποίο ισχύει:

- το κατώφλι w_0 το τοποθετεί ακριβώς στη μέση ανάμεσα στις δύο κλάσεις, οπότε $\gamma_0=\gamma_1$, και
- η κλιμάκωση των **w** και w_0 είναι τέτοια ώστε

$$\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_0 \begin{cases} \leq -1, & \alpha\nu\ \boldsymbol{x} \in C_0 \\ \geq 1, & \alpha\nu\ \boldsymbol{x} \in C_1 \end{cases}$$

Συνεπώς για ένα κανονικό διαχωριστικό υπερεπίπεδο έχουμε

$$\min_{x \in C_0} |\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\boldsymbol{0}}| = \min_{x \in C_1} |\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + w_{\boldsymbol{0}}| = 1, \text{ και}$$

$$\gamma_0=\gamma_1=1/||\mathbf{w}||$$

οπότε το περιθώριο ταξινόμησης δίνεται από τη σχέση,

$$\gamma = \frac{2}{\|w\|}$$

Έστω ότι διαθέτουμε P ζεύγη προτύπων, στόχων: $(x_1,d_1), \dots ,(x_P,d_P)$, όπου $d_i=-1$ αν x_i ανήκει στη C_0 και $d_i=1$ αν x_i ανήκει στη C_1. Τότε οι ανισότητες που περιγράφουν το κανονικό διαχωριστικό υπερεπίπεδο απλοποιούνται στην εξής μορφή:

$$d_i(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x_i} + w_0) \geq 1, \qquad i = 1, \dots , P$$

η οποία αποτελεί και το βασικό περιορισμό του προβλήματος που καλούμαστε να λύσουμε.

Άρα, η καταλληλότερη λύση του προβλήματος είναι το κανονικό διαχωριστικό υπερεπίπεδο της παραπάνω μορφής το οποίο μεγιστοποιεί το περιθώριο γ, ή διαφορετικά ελαχιστοποιεί τη νόρμα $||\mathbf{w}||$ ή $||\mathbf{w}||^2$. Έτσι η αναζήτηση της καταλληλότερης λύσης μετατρέπεται στο παρακάτω πρόβλημα βελτιστοποίησης:

Υπολόγισε το ελάχιστο της συνάρτησης:

$$J(w, w_0) = \frac{1}{2}\|w\|^2$$

υπό τους περιορισμούς των P ανισοτήτων που δίνονται προηγουμένως.

Συνεπώς η εύρεση των διανυσμάτων υποστήριξης και κατ' επέκταση η όλη λογική λειτουργίας των SVM, αποτελεί ένα πρόβλημα τετραγωνικού προγραμματισμού. Η ελαχιστοποίηση της τετραγωνικής συνάρτησης υπό τους περιορισμούς γίνεται με τη χρήση των πολλαπλασιαστών Lagrange, δημιουργώντας την παρακάτω συνάρτηση κόστους Lagrange:

$$L(\boldsymbol{w}, w_0, \lambda_1, \dots, \lambda_P) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{P} \lambda_i \left[d_i(\boldsymbol{w^T x_i} + w_0) - 1\right]$$

Η συνάρτηση L πρέπει να ελαχιστοποιηθεί ως προς τα w και να μεγιστοποιηθεί ως προς τα λ_i. Στο βέλτιστο σημείο πρέπει να ισχύουν οι παρακάτω συνθήκες, οι οποίες ονομάζονται συνθήκες Karush-Kuhn-Tucker:

$$\frac{\partial L}{\partial w_0} = 0, \frac{\partial L}{\partial w} = 0$$

και

$$\lambda_i[d_i(\boldsymbol{w^T x_i} + w_0) - 1] = 0, \qquad i = 1, \dots, P$$

οι οποίες μεταφράζονται στις παρακάτω εξισώσεις:

$$\sum_{i=1}^{P} \lambda_i \, d_i = 0$$

$$\boldsymbol{w} = \sum_{i=1}^{P} \lambda_i \, d_i \boldsymbol{x_i}$$

Από αυτές υπολογίζεται η βέλτιστη τιμή του w, η οποία καθορίζει με βάση τις αντίστοιχες προαναφερθείσες σχέσεις τη βέλτιστη γραμμική διαχωριστική επιφάνεια:

$$g^*(x) = \boldsymbol{w^T x_i} + w_0 = \sum_{i=1}^{P} \lambda_i \, d_i \boldsymbol{x_i x} + w_0$$

Επίσης μπορούμε πλέον να υπολογίσουμε τη βέλτιστη πόλωση w_0 με βάση την (79). Πράγματι, αν το x_i είναι ένα διάνυσμα υποστήριξης τότε:

$$d_i(\boldsymbol{w^T x_i} + w_0) = 1$$

και ύστερα από πράξεις έχουμε:

$$w_0 = \frac{1}{d_i} - \boldsymbol{w^T x_i}$$

Για λόγους αριθμητικής ευστάθειας η εκτίμηση της πόλωσης w_0 είναι καλύτερα να υπολογίζεται από τον μέσο όρο της προηγούμενης εξίσωσης για όλα τα διανύσματα υποστήριξης, δηλαδή:

$$w_0 = \frac{1}{|I_{sv}|} \sum_{i \in I_{sv}} (\frac{1}{d_i} - \boldsymbol{w^T x_i})$$

όπου ορίσαμε το σύνολο

$$I_{sv} = \{i: x_i = \delta\iota\acute{\alpha}\nu\upsilon\sigma\mu\alpha\ \upsilon\pi o\sigma\tau\acute{\eta}\rho\iota\xi\eta\varsigma\}$$

Παρατηρούμε πως οι μόνοι πολλαπλασιαστές λ_i που μπορούν να είναι διάφοροι του μηδενός (και θετικοί αφού $\lambda_i \geq 0$ ούτως ή άλλως) είναι αυτοί που αντιστοιχούν σε κάποιο διάνυσμα υποστήριξης x_i. Για όλα τα υπόλοιπα διανύσματα προτύπων, που αποτελούν όπως είναι εύκολα κατανοητό και τη συντριπτική πλειοψηφία, ισχύει ότι $\lambda_i=0$. Συνεπώς το διάνυσμα λύσης **w**, *είναι ένας θετικός γραμμικός συνδυασμός των διανυσμάτων υποστήριξης.* Αυτή αποτελεί μια πολύ σημαντική παρατήρηση γενικά, αλλά το πρόβλημα είναι πως δε γνωρίζουμε από την αρχή ποια είναι τα διανύσματα υποστήριξης, δηλαδή το σύνολο I_{sv}, καθώς αυτό γίνεται γνωστό ύστερα από τη λύση του προβλήματος.

Με βάση τα παραπάνω έχουμε:

$$\frac{1}{2}\|w\|^2 = \frac{1}{2} w^T w = \frac{1}{2} \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_i \lambda_j d_i d_j x_i^T x_j$$

και

$$\sum_{i=1}^{P} \lambda_i [d_i (w^T x_i + w_0) - 1] = \sum_{i=1}^{P} \lambda_i d_i \sum_{j=1}^{P} \lambda_j d_j x_j^T x_i + w_0 \sum_{i=1}^{P} \lambda_i d_i - \sum_{i=1}^{P} \lambda_i$$
$$= \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_i \lambda_j d_i d_j x_i^T x_j - \sum_{i=1}^{P} \lambda_i$$

Συνεπώς η συνάρτηση Lagrange γίνεται:

$$L(\lambda_1, \dots, \lambda_P) = \sum_{i=1}^{P} \lambda_i - \frac{1}{2} \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_i \lambda_j d_i d_j x_i^T x_j$$

και η συνάρτηση του δυϊκού προβλήματος είναι:

$$L^d(\lambda_1, \dots, \lambda_P) = \frac{1}{2} \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_i \lambda_j d_i d_j x_i^T x_j - \sum_{i=1}^{P} \lambda_i$$

Σχετικά με τη λύση του τετραγωνικού προβλήματος θα πρέπει με την ευκαιρία να τονισθεί ότι όσο αυξάνεται ο αριθμός των προτύπων, ο κλασσικός αλγόριθμος τετραγωνικού προγραμματισμού γίνεται απαγορευτικά αργός και επίσης οι ανάγκες σε μνήμη είναι τεράστιες. Αυτά τα προβλήματα έχουν ωθήσει στη δημιουργία αποδοτικότερων αλγορίθμων για την επίλυση τέτοιων προβλημάτων, οι οποίοι μπορούν να αναζητηθούν στην αντίστοιχη βιβλιογραφία.

Μη διαχωρίσιμα προβλήματα

Η μέθοδος SVM που έχει παρουσιαστεί μέχρι στιγμής είναι κατάλληλη για την επίλυση γραμμικά διαχωρίσιμων προβλημάτων. Βασίζεται στην υπόθεση ότι το πρόβλημα είναι γραμμικά διαχωρίσιμο και έτσι ορίζονται τόσο το κανονικό διαχωριστικό υπερεπίπεδο όσο και τα διανύσματα υποστήριξης. Ωστόσο, πολλά προβλήματα δεν είναι γραμμικά διαχωρίσιμα.

Για τις περιπτώσεις αυτές, η μέθοδος μπορεί να επεκταθεί για να δώσει λύσεις, εισάγοντας στο αρχικό πρόβλημα τετραγωνικού προγραμματισμού τις μεταβλητές χαλαρότητας ξ_i για κάθε έναν από τους περιορισμούς:

$$d_i(\boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{i}} + w_0) \geq 1 - \xi_i, \qquad i = 1, \dots, P$$

$$\xi_i \geq 0$$

Αν η μεταβλητή ξ_i είναι μεγαλύτερη του 1 το πρότυπο $\mathbf{x_i}$ ταξινομείται λανθασμένα. Άρα το άθροισμα όλων των μεταβλητών χαλαρότητας μας δίνει το άνω όριο του πλήθους των προτύπων που ταξινομούνται λάθος.

Αντίστοιχα, η αρχική συνάρτηση κόστους τροποποιείται ώστε να λάβει υπ' όψιν της τις νέες μεταβλητές:

$$J_{ns}(w, w_0) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{P} \xi_i$$

Ο συνδυασμός της νέας συνάρτησης κόστους και των περιορισμών όπως αυτά δόθηκαν παραπάνω μορφώνει το νέο πρόβλημα τετραγωνικού προγραμματισμού το οποίο βρίσκει το βέλτιστο διαχωριστικό υπερεπίπεδο για τις μη διαχωρίσιμες κατηγορίες.

Η παράμετρος C είναι βάρος του κόστους των λανθασμένων ταξινομήσεων και επιλέγεται από το χρήστη. Για C=0 αγνοούμε τελείως τις παραμέτρους χαλαρότητας, και δε μας ενδιαφέρει αν έχουμε λανθασμένες ταξινομήσεις, ενώ για μεγάλες τιμές του C δίνουμε μεγάλη έμφαση. Η συνάρτηση κόστους τροποποιείται αντίστοιχα ώστε να συμπεριλαμβάνει τις μεταβλητές ξ_i και αυτό γίνεται με την εισαγωγή των πολλαπλασιαστών Lagrange μ_i. Αντίστοιχα τροποποιούνται και οι συνθήκες Karush-Kuhn-Tucker. Η νέα συνάρτηση κόστους της μορφής Lagrange είναι:

$$L_{ns}(w, w_0, \lambda_1, \dots, \lambda_P, \xi_1, \dots, \xi_P) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{P}\xi_i - \sum_{i=1}^{P}\lambda_i[d_i(w^T x_i + w_0) - 1 + \xi_i] - \sum_{i=1}^{P}\mu_i\xi_i$$

και η οποία πρέπει να μεγιστοποιηθεί ως προς τις παραμέτρους λ_i.

Οι συνθήκες Karush-Kuhn-Tucker είναι:

$$\frac{\partial L_{ns}}{\partial w} = w - \sum_{i=1}^{P}\lambda_i d_i x_i = 0$$

$$\frac{\partial L_{ns}}{\partial w_0} = -\sum_{i=1}^{P}\lambda_i d_i = 0$$

$$\frac{\partial L_{ns}}{\partial \xi_i} = C - \lambda_i - \mu_i = 0$$

$$\frac{\partial L_{ns}}{\partial \lambda_i} = d_i(w^T x_i + w_0) - 1 + \xi_i \geq 0$$

$$\frac{\partial L_{ns}}{\partial \mu_i} = \xi_i \geq 0$$

$$\lambda_i, \mu_i \geq 0$$

$$\lambda_i\{d_i(w^T x_i + w_0) - 1 + \xi_i\} = 0$$

$$\mu_i\xi_i = 0$$

Η συνάρτηση Lagrange του δυϊκού προβλήματος είναι:

$$L_{ns}^{d} = -L_{ns} = \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P}\lambda_i\,\lambda_j\,d_i d_j \boldsymbol{x}_i{}^T\boldsymbol{x}_j - \sum_{i=1}^{P}\lambda_i$$

η οποία πρέπει να ελαχιστοποιηθεί ως προς τις παραμέτρους λ_i και είναι αυτή που τελικά χρησιμοποιείται για την επίλυση του προβλήματος, μαζί με τους περιορισμούς Karush-Kuhn-Tucker και τον $0 \le \lambda_i \le C$ για i=1,… ,P. "Όμως πως καταφέρνει ένα SVM να λύσει αυτού του τύπου τα προβλήματα;

Η όλη μεθοδολογία βασίζεται στο θεώρημα του Cover, σύμφωνα με το οποίο κάθε πολυδιάστατος χώρος με μη γραμμικά διαχωρίσιμα πρότυπα, μπορεί να μετασχηματιστεί σε ένα νέο χώρο στον οποίο τα πρότυπα είναι γραμμικά διαχωρίσιμα με υψηλή πιθανότητα, αρκεί ο μετασχηματισμός να είναι μη γραμμικός και ο νέος αυτός χώρος να έχει την απαραίτητη διάσταση. Η συνάρτηση μετασχηματισμού είναι η Φ(.) και μετά την εφαρμογή της, τα πρότυπα έχουν περιέλθει σε ένα χώρο όπου είναι γραμμικά διαχωρίσιμα. Έτσι ακολουθείται ακριβώς η λογική επίλυσης που παρουσιάστηκε για τα γραμμικά διαχωρίσιμα προβλήματα μόνο που κάθε πρότυπο x_i αντικαθίσταται από το $\Phi(x_i)$.

Η βέλτιστη διαχωριστική επιφάνεια είναι:

$$g^*(\boldsymbol{x}) = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Phi}(\boldsymbol{x}) + w_0 = \sum_{i=1}^{P}\lambda_i d_i \boldsymbol{\Phi}(\boldsymbol{x}_i)^T\boldsymbol{\Phi}(\boldsymbol{x}) + w_0$$

το κατώφλι είναι:

$$w_0 = \frac{1}{|I_{sv}|}\sum_{i\in I_{sv}}\left(\frac{1}{d_i} - \sum_{j=1}^{P}\lambda_j d_j \boldsymbol{\Phi}(\boldsymbol{x}_j)^T\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}})\right)$$

και η συνάρτηση κόστους είναι:

$$L(\lambda_1,\dots,\lambda_P) = \sum_{i=1}^{P}\lambda_i - \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P}\lambda_i\,\lambda_j\,d_i d_j \boldsymbol{\Phi}(\boldsymbol{x}_i)^T\boldsymbol{\Phi}\left(\boldsymbol{x}_{\boldsymbol{j}}\right)$$

Είναι προφανές πως σε όλες τις εξισώσεις εμφανίζονται εσωτερικά γινόμενα της μορφής $\Phi(x)^T\Phi(y)$, αλλά η συνάρτηση μετασχηματισμού Φ(.) ποτέ δεν εμφανίζεται μόνη της. Λόγω αυτού ορίζουμε τη συνάρτηση

$$k(\mathbf{x},\mathbf{y}) = \Phi(\mathbf{x})^T\,\Phi(\mathbf{y})$$

η οποία ονομάζεται συνάρτηση πυρήνα.

Με χρήση αυτής της συνάρτησης μπορούν να απλοποιηθούν αρκετά οι πράξεις καθώς τα Φ(x) και Φ(y) μπορεί να έχουν και άπειρες διαστάσεις, ενώ η k(.,.) υπολογίζεται γενικά σε μικρό αριθμό πράξεων. Επίσης δε χρειάζεται να γνωρίζουμε ακριβώς ποια είναι η συνάρτηση μετασχηματισμού (κάτι που για τα περισσότερα προβλήματα είναι πολύ δύσκολο), αλλά να γνωρίζουμε απλά ότι υπάρχει έστω και σε άπειρες διαστάσεις. Αυτό που μας ενδιαφέρει είναι να χρησιμοποιήσουμε την καταλληλότερη, για την περίπτωση που μας ενδιαφέρει, συνάρτηση πυρήνα. Η πιο συνηθισμένες συναρτήσεις πυρήνα είναι η Γκαουσιανή RBF, η πολυωνυμική, η σιγμοειδής, η αντίστροφη πολυτετραγωνική κ.α.. Συνεπώς ο σκοπός του εκάστοτε σχεδιαστή είναι να βρει ποια συνάρτηση πυρήνα k, ταιριάζει καλύτερα στο είδος του προβλήματος που αντιμετωπίζει, ώστε να τη χρησιμοποιήσει για να επιτύχει την μέγιστη ακρίβεια ταξινόμησης.

Με βάση τα παραπάνω μπορούμε να πούμε πως τα SVM μπορούν να χρησιμοποιηθούν με επιτυχία στην ταξινόμηση τόσο των γραμμικά διαχωρίσιμων κατηγοριών όσο και των μη γραμμικά διαχωρίσιμων, με τις εξής μετατροπές στις παραμέτρους:

- Στις γραμμικά διαχωρίσιμες χρησιμοποιείται η γραμμική συνάρτηση πυρήνα $k(x,y) = x^T y$ (δηλαδή δε χρησιμοποιεί συνάρτηση πυρήνα) και ο περιορισμός $\lambda_i \leq C$ δεν υπάρχει, ή διαφορετικά C άπειρο.
- Στις μη γραμμικά διαχωρίσιμες κατηγορίες χρησιμοποιείται μια μη γραμμική συνάρτηση της επιλογής μας και ο περιορισμός $\lambda_i \leq C$ με το όριο C να καθορίζεται από το χρήστη και να υποδηλώνει τη βαρύτητα που δίνουμε στις λάθος ταξινομήσεις. Μεγάλες τιμές σημαίνει πως απαιτούμε λίγες λάθος ταξινομήσεις, ενώ μικρές σημαίνει πως έχουμε χαλαρότερα κριτήρια.

Όπως αναφέραμε και προηγουμένως, τα SVM καλούνται να λύσουν ένα πρόβλημα τετραγωνικού προγραμματισμού το οποίο στην πιο γενική του μορφή είναι:

Υπολόγισε το μέγιστο της συνάρτησης:

$$L_{SVM}(\lambda_1, \dots, \lambda_P) = \sum_{i=1}^{P} \lambda_i - \frac{1}{2} \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_i q_{ij} \lambda_j$$

υπό τους περιορισμούς:

$$0 \leq \lambda_i \leq C \text{ και } \sum_{i=1}^{P} \lambda_i \, d_i = 0$$

όπου τα στοιχεία q_{ij} είναι της μορφής $q_{ij}=d_i d_j k(\boldsymbol{x_i},\boldsymbol{x_j})$.

Το μεγάλο μειονέκτημα των SVM είναι ο τεράστιος υπολογιστικός φόρτος καθώς το μέγεθος του πίνακα $Q=[q_{ij}]$ είναι πολύ μεγάλο. Για το λόγο αυτό έχουν προταθεί διάφορες βελτιστοποιήσεις με σκοπό την επιτάχυνση της μεθόδου, όπως η μέθοδος του τεμαχισμού, η μέθοδος του Osuna και το SVMlight, η μέθοδος ακολουθιακής ελάχιστης βελτιστοποίησης κ.α. Οι μέθοδοι αυτές είναι αρκετά επιτυχημένες αν και όχι τέλειες και για περαιτέρω λεπτομέρειες μπορεί κανείς να απευθυνθεί στην αντίστοιχη βιβλιογραφία.

Συνοψίζοντας τις παραπάνω γνώσεις μπορούμε να πούμε πως τα SVM είναι πολύ αποδοτικά στην επίλυση προβλημάτων ταξινόμησης οποιουδήποτε τύπου, χάρη στην ικανότητα τους να αλλάζουν χώρο διαστάσεων μέχρι να βρουν εκείνο στον οποίο τα πρότυπα καθίστανται γραμμικά διαχωρίσιμα. Μάλιστα, σε αντίθεση με τα νευρωνικά, η διαστασιμότητα του μοντέλου είναι ανεξάρτητη της πολυπλοκότητας του και συνεπώς μπορούμε να μεταβούμε σε χώρους πολύ μεγάλων διαστάσεων εντελώς φυσικά και χωρίς δυσκολία. Το μεγάλο τους μειονέκτημα είναι το υπολογιστικό κόστος καθώς αύξηση της διαστασιμότητας συνεπάγεται και αύξηση των πράξεων αλλά και των αναγκών σε χώρους μνήμης για τη διατήρηση των αποτελεσμάτων. Το μειονέκτημα αυτό αντιμετωπίζεται με τη χρήση των πυρήνων και την επίλυση του δυϊκού προβλήματος σε πρώτη φάση, αλλά και με τη χρήση βελτιστοποιημένων εκδόσεων του αρχικού αλγορίθμου SVM.

6.3 K-MEANS

Η μέθοδος K-Means αποτελεί μια διαδικασία ανάλυσης προτύπων σε ομάδες, που στόχο έχει την ταξινόμηση n προτύπων $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$ σε k ομάδες $X_1, X_2, \ldots, X_k$, όπου n>k. Είναι μη-επιβλεπόμενη μέθοδος μάθησης και είναι επαναληπτικής φύσης. Σε κάθε ομάδα αντιστοιχεί ένα κέντρο το οποίο και την εκπροσωπεί. Ο αριθμός των τελικών ομάδων είναι προκαθορισμένος από το χρήστη. Για κάθε πρότυπο, βρίσκουμε την ευκλείδεια απόσταση από όλα τα κέντρα των ομάδων και τελικά τοποθετούμε το πρότυπο στην ομάδα από την οποία η απόσταση αυτή είναι η μικρότερη. Με άλλα λόγια, κάθε πρότυπο τοποθετείται στην ομάδα για την οποία απέχει λιγότερο από το κέντρο της. Ξεκινά με τυχαία αρχικά διανύσματα κέντρων και ουσιαστικά προσπαθεί να βρει την ισορροπία μεταξύ των δύο κάτωθι απαιτήσεων:

1. για κάθε i=1,…,K, η κλάση X_i αποτελείται από τα πρότυπα $\mathbf{x}^{(P)}$ τα οποία βρίσκονται πιο κοντά στο διάνυσμα c_i σε σχέση με όλα τα άλλα διανύσματα c_j , j≠I, δηλαδή $\|\boldsymbol{x}^{(p)} - \boldsymbol{c}_i\| = \min_j \|\boldsymbol{x}^{(p)} - \boldsymbol{c}_j\|$
2. για κάθε i=1,…, K, το διάνυσμα $\mathbf{c}_i$ είναι ο μέσος όρος των προτύπων που ανήκουν σε κάθε κλάση X_i, δηλαδή:

$$\boldsymbol{c_i} = \frac{1}{|X_i|} \sum_{p \in X_i} \boldsymbol{x}^{(\boldsymbol{P})}$$

όπου $|X_i|$ είναι το πλήθος των προτύπων της κλάσης. Το διάνυσμα $\mathbf{c_i}$ λέγεται κέντρο της κλάσης.

Τα κέντρα αρχικά θα πρέπει να απέχουν αρκετά μεταξύ τους. Μόλις ολοκληρωθεί η διαδικασία την πρώτη φορά για όλα τα πρότυπα, μπορούμε να πούμε πως έχει γίνει μια πρώιμη κατηγοριοποίηση. Με βάση αυτή, επαναϋπολογίζονται τα κέντρα ώστε να αποτελούν πραγματικούς μέσους των δημιουργηθέντων από το προηγούμενο βήμα υποσυνόλων και για τα νέα κέντρα επαναλαμβάνεται η διαδικασία. Η ολοκλήρωση της μεθόδου επέρχεται όταν παρέλθει μια επανάληψη δίχως μετακίνηση κανενός μέσου, οπότε και το σύστημα έχει έρθει σε ισορροπία.

Προφανώς σκοπός της μεθόδου είναι η ελαχιστοποίηση μιας συνάρτησης, η οποία στην προκειμένη περίπτωση είναι το τετραγωνικό σφάλμα της απόστασης του εκάστοτε προτύπου από το κάθε κέντρο.

Παρόλο που ο αλγόριθμος πάντοτε συγκλίνει, δεν είναι απαραίτητο ότι το σημείο ισορροπίας του αποτελεί και τη βέλτιστη γενική λύση. Δηλαδή συχνά συγκλίνει σε τοπικά ελάχιστα. Προφανώς, ο σημαντικότερος παράγοντας είναι η αρχική επιλογή των κέντρων, επιλογή στην οποία παρουσιάζει μεγάλη ευαισθησία. Για την αντιμετώπιση αυτού του προβλήματος έχει προταθεί η πολλαπλή εφαρμογή του αλγορίθμου στο εκάστοτε σύνολο, ώστε να δημιουργηθεί μια πιο γενική αντικειμενική εικόνα για το που ανήκει το κάθε στοιχείο. Επίσης επειδή η όλη μέθοδος στηρίζεται σε υπολογισμό αποστάσεων, η ακρίβεια αυτών παίζει σημαντικό ρόλο. Άρα καλό είναι να χρησιμοποιούνται κανονικοποιημένα δεδομένα καθώς αυτό διευκολύνει στον ακριβέστερο υπολογισμό των αποστάσεων. Ένα άλλο μειονέκτημα διαφορετικού είδους, είναι η αδυναμία του αλγορίθμου να λάβει υπ’ όψιν του οποιαδήποτε πληροφορία σχετικά με την εντροπία των δεδομένων, ή διαφορετικά τη συσχέτισή τους, ώστε να επιτύχει καλύτερα αποτελέσματα.

6.4 Fuzzy C-Means Clustering (FCM)

Η έννοια του fuzzy clustering μεταφράζεται στα ελληνικά ως ασαφής κατηγοριοποίηση. Βασίζεται στη γενικότερη φιλοσοφία της ασαφούς λογικής (fuzzy logic), και συνεπώς ένα βασικό χαρακτηριστικό της είναι ότι είναι «μαλακού» τύπου κατηγοριοποίηση. Αυτό πρακτικά σημαίνει ότι κάθε σημείο ανήκει μερικώς (σε διαφορετικό βαθμό) σε πολλές κατηγορίες, αντί να ανήκει ολοκληρωτικά σε μία και μοναδική. Συνεπώς τα σημεία στα άκρα που βρίσκονται στα άκρα μιας κατηγορίας μπορούν να ανήκουν σε αυτή, απλά σε *μικρότερο βαθμό* σε σχέση

με τα σημεία που βρίσκονται στο κέντρο της. Ταυτόχρονα όμως ανήκουν και στις γειτονικές κατηγορίες, σε κάποιο βαθμό. Για κάθε σημείο λοιπόν x, έχουμε ένα συντελεστή $u_k(x)$ που προσδιορίζει το βαθμό κατά τον οποίο ανήκει το σημείο στην κατηγορία k. Το άθροισμα όλων των συντελεστών για κάθε σημείο έχει τιμή 1.

$$\forall x \left(\sum_{k=1}^{\alpha\rho.\kappa\alpha\tau\eta\gamma o\rho\iota\acute{\omega}\nu} u_k(x) = 1 \right)$$

Στον αλγόριθμο αυτό, το κέντρο κάθε κατηγορίας προκύπτει από το μέσο όρο όλων των σημείων, πολλαπλασιασμένα με το βαθμό κατά τον οποίο κάθε ένα ανήκει στη συγκεκριμένη κατηγορία:

$$c_k = \frac{\sum_x u_k(x)^m \, x}{\sum_x u_k(x)^m}$$

Ο συντελεστής u, ο οποία διαφορετικά μπορεί να ονομαστεί και βαθμός συμμετοχής του σημείου, σχετίζεται με το αντίστροφο της απόστασης του σημείου από το κέντρο της κατηγορίας:

$$u_k(x) = \frac{1}{d(c_k, x)}$$

Αφού υπολογιστούν όλοι οι βαθμοί συμμετοχής ενός σημείου κανονικοποιούνται και «ασαφοποιούνται» με τη χρήση μιας πραγματικής παραμέτρου *m*>1, ώστε το άθροισμα τους να είναι 1. Το πως πραγματοποιείται αυτό φαίνεται στον παρακάτω τύπο:

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(c_k, x)}{d(c_j, x)}\right)^{2/(m-1)}}$$

Για *m* ίσο με 2, η περίπτωση είναι παρόμοια με τη γραμμική κανονικοποίηση του διανύσματος συντελεστών κάθε σημείου, με σκοπό το άθροισμα των επιμέρους στοιχείων να γίνει 1. Όταν το *m* είναι κοντά στο 1, το κέντρο της κατηγορίας που βρίσκεται πιο κοντά στο σημείο λαμβάνει πολύ μεγαλύτερη βαρύτητα από τα υπόλοιπα, και ο αλγόριθμος ομοιάζει με τον k-means.

Γενικά ο FCM είναι παρόμοιος με τον K-means και μπορούμε να πούμε πως αποτελεί την ασαφή εκδοχή του δεύτερου. Έτσι και εδώ συμβαίνουν τα εξής:

1. Επιλέγεται ο αριθμός των κατηγοριών
2. Σε κάθε σημείο αντιστοιχίζονται τυχαίοι αρχικοί βαθμοί συμμετοχής

3. Υπολόγισε τα κέντρα των κατηγοριών
4. Με βάση τα νέα κέντρα υπολόγισε τους βαθμούς συμμετοχής κάθε σημείου
5. Επανέλαβε από το 3 μέχρι ο αλγόριθμος να συγκλίνει, δηλαδή σε μια επανάληψη να μη παρατηρηθεί καμία μεταβολή μεγαλύτερη από μια πολύ μικρή σταθερά ε, στις τιμές των συντελεστών.

Ο αλγόριθμος ελαχιστοποιεί την ενδό-κατηγοριακή ποικιλομορφία, αλλά πέρα από τις ομοιότητες έχει και τα ίδια προβλήματα με τον k-means. Το ελάχιστο είναι τοπικό ελάχιστο, και τα αποτελέσματα εξαρτώνται σε μεγάλο βαθμό από την αρχική επιλογή βαρών. Σαν εναλλακτική λύση προτείνεται η εφαρμογή του αλγορίθμου expectation-maximization, ο οποίος έχει καλύτερες ιδιότητες σύγκλισης διατηρώντας το βασικό χαρακτηριστικό του FCM, που είναι η μερική συμμετοχή σε διάφορες κατηγορίες.

6.5 K-Nearest Neighbors Clustering (K-NN)

Στην αναγνώριση προτύπων, ο αλγόριθμος K-NN είναι μια μέθοδος κατηγοριοποίησης αντικειμένων με βάση τα *k* κοντινότερα σε αυτά πρότυπα στο χώρο των χαρακτηριστικών. Ο αλγόριθμος K-NN είναι ένα είδος μάθησης βασισμένο σε στιγμιότυπα, γνωστό και ως βαριεστημένη μάθηση όπου η συνάρτηση προσεγγίζεται μόνο τοπικά και όλοι οι υπολογισμοί αναβάλλονται μέχρι τη στιγμή της κατηγοριοποίησης. Αποτελεί έναν από τους απλούστερους αλγορίθμους της μηχανικής μάθησης· ένα αντικείμενο κατηγοριοποιείται με βάση την πλειοψηφία των γειτόνων του, με το αντικείμενο να οδεύει προς την κατηγορία που υπερισχύει ανάμεσα στους *k* κοντινότερους γείτονές του, (όπου *k* είναι ένας θετικός, γενικά μικρός ακέραιος). Εάν *k*=1, τότε το αντικείμενο απλά τοποθετείται στην κατηγορία του κοντινότερου γείτονα.

Η ίδια μέθοδος μπορεί να χρησιμοποιηθεί και για αναδρομή, απλά με το να αναθέσουμε την τιμή της ιδιότητάς του προτύπου στη μέση τιμή των τιμών των *k* κοντινότερων γειτόνων του. Μπορεί να είναι ωφέλιμη η χρήση βαρών στην συνεισφορά των γειτόνων, ώστε οι κοντινοί γείτονες να συνεισφέρουν περισσότερο στη μέση τιμή απ' ότι οι πιο απομακρυσμένοι. Ένας τυπικός τρόπος να πραγματοποιηθεί αυτό είναι να δοθεί σε γείτονα ένα βάρος 1/*d*, όπου *d* είναι η απόσταση από το ζητούμενο πρότυπο. Αυτό αποτελεί μια γενίκευση της γραμμικής παρεμβολής.

Ως γείτονες θεωρούνται εκείνα τα πρότυπα για τα οποία γνωρίζουμε την πραγματική τους κατηγορία, και από αυτό το υποσύνολο προέρχονται τα αποτελέσματα της μεθόδου. Αυτό μπορούμε να πούμε πως αποτελεί το υποσύνολο εκπαίδευσης του αλγορίθμου, παρόλο που δεν πλήρη τις αυστηρές προδιαγραφές τις τυπικής θεώρησης των συνόλων εκπαίδευσης (όπως ας πούμε την ύπαρξη συγκεκριμένου βήματος εκπαίδευσης). Ο αλγόριθμος είναι ευαίσθητος

στην τοπική δομή των δεδομένων. Το υποσύνολο εκπαίδευσης αποτελείται λοιπόν από διανύσματα στο πολυδιάστατο χώρο χαρακτηριστικών, κάθε ένα εκ των οποίο έχει και μια ταμπέλα που δίνει την πληροφορία για την κατηγορία στην οποία ανήκει.

Η φάση εκπαίδευσης συνίσταται από τον προσδιορισμό του υποσυνόλου εκπαίδευσης (γειτόνων) και την αποθήκευση των ανάλογων προτύπων και των αντίστοιχων πληροφοριών κατηγοριοποίησης τους. Στη φάση κατηγοριοποίησης, ένα άγνωστο πρότυπο κατηγοριοποιείται με το να τοποθετείται στην κατηγορία που ανήκει η πλειοψηφία από τους *k* κοντινότερους γείτονές του, όπου *k* μια σταθερά προκαθορισμένη από τον χρήστη.

Συνήθως η Ευκλείδεια απόσταση χρησιμοποιείται ως μέθοδος υπολογισμού των αποστάσεων, ωστόσο είναι εφαρμόσιμη μόνο σε συνεχείς μεταβλητές. Για τις διακριτές μεταβλητές μπορούμε να χρησιμοποιήσουμε την απόσταση Hamming ή άλλες αντίστοιχες μεθόδους.

Το μειονέκτημα της βασικής πλειοψηφικής κατηγοριοποίησης είναι πως οι κατηγορίες με τα πιο συχνά πρότυπα τείνουν να κυριαρχούν στην πρόβλεψη του αγνώστου προτύπου, καθώς είναι πιθανότερο να εμφανιστούν μέσα στους *k* κοντινότερους γείτονες, ειδικά αν το *k* δεν έχει μικρές τιμές. Ένας τρόπος να αντιμετωπιστεί αυτό είναι να τοποθετηθούν βάρη ώστε τα πρότυπα που βρίσκονται πιο κοντά (οι πιο κοντινοί γείτονες δηλαδή) να έχουν μεγαλύτερη αξία για την πρόβλεψη της κατηγορίας.

Η βέλτιστη τιμή του *k* εξαρτάται από τα δεδομένα. Γενικά μεγαλύτερες τιμές του *k* μειώνουν την επίδραση του θορύβου στην κατηγοριοποίηση, αλλά κάνουν πιο αδρές τις διαχωριστικές γραμμές μεταξύ των κατηγοριών. Μια καλή τιμή για το *k* μπορεί να βρεθεί χρησιμοποιώντας διάφορες ευριστικές μεθόδους. Στη συνηθισμένη περίπτωση των 2 κατηγοριών, όπως στην εργασία μας δηλαδή, είναι προτιμότερο να επιλέγουμε *k* μονό αριθμό ώστε να αποφεύγονται οι ισοψηφίες.

Η απόδοση του αλγορίθμου μπορεί να μειωθεί δραματικά από την παρουσία θορυβωδών ή άσχετων χαρακτηριστικών, γι’ αυτό και υπενθυμίζεται με την ευκαιρία η σημασία της ύπαρξης ενός υποσυστήματος επιλογής χαρακτηριστικών. Πιο αποτελεσματικές μέθοδοι φαίνεται να είναι αυτές που βασίζονται στους εξελικτικούς αλγορίθμους και στην αμοιβαία πληροφορία, αλλά κάθε περίπτωση είναι ξεχωριστή και πρέπει να αντιμετωπίζεται ανάλογα.

7. ΠΕΙΡΑΜΑΤΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

Εισαγωγή

Αφού πλέον τοποθετήθηκε το θεωρητικό υπόβαθρο πάνω στο οποίο στηρίχθηκε η εργασία, μπορούμε πλέον να περάσουμε στις λεπτομέρειες της υλοποίησης του συστήματος κατηγοριοποίησης. Ακολουθεί η μελέτη όλων των σχημάτων επιλογής χαρακτηριστικών και κατηγοριοποίησης που χρησιμοποιήθηκαν στην εργασία, σε μια προσπάθεια να αντιληφθούμε και να καταγράψουμε τον τρόπο που οι διάφορες παράμετροί τους επηρεάζουν την απόδοση του συστήματος μας. Ακόμα γίνεται μια μικρή αναφορά και επανάληψη των βασικών της κάθε υλοποίησης.

Να προσθέσουμε εδώ ότι η βασική συνάρτηση αξιολόγησης J που χρησιμοποιήθηκε για την επιλογή χαρακτηριστικών βασίστηκε στην κατηγοριοποίηση FCM. Αυτή αποτέλεσε τον πυρήνα του συστήματός μας. Διατηρήθηκε κοινή σε όλες τις περιπτώσεις ώστε να έχουμε ένα αντικειμενικό μέτρο σύγκρισης των διαφόρων μεθόδων. Όμως αυτό δεν έγινε τυχαία, διότι η FCM εμφάνισε την καλύτερη συμπεριφορά από πλευράς απόδοσης και χρόνου ολοκλήρωσης. Εκτός από αυτήν δοκιμάστηκαν ως πυρήνες νευρωνικά δίκτυα και SVM, τα οποία και ήταν πολύ πιο αργά από την FCM καθώς έπρεπε σε κάθε χρήση τους να εκπαιδεύονται, αλλά και στην καλύτερη περίπτωση παρουσίαζαν τις ίδιες επιδόσεις με αυτή. Για τους λόγους αυτούς κρίθηκε ασύμφορη η περαιτέρω χρήση τους.

Επίσης κατά τη λήψη των πειραματικών τιμών χρησιμοποιήθηκε στην προεπεξεργασία απλή κανονικοποίηση και όχι η παράγωγος με γκαουσιανό πυρήνα. Αφού βρεθεί η σύσταση των παραμέτρων θα δοκιμαστεί και η παράγωγος για να δούμε πως επιδρά.

7.1 Χαρακτηριστικά Συστήματος Υλοποίησης

Όλοι οι αλγόριθμοι και οι διαδικασίες υλοποιήθηκαν στο περιβάλλον της Matlab, έκδοση 7.6.0 (R2008a). Ο υπολογιστής που χρησιμοποιήθηκε για την πραγματοποίηση των εξομοιώσεων αποτελούνταν από έναν επεξεργαστή Intel Q9450@2.67GHz, είχε RAM 3.2 GB και έτρεχε λειτουργικό Windows XP SP3.

7.2 SFFS

Αρχικά υλοποιήθηκε ο FGA SFFS. Ο αριθμός των στοιχείων εκκίνησης που επιλέχθηκε ήταν το 2. Συνεπώς το σύστημα βρίσκει πρώτα το καλύτερο στοιχείο και στη συνέχεια βρίσκει το στοιχείο εκείνο το οποίο συνδυαζόμενο με το αρχικό δίνει το βέλτιστο υποσύνολο 2 στοιχείων.

Όπως προαναφέρθηκε και κατά τη θεωρητική ανάλυση τα υποσύνολα αυτά, λόγω αυτού του σειριακού τρόπου κατασκευής, είναι πολύ πιθανόν να μην είναι τα βέλτιστα για το μέγεθός τους.

Ύστερα ο αλγόριθμος εισέρχεται στο επαναληπτικό του στάδιο. Σε πρώτη φάση, με την ίδια λογική προσθέτει ένα νέο στοιχείο στο υπάρχον υποσύνολο. Στη συνέχεια ψάχνει να βρει το λιγότερο σημαντικό στοιχείο του υποσυνόλου, δηλαδή το στοιχείο εκείνο το οποίο όταν αφαιρεθεί από το υποσύνολο λύσης θα μειώσει σε μικρότερο βαθμό την τιμή της συνάρτησης αξιολόγησης J. Συνεπώς αφαιρούμε ένα στοιχείο τη φορά και παρατηρούμε την τιμή που δίνει η απουσία του στην J. Το στοιχείο που όταν αφαιρεθεί δίνει τη μεγαλύτερη J είναι και το στοιχείο που προτείνεται για αποχώρηση. Ο μόνος λόγος για να παραμείνει το στοιχείο είναι να είναι το στοιχείο που αμέσως έχει προστεθεί στο προηγούμενο βήμα. Σε αυτή την περίπτωση θεωρούμε πως η επανάληψη ολοκληρώθηκε και ξεκινά μια νέα. Παρακάτω, στο τρίτο και τελευταίο στάδιο, ο αλγόριθμος επαναλαμβάνει την ίδια αφαιρετική διαδικασία μόνο που αυτή τη φορά το υποψήφιο στοιχείο αποχωρεί μόνο αν η τιμή που δίνει στην J είναι μεγαλύτερη από την τιμή που είχε στην προηγούμενη επανάληψη, ειδάλλως ολοκληρώνεται η τρέχουσα επανάληψη και ξεκινά μια νέα.

Ακολουθούν σε συνοπτική μορφή τα προαναφερθέντα:

0. Δημιούργησε ένα υποσύνολο με 3 στοιχεία (ή k στοιχεία γενικά).
1. Πρόσθεσε ένα στοιχείο (k=k+1).
2. Αναγνώρισε το λιγότερο σημαντικό στοιχείο του νεοσυσταθέντος υποσυνόλου λύσης:
    - Αν είναι το στοιχείο που προστέθηκε στο αμέσως προηγούμενο βήμα, διατήρησε το και ολοκλήρωσε την επανάληψη επιστρέφοντας στο βήμα 1.
    - Αν είναι κάποιο από τα υπόλοιπα στοιχεία αφαίρεσε το κανονικά (k=k-1).
3. Για το μειωμένο υποσύνολο λύσης ξαναβρές το λιγότερο σημαντικό στοιχείο:
    - Αν η τιμή της J μετά την αφαίρεση του είναι μεγαλύτερη από την τιμή που είχε η J κατά την προηγούμενη επανάληψη τότε αφαίρεσε το υποψήφιο στοιχείο και επανέλαβε (k=k-1).
    - Διαφορετικά ολοκλήρωσε την επανάληψη και πήγαινε στο βήμα 1.

Σε κάθε περίπτωση ο αριθμός των στοιχείων του υποσυνόλου πρέπει να είναι μεγαλύτερος ή ίσος με τον αρχικό αριθμό, δηλαδή τον αριθμό των στοιχείων εκκίνησης (k πάντα μεγαλύτερο ή ίσο του 2 για τη συγκεκριμένη εργασία). Συνεπώς αν σε κάποιο από τα αφαιρετικά στάδια καταλήξει να έχει τη μικρότερη δυνατή τιμή τότε ο αλγόριθμος αυτόματα ολοκληρώνει την επανάληψη του και επιστρέφει στο βήμα 1.

Ο αλγόριθμος ολοκληρώνεται όταν συμπληρωθεί συγκεκριμένος αριθμός επαναλήψεων (εποχών), ο οποίος επιλέχθηκε να είναι το 30. Η τιμή αυτή αποτελεί ένα καλό ισοζύγιο μεταξύ ελευθερίας για καλύτερη έρευνα του χώρου των λύσεων και συνεπώς πραγματοποίηση καλύτερων επιλογών, και χρονικού-υπολογιστικού κόστους. Επίσης να τονίσουμε ότι πολύ γρήγορα ο αλγόριθμος έφθανε σε μια αποδεκτά καλή τιμή, οπότε το μεγαλύτερο μέρος του χρόνου του το περνά στην προσπάθεια του να βελτιστοποιήσει τοπικά την απόδοση.

Όπως είναι προφανές λοιπόν ο αλγόριθμος αυξάνει το μέγεθος του υποσυνόλου λύσης (τον αριθμό χαρακτηριστικών που συγκροτούν το υποσύνολο λύσης) μέχρι να βρει μια περιοχή αριθμών που να δίνει καλές τιμές στη συνάρτηση αξιολόγησης και ύστερα δίνει μεγαλύτερη έμφαση στην εύρεση της σύστασης του βέλτιστου υποσυνόλου σε αυτήν την περιοχή μεγεθών. Επίσης πρέπει να έχουμε υπ' όψιν πως ο αριθμός ελαχίστων στοιχείων είναι περισσότερο ένας δείκτης ελευθερίας του συστήματος να καθορίσει από μόνο του το βέλτιστο μέγεθός του, παρά μια αναγκαστική τιμή που πρέπει να έχει το βέλτιστο υποσύνολο(αυτό θα φανεί και στις μετρήσεις). Είναι συχνό φαινόμενο το μέγεθος του προτεινόμενου υποσυνόλου να είναι κοντά στο ελάχιστο ή το ελάχιστο, αλλά αυτό δε σημαίνει ότι είναι απίθανο να περιέχει πολλά παραπάνω στοιχεία, απλά όσο αυξάνεται αυτός ο αριθμός τόσο μειώνεται η πιθανότητα να αυξηθεί εκ νέου, καθώς είναι προφανές πως οι συνθήκες για μόνιμη αύξηση είναι πιο δύσκολα υλοποιήσιμες από αυτές για την πραγματοποίηση αφαίρεσης. Αυτό έρχεται σε αντίθεση με τον τρόπο λειτουργίας των γενετικών αλγορίθμων. Για το λόγο αυτό, επιλέχθηκε ως κριτήριο τερματισμού ο αριθμός των επαναλήψεων και όχι η ολοκλήρωση όταν αυξηθεί το μέγεθος του υποσυνόλου λύσης κατά ένα συγκεκριμένο αριθμό, +2 ή +3 ας πούμε από την αρχική τιμή. Συνεπώς όσο μικρότερος είναι αυτός ο αριθμός τόσο πιο ελεύθερο είναι το σύστημα να ψάξει το χώρο των λύσεων για την καλύτερη δυνατή απάντηση.

Οι παράμετροι του αλγορίθμου είναι:

- ο αριθμός των αρχικών στοιχείων (και ελαχίστων αποδεκτών στοιχείων)
- ο αριθμός των εποχών=30 (όπως εξηγήθηκε παραπάνω)

7.3 Παράμετροι SFFS

Το μέγεθος που μετρήσαμε και θέσαμε ως κριτήριο για την επίδραση της κάθε παραμέτρου στο σύστημα, είναι η μεγιστοποίηση της τιμής της συνάρτησης αξιολόγησης J(με απόλυτη μέγιστη τιμή το 1) του κάθε φορά προτεινόμενου υποσυνόλου από τον αλγόριθμο.

Αριθμός αρχικών/ελαχίστων στοιχείων

Η παράμετρος αυτή καθορίζει πόσα στοιχεία θα έχει το υποσύνολο τη στιγμή που αρχίζει πάνω του να εκτελείται ο αλγόριθμος. Ταυτόχρονα είναι και ο ελάχιστος αριθμός στοιχείων

που μπορεί να έχει το τελικό προτεινόμενο υποσύνολο λύσης, δηλαδή ο αλγόριθμος θα πρέπει να καταλήξει στο βέλτιστο υποσύνολο προκαθορισμένου ελάχιστου μεγέθους ή αν δεν ικανοποιείται από αυτό, να ψάξει μια λύση με περισσότερα στοιχεία. Τα αρχικά στοιχεία προστίθενται με σειριακά με βάση το πιο είναι το καλύτερο κάθε φορά σε σχέση με το ήδη υπάρχον υποσύνολο. Όπως έχει εξηγηθεί στο θεωρητικό κομμάτι αυτός ο τρόπος απέχει από τον βέλτιστο.

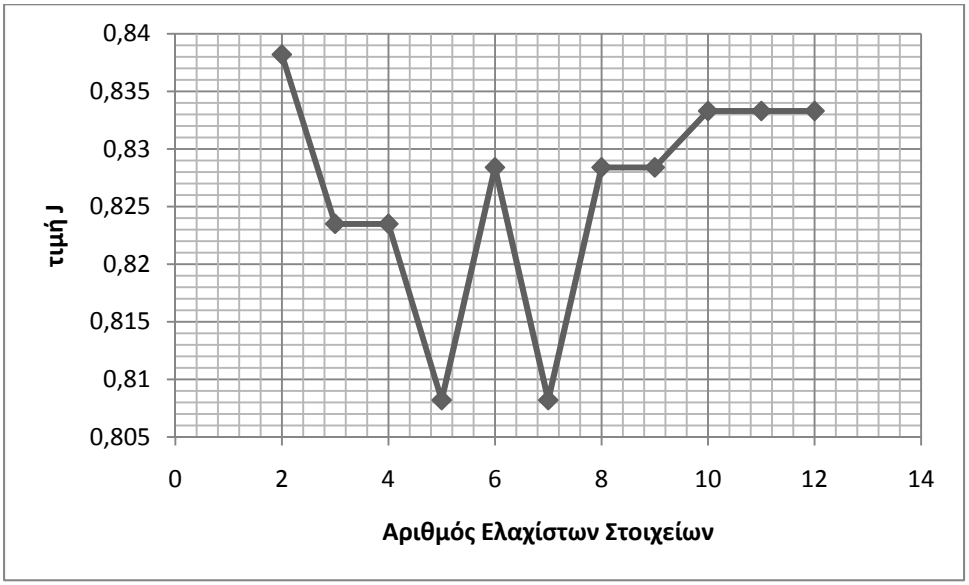
Η προτεινόμενη τιμή του κατασκευαστή είναι το 2. Ξεκινώντας από αυτή αυξάναμε κατά 1 κάθε φορά μέχρι την τιμή 12. Για κάθε τιμή της παραμέτρου εκτελέσαμε 3 επαναλήψεις ώστε να έχουμε ένα μέσο όρο της απόδοσης και να μειώσουμε την επίδραση τυχαίων αποκλίσεων από την τυπική συμπεριφορά σε κάθε σημείο, σε κάποιο λογικά εφικτό βαθμό. Ακολουθεί ο πίνακας με τις μετρήσεις:

<u>Πίνακας 7.1:</u> Μετρήσεις SFFS κανονικοποιημένων δεδομένων

K	Javg	J
2	0,8382	0,8382
		0,8382
		0,8382
3	0,8235	0,8382
		0,8235
		0,8088
4	0,8235	0,8382
		0,8382
		0,7941
5	0,8082	0,7941
		0,7941
		0,8382
6	0,8284	0,8235
		0,8235
		0,8382
7	0,8082	0,8382
		0,7941
		0,7941
8	0,8284	0,8235
		0,8235
		0,8382
9	0,8284	0,8235
		0,8382
		0,8235
10	0,8333	0,8382

		0,8235
		0,8382
11	0,8333	0,8382
		0,8235
		0,8382
12	0,8333	0,8235
		0,8382
		0,8382

και η οπτικοποίηση των μέσων τιμών τους στην εξής γραφική παράσταση:



Σχήμα 7.1: Γραφική Παράσταση Αριθμού Ελαχίστων Στοιχείων-Τιμής συνάρτησης αξιολόγησης J για τον SFFS με κανονικοποιημένα δεδομένα

Παρατηρούμε πως ο αλγόριθμος εμφανίζει την καλύτερη μέση τιμή του για τη μικρότερη (και προτεινόμενη)τιμή 2 με J=0,8382. Αυτή είναι και η ολική μέγιστη τιμή που εμφανίζεται στις μετρήσεις. Στο σημείο αυτό παρουσιάζει λοιπόν και τη μεγαλύτερη σταθερότητα, καθώς επιτυγχάνει το μέγιστο σε όλες τις επαναλήψεις. Γενικά από την τιμή αυτή και μετά εμφανίζει μια πτωτική πορεία μέχρι να αρχίσει να αυξάνεται ξανά και να σταθεροποιείται τελικά από το 10 και μετά στο τοπικό μέγιστο, το 0,8333.

Γενικά, η αύξηση του αριθμού των στοιχείων προκάλεσε αύξηση στο χρόνο ολοκλήρωσης του αλγορίθμου, ιδιαίτερα για 10 στοιχεία και πάνω. Αυτό οφείλεται στο ότι ο αλγόριθμος πρέπει να χειρίζεται περισσότερα στοιχεία, γεγονός που συνεπάγεται περισσότερες επαναλήψεις των βασικών διαδικασιών του. Χοντρικά για λίγα στοιχεία χρειάζονταν 2-3 λεπτά για να ολοκληρωθεί, ενώ για τα 12 στοιχεία περίπου 5 λεπτά. Επίσης να τονίσουμε πως τα

προτεινόμενα υποσύνολα αρκετές φορές περιείχαν περισσότερα των ελαχίστων στοιχεία. Συνεπώς μια μικρότερη τιμή δίνει στην πραγματικότητα τη δυνατότητα για μεγαλύτερη ευελιξία στο σύστημα, ώστε να καταλήξει στο καλύτερο υποσύνολο, παρά καθορίζει την απόδοση αυτή κάθε αυτή.

Στη συνέχεια πραγματοποιήσαμε και μετρήσεις για το σύνολο δεδομένων που προέκυψε από την παραγώγιση και οι οποίες παρουσιάζονται στον παρακάτω πίνακα:

<u>Πίνακας 7.2:</u> Μετρήσεις SFFS παραγωγιζόμενων δεδομένων

Αριθμός Αρχικών Στοιχείων	J
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1

Παρατηρούμε πως το σύνολο που προέρχεται από την παραγώγιση των δεδομένων δίνει την τέλεια απόδοση στο σύστημα για οποιοδήποτε αριθμό αρχικών στοιχείων. Αυτό σημαίνει πως το αυτό το σύνολο δεδομένων είναι πιο φιλικό για το σύστημα κατηγοριοποίησης. Επίσης το βέλτιστο υποσύνολο λύσης που προκύπτει από αυτήν την προεπεξεργασία φαίνεται να έχει μεγαλύτερο μέγεθος (πάνω από 10 στοιχεία) απ’ ότι με την απλή κανονικοποίηση. Στο μέγεθος αυτό φαίνεται ότι πάντοτε καταλήγει ο αλγόριθμος, καθώς ξεκινά με μικρότερο αριθμό στοιχείων.

Η τιμή που επελέχθει λοιπόν για την παράμετρο αυτή είναι το 2, καθώς συνδυάζει το μικρότερο χρόνο τρεξίματος και την καλύτερη δυνατή απόδοση και για τις δύο περιπτώσεις προεπεξεργασίας δεδομένων.

7.4 HGA

Στη συνέχεια υλοποιήθηκε ο Υβριδικός Γενετικός Αλγόριθμος, όπως παρουσιάζεται στο αντίστοιχο κεφάλαιο της θεωρίας. Το είδος του αλγορίθμου που υλοποιήθηκε στην εργασία

είναι σταθερής γενεάς, δηλαδή σε κάθε γενεά ένα μόνο μέρος του πληθυσμού αντικαθίσταται από τους απογόνους. Στη συγκεκριμένη περίπτωση έχουμε τη γέννηση ενός μόνο απογόνου από κάθε ζευγάρι και συνεπώς σε κάθε γενεά αντικαθίσταται ένα μόνο χρωμόσωμα. Αυτό δεν είναι απαραίτητα το λιγότερο κατάλληλο.

Πιο συγκεκριμένα, αν ο απόγονος είναι πιο κατάλληλος και από τους δύο γονείς, αντικαθιστά τον περισσότερο κατάλληλο, αν η καταλληλότητα του είναι ανάμεσα τους αντικαθιστά τον λιγότερο κατάλληλο γονέα και τέλος αν είναι λιγότερο κατάλληλος και από τους δύο γονείς αντικαθιστά το χειρότερο χρωμόσωμα του πληθυσμού. Τα προαναφερθέντα αποτελούν και τη συνθήκη αντικατάστασης που χρησιμοποιήθηκε κατά την υλοποίηση του αλγορίθμου.

Επίσης, ο αλγόριθμος έχει και μια λειτουργία τοπικής βελτιστοποίησης του απογόνου. Η θεωρητική περιγραφή της έλαβε χώρα στο αντίστοιχο προηγούμενο κεφάλαιο. Συνοπτικά, αυτό που κάνει είναι να προσθέτει *r* φορές τα πιο σημαντικά χαρακτηριστικά που δε βρίσκονται στο επιλεγμένο σύνολο και να αφαιρεί τις ίδιες φορές τα λιγότερο σημαντικά χαρακτηριστικά από το επιλεγμένο υποσύνολο. Φαινομενικά λοιπόν μπορεί να μην επηρεάζει τον αλγόριθμο, επί της ουσίας όμως έχει ευεργετικές ιδιότητες. Η χρήση της προφανώς βελτιώνει την ποιότητα του απογόνου τοπικά, δηλαδή προκαλεί μικροαλλαγές στη σύσταση του με σκοπό την αύξηση της καταλληλότητας του. Το βάθος της βελτιστοποίησης *r*, είναι σημαντικός παράγοντας για την αποτελεσματικότητα της, καθώς όσο μεγαλύτερο είναι αυτό, τόσο περισσότερο πιθανό είναι να βελτιωθεί ο απόγονος. Πέρα από αυτό όμως, επιτελεί και μια άλλη σημαντική λειτουργία που είναι η διατήρηση της σταθερότητας του αριθμού *d* των γονιδίων στην προεπιλεγμένη τιμή της, καθώς αν ανιχνευθεί απόκλιση (λόγω διασταύρωσης ή μετάλλαξης) επιδρούν οι ανάλογοι τελεστές μεμονωμένα και επαναφέρουν τη ζητούμενη ισορροπία. Το μεγάλο και μοναδικό μειονέκτημα είναι το υπολογιστικό κόστος, καθώς ο αλγόριθμος ακόμα για την πιο μικρή τιμή του *r*, το *r*=1, χρειάζεται τον τετραπλάσιο χρόνο κατά μέσο όρο για να ολοκληρωθεί.

Ας δούμε όμως και τις υπόλοιπες λεπτομέρειες της υλοποίησης.

Οι παράμετροι του συστήματος είναι:

- P: αριθμός χρωμοσωμάτων πληθυσμού
- d: αριθμός ενεργών γονιδίων (επιλεγμένων χαρακτηριστικών)
- p_m: πιθανότητα μετάλλαξης
- αριθμός συνολικών γενεών
- ο αριθμός των σημείων διασταύρωσης
- r: το βάθος της τοπικής βελτιστοποίησης
- η ποινή w
- q: η πίεση επιλογής

Αρχικά ο αλγόριθμος δημιουργεί έναν αριθμό *P* χρωμοσωμάτων και στη συνέχεια δίνει σε *d* τυχαία γονίδια (χαρακτηριστικά) την τιμή 1 ενώ όλα τα υπόλοιπα γίνονται 0. Ο αριθμός *d* είναι προκαθορισμένος από τον χρήστη και χάρη στον υβριδικό αλγόριθμο που χρησιμοποιούμε υπάρχουν ενσωματωμένοι μηχανισμοί για τη διατήρηση του σε αυτή σταθερή τιμή με πολύ καλή επιτυχία. Η τιμή της πιθανότητας διασταύρωσης είναι 1 (συνεπώς σε κάθε επανάληψη έχουμε διασταύρωση και ποτέ δε συμβαίνει κλωνοποίηση), ενώ τη τιμή της πιθανότητας μετάλλαξης ορίστηκε στο p_m=0.1.

Αφού λοιπόν ο αλγόριθμος δημιουργήσει τον αρχικό τυχαίο γονεϊκό πληθυσμό και αξιολογήσει τα χρωμοσώματα, εκκινά τις γνωστές διαδικασίες επιλογής, διασταύρωσης, μετάλλαξης καθώς και των διαδικασιών τοπικής βελτιστοποίησης. Στη συνέχεια το νέο χρωμόσωμα αντικαθιστά, ανάλογα με την καταλληλότητα του κάποιο από τα γονεϊκά και ο αλγόριθμος επαναλαμβάνεται.

Ο αλγόριθμος ολοκληρώνεται μετά το πέρας συγκεκριμένου αριθμού γενεών, οπότε και από το σύνολο των χρωμοσωμάτων του πληθυσμού, επιλέγεται το καταλληλότερο ως λύση. Τα γονίδια αυτού του χρωμοσώματος που έχουν την τιμή 1, αποτελούν και τα προτεινόμενα χαρακτηριστικά του αλγορίθμου και συνιστούν το υποσύνολο που θα τροφοδοτήσει τον κατηγοριοποιητή του επόμενου σταδίου. Παρουσιάζονται παρακάτω συνοπτικά τα βήματα:

1. Δημιουργία αρχικού πληθυσμού
2. Επιλογή 2 χρωμοσωμάτων από το γονεϊκό πληθυσμό
3. Εφαρμογή διασταύρωσης και μετάλλαξης για τη δημιουργία του απογόνου
4. Εφαρμογή τοπικών βελτιώσεων στον απόγονο
5. Αντικατάσταση κάποιου χρωμοσώματος του γονεϊκού πληθυσμού με βάση τη συνθήκη αντικατάστασης
6. Επανάληψη του αλγορίθμου από το βήμα 2 μέχρι τη συμπλήρωση συγκεκριμένου αριθμού γενεών

Γενικά ο αλγόριθμος είναι πιο αργός από τον SFFS και τα ποσοστά επιτυχίας του υψηλά και σταθερά. Αυτό οφείλεται στη διαδικασία την τοπικής βελτιστοποίησης. Δίχως αυτή ο αλγόριθμος παρουσιάζει πολύ γρήγορο χρόνο ολοκλήρωσης και ικανοποιητική απόδοση, η οποία όμως χαρακτηρίζεται από υψηλές διακυμάνσεις στην τιμή της. Για το λόγο αυτό κρίνεται θετική η προσθήκη αυτής της διαδικασίας για το σύστημα, αν και αντικειμενικά ο επιπλέον χρόνος είναι πολύ μεγάλος σε σχέση με την απλή έκδοση του αλγορίθμου.

7.5 Παράμετροι HGA

Ο αλγόριθμος HGA που χρησιμοποιείται στην εργασία έχει αρκετές παραμέτρους. Αυτό, πέρα από το χρόνο που χρειάζεται για τη διερεύνηση τους, εισαγάγει και ένα νέο θέμα, το γεγονός πως κάποιες παράμετροι είναι πιθανό να συσχετίζονται. Αυτή η συσχέτιση, είναι πολύ δύσκολο να εκτιμηθεί από τη πειραματική διαδικασία, καθώς σε αυτή μεταβάλλουμε μόνο μια παράμετρο και διατηρούμε όλες τις άλλες σταθερές. Ο μόνος τρόπος να αντιληφθούμε κάτι τέτοιο είναι να υπάρχει σαφής υπόνοια από τη θεωρία. Για το λόγο αυτό, ο συνδυασμός των βέλτιστων τιμών των παραμέτρων που θα προτείνονται στην υλοποίηση, είναι εξαιρετικά πιθανό να δίνουν και τα καλύτερα αποτελέσματα, όχι όμως 100% σίγουρο.

Γενικά όπως προαναφέραμε, όλες οι παράμετροι διατηρούνται σταθερές στις προτεινόμενες τιμές του δημιουργού του αλγορίθμου και μεταβάλλεται κάθε φορά μόνο μια με σκοπό την εύρεση της περιοχής τιμών που παρουσιάζει καλύτερη απόδοση, ή τέλος πάντων να παρατηρηθεί ο τρόπος με τον οποίο επηρεάζει την απόδοση γενικότερα. Το κριτήριο που επιλέξαμε είναι η βελτιστοποίηση της καταλληλότητας του καλύτερου και συνεπώς προτεινόμενου χρωμοσώματος. Η καταλληλότητα του χρωμοσώματος δεν είναι τίποτα άλλο από τη συνάρτηση αξιολόγησης J, όπως αυτή εμφανίζεται στα υπόλοιπα FSA και σκιαγραφεί επί της ουσίας την απόδοση του συστήματος επιλογής χαρακτηριστικών. Ο χρόνος ολοκλήρωσης ενός τρεξίματος με τις προτεινόμενες τιμές είναι περίπου 5 λεπτά. Βέβαια, καθώς μεταβάλλονται οι διάφορες παράμετροι είναι πιθανόν να αυξηθεί ο χρόνος τρεξίματος. Σε αυτές τις περιπτώσεις θα γίνεται σχετική αναφορά.

Οι προτεινόμενες τιμές για κάθε παράμετρο είναι:

- P (αριθμός χρωμοσωμάτων)=20
- d (αριθμός γονιδίων)=10
- p_m (παράμετρος μετάλλαξης)=0,1
- T (Αριθμός Γενεών)= καμία (ανάλογα με το σύνολο δεδομένων)
- ο αριθμός των σημείων διασταύρωσης=3
- r (βάθος τοπικής βελτιστοποίησης)=2
- w (η ποινή)=0,25
- q (η πίεση επιλογής)=0,5

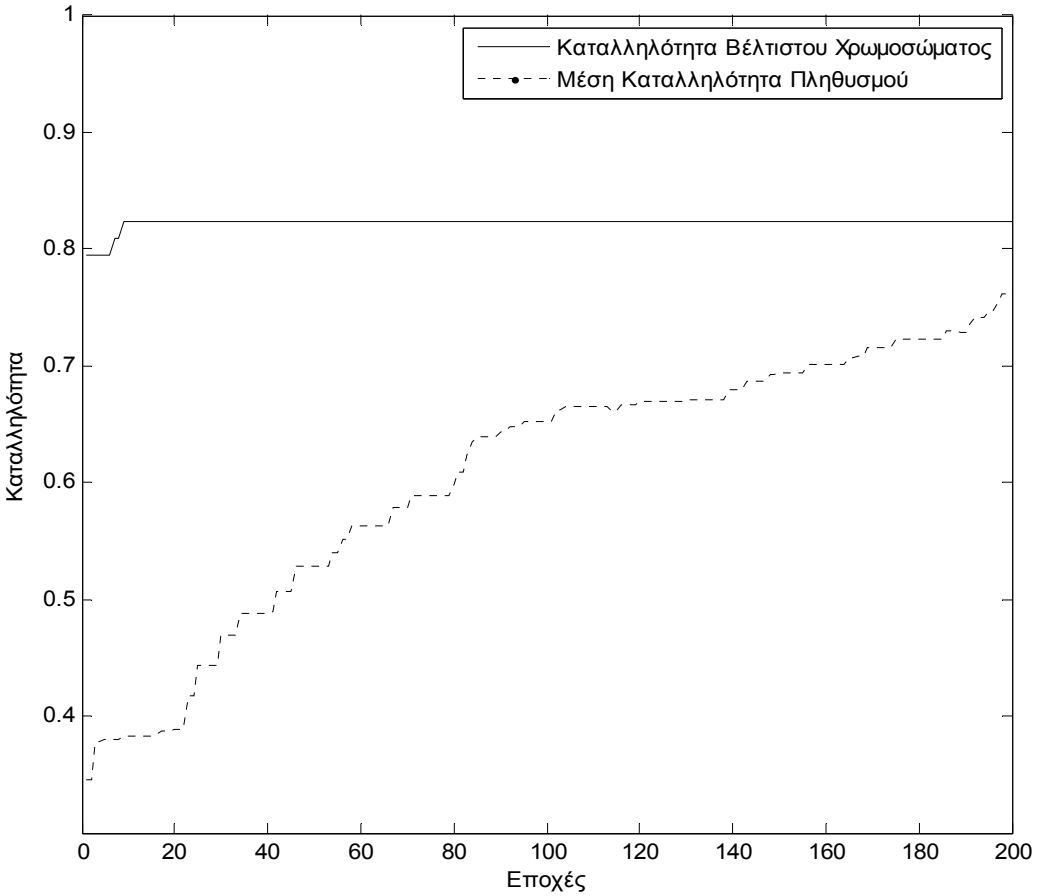
Ας δούμε λοιπόν πως επηρεάζει η μεταβολή κάθε μιας την απόδοση του συστήματος

7.5.1 Αριθμός Γενεών

Ο αριθμός των γενεών είναι με πιο απλά λόγια το χρονικό περιθώριο που έχει ο αλγόριθμος για να κάνει την έρευνά του και η λογική λέει πως όσο μεγαλύτερο είναι αυτό, τόσο καλύτερη απόδοση θα έχει το σύστημα. Στην πράξη, καθώς οι γενετικοί αλγόριθμοι βασίζονται σε ένα βαθμό και στην τύχη είναι πιθανό να δημιουργηθεί ένα καλό χρωμόσωμα πολύ νωρίς ή να

χρειάζεται περισσότερες γενιές από τις προκαθορισμένες μέχρι να εμφανιστεί. Ο στόχος μας είναι να διερευνήσουμε το κατά πόσο αυτό συμβαίνει στο σύνολο δεδομένων της εργασίας.

Φυσιολογικά ένας απλός ΓΑ χρειάζεται μεγάλο αριθμό γενεών για να συγκλίνει σε μια καλή λύση. Στο συγκεκριμένο υβριδικό ΓΑ όμως χρησιμοποιούμε την τοπική βελτιστοποίηση και αυτό μας δίνει το δικαίωμα να μειώσουμε τον αριθμό των γενεών σε μια πολύ μικρότερη της τυπικής τιμή. Λαμβάνοντας αυτά υπ' όψιν, για τη μέτρηση θέσαμε των αριθμό των γενεών στο 200. Τα μεγέθη των οποίων τη μεταβολή των τιμών μετρήσαμε είναι η καταλληλότητα του καλύτερου χρωμοσώματος και η μέση καταλληλότητα του πληθυσμού. Ακολουθεί η αντίστοιχη γραφική παράσταση:



Σχήμα 7.2: Γραφική παράσταση της μέσης καταλληλότητας του πληθυσμού και της καταλληλότητας του βέλτιστου χρωμοσώματος σε σχέση με τον αριθμό εποχών

Παρατηρούμε πως η καταλληλότητα του βέλτιστου χρωμοσώματος αυξάνει πολύ γρήγορα (μέσα σε 2-3 εποχές) και στη συνέχεια σχεδόν σταθεροποιείται στην περιοχή τιμών γύρω από

το 0,8. Από την εποχή 18 και έπειτα λαμβάνει τη μέγιστη τιμή της fitness=0,8283 και τη διατηρεί μέχρι και το τέλος του τρεξίματος. Δηλαδή όσες εποχές και αν περάσουν δεν είναι δυνατόν να επιτευχθεί μια καλύτερη τιμή. Από την άλλη πλευρά, η μέση καταλληλότητα του πληθυσμού αυξάνεται συνεχώς με το πέρας των γενεών, αρχικά πιο γρήγορα και στη συνέχεια με μικρότερο ρυθμό.

Από τα παραπάνω καταλήγουμε στο συμπέρασμα ότι δε χρειαζόμαστε ένα μεγάλο αριθμό γενεών για να πάρουμε τη βέλτιστη λύση που χρειαζόμαστε. Αυτό οφείλεται σε ένα πολύ μεγάλο βαθμό στην ύπαρξη της τοπικής βελτιστοποίησης. Συνεπώς, αν και χρονοβόρα σαν διαδικασία, βοηθά το σύστημα πολύ γρήγορα να βρει τη μέγιστη τιμή του και συνεπώς αντί να χρειαζόμαστε χιλιάδες γενιές, όπως χρειάζεται ένας τυπικός ΓΑ, μπορούμε με μερικές δεκάδες να πετύχουμε την καλύτερη λύση διατηρώντας παράλληλα το μέγεθος του υποσυνόλου λύσης στα προκαθορισμένα επίπεδα, όπως έχει εξηγηθεί και στη θεωρία. Επίσης είναι πιθανό η τιμή να μειωθεί και άλλο σε περίπτωση που χρησιμοποιηθεί ένα καλύτερο σύνολο παραμέτρων, όπως αυτό προκύψει από την περαιτέρω διερεύνησή τους.

Ως τιμή λοιπόν της παραμέτρου θα χρησιμοποιήσουμε το 35. Αυτή η τιμή, η οποία είναι σχεδόν διπλάσια από την τιμή για την οποία επετεύχθη το μέγιστο με τις προτεινόμενες τιμές, μας επιτρέπει να έχουμε ένα καλό περιθώριο έρευνας, σε περίπτωση που αρχικά ο αλγόριθμος δυσκολεύεται να συγκλίνει σε μια καλή λύση, διατηρώντας παράλληλα τον χρόνο τρεξίματος σε λογικά λειτουργικά πλαίσια (περίπου 3 λεπτά).

7.5.2 Αριθμός χρωμοσωμάτων P

Ο αριθμός των χρωμοσωμάτων είναι στην πράξη ο αριθμός των υποσυνόλων τα οποία αποτελούν τις πιθανές λύσεις του προβλήματος επιλογής χαρακτηριστικών. Μεγαλύτερος αριθμός δίνει τη δυνατότητα για περισσότερους συνδυασμούς υποσυνόλων (ζευγαρώματα), άρα βοηθά στην ύπαρξη γενετικής ποικιλομορφίας και συνεπώς είναι πιθανότερο να δημιουργηθούν απόγονοι με μεγάλη καταλληλότητα. Από την άλλη είναι πιθανό τα κατάλληλα χρωμοσώματα να μη ζευγαρώνουν τόσο συχνά και έτσι να καθυστερείται η εξελικτική διαδικασία, αν και αυτό εξαρτάται και από την παράμετρο της πίεσης επιλογής q.

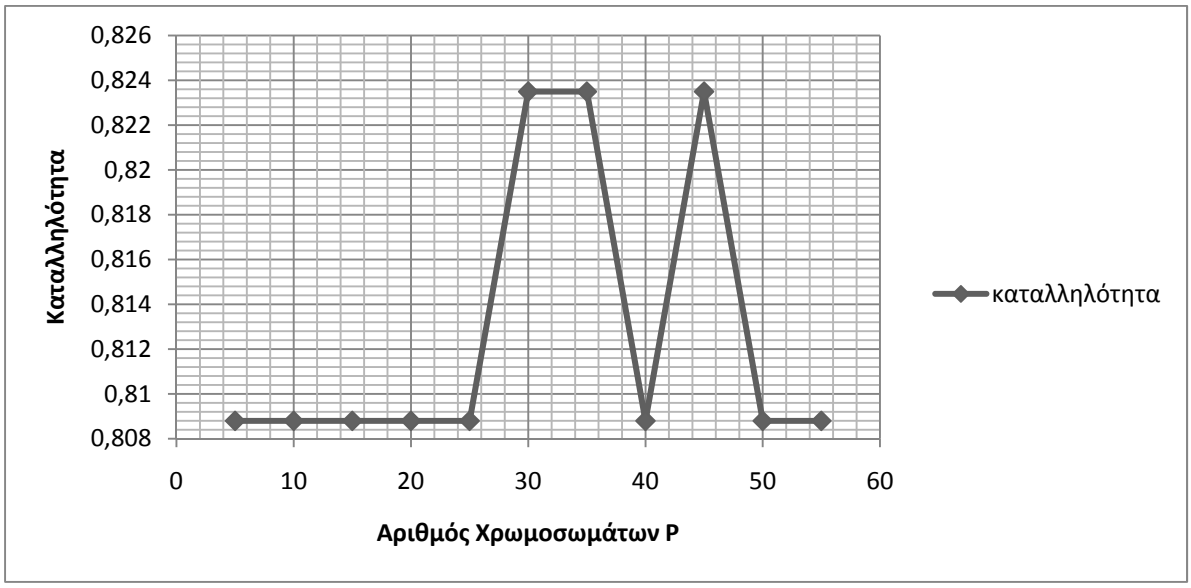
Η προτεινόμενη τιμή της παραμέτρου P είναι το 20. Μεταβάλλαμε την τιμή της από 5 έως 55 με βήμα 5. Στον παρακάτω πίνακα συνοψίζονται οι μετρήσεις:

<u>Πίνακας 7.3</u>: Μετρήσεις αριθμού χρωμοσωμάτων-καταλληλότητας

P	καταλληλότητα
5	0,8088
10	0,8088
15	0,8088
20	0,8088

25	0,8088
30	0,8235
35	0,8235
40	0,8088
45	0,8235
50	0,8088
55	0,8088

και ακολουθεί η οπτικοποίηση τους στην εξής γραφική παράσταση:



Σχήμα 7.3: Γραφική παράσταση της καταλληλότητας του βέλτιστου χρωμοσώματος σε σχέση με τον αριθμό των χρωμοσωμάτων

Παρατηρούμε πως η μεταβολή στην τιμή αυτής της παραμέτρου επηρεάζει την καταλληλότητα. Φαίνεται να υπάρχει μια περιοχή καλών τιμών μεταξύ 30 και 35 καθώς και κοντά στο 45. Αυτό που δε φαίνεται όμως να υπάρχει, είναι μια συγκεκριμένη αναλογία μεταξύ των 2, τουλάχιστον όχι για το συγκεκριμένο σύνολο δεδομένων. Η μέγιστη καταλληλότητα που επετεύχθη είναι 0,8235.

Πέρα από όλα αυτά, η αύξηση του αριθμού των χρωμοσωμάτων του πληθυσμού προκάλεσε ανάλογη μικρή αύξηση του χρόνου τρεξίματος, καθώς ο αλγόριθμος διαχειρίζεται περισσότερα στοιχεία. Λαμβάνοντας αυτό υπ’ όψιν, θεωρούμε πως η ιδανικότερη τιμή για τον πληθυσμό είναι η μικρότερη δυνατή από αυτές που επιτυγχάνεται το μέγιστο, δηλαδή το P=30.

7.5.3 Αριθμός Γονιδίων d

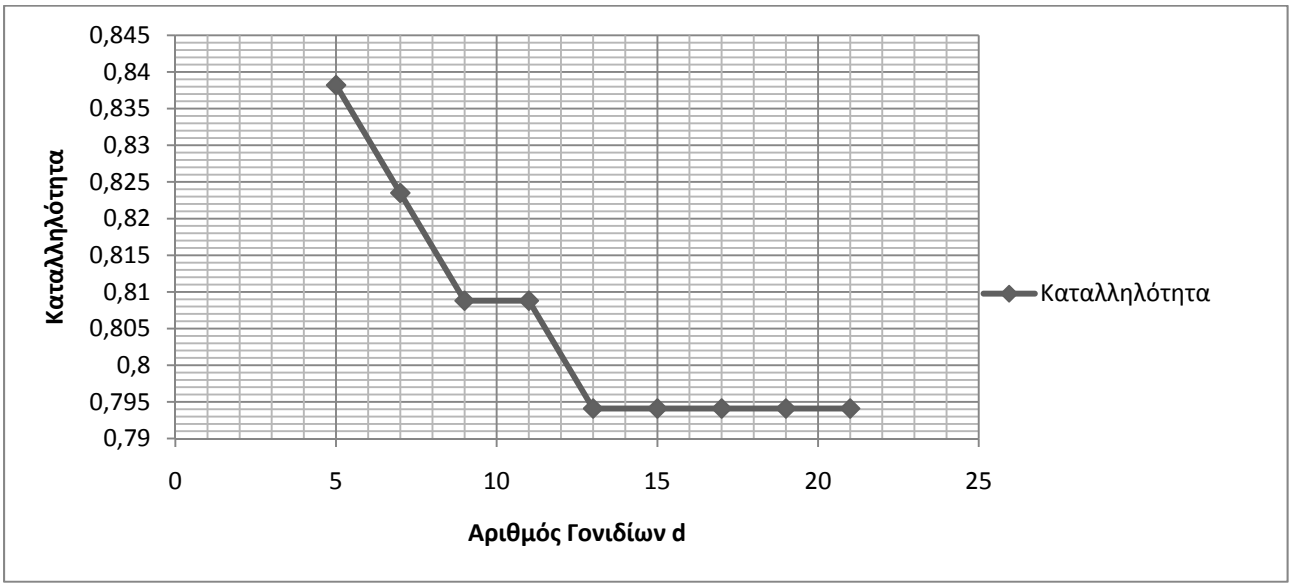
Ο αριθμός των γονιδίων είναι ο αριθμός των επιλεγμένων χαρακτηριστικών σε κάθε υποσύνολο (χρωμόσωμα) και επί της ουσίας αυτή η παράμετρος καθορίζει το μέγεθος του προτεινόμενου υποσυνόλου λύσης.

Η προτεινόμενη τιμή της παραμέτρου d είναι το 10. Μεταβάλλαμε την τιμή της από το 5 έως το 21 με βήμα 2. Να προσθέσουμε εδώ ότι λόγω της υλοποίησης κάποιων έτοιμων συναρτήσεων της Matlab, οι οποίες δε σχετίζονται άμεσα με τη φύση του αλγόριθμου, ο αριθμός των γονιδίων δεν μπορούσε να είναι μικρότερος από 5, καθώς τότε δημιουργούνταν πρόβλημα με το indexing των πινάκων. Στον παρακάτω πίνακα συνοψίζονται οι μετρήσεις:

Πίνακας 7.4: Μετρήσεις αριθμού γονιδίων-καταλληλότητας

d	J
5	0,8382
7	0,8235
9	0,8088
11	0,8088
13	0,7941
15	0,7941
17	0,7941
19	0,7941
21	0,7941

και ακολουθεί η οπτικοποίηση τους στην εξής γραφική παράσταση:



Σχήμα 7.4: Γραφική παράσταση αριθμού γονιδίων-καταλληλότητας βέλτιστου χρωμοσώματος

Για τη συγκεκριμένη παράμετρο παρατηρούμε πως η αύξηση του αριθμού των γονιδίων προκαλεί σταθερή γενικά μείωση της καταλληλόλητας του καλύτερου χρωμοσώματος, μέχρι την σταθεροποίηση της στην τιμή 0,7941. Η μέγιστη τιμή επετεύχθη για τη μικρότερη τιμή των μετρήσεων d=5 και είναι 0,8382. Είναι γνωστό πως ο αριθμός των γονιδίων αντιστοιχεί στα επιλεγμένα χαρακτηριστικά του υποσυνόλου λύσης και συνεπώς υπονοείται από τις μετρήσεις πως το μέγεθος του βέλτιστου υποσυνόλου δεν είναι μπορεί να είναι μεγάλο. Η μεταβολή σε αυτήν την παράμετρο δεν επηρέασε με κάποιον εμφανή τρόπο το χρόνο τρεξίματος.

Με βάση τα παραπάνω επιλέξαμε να δώσουμε στην παράμετρο αυτή την τιμή d=5.

7.5.4 Συντελεστής Μετάλλαξης p_m

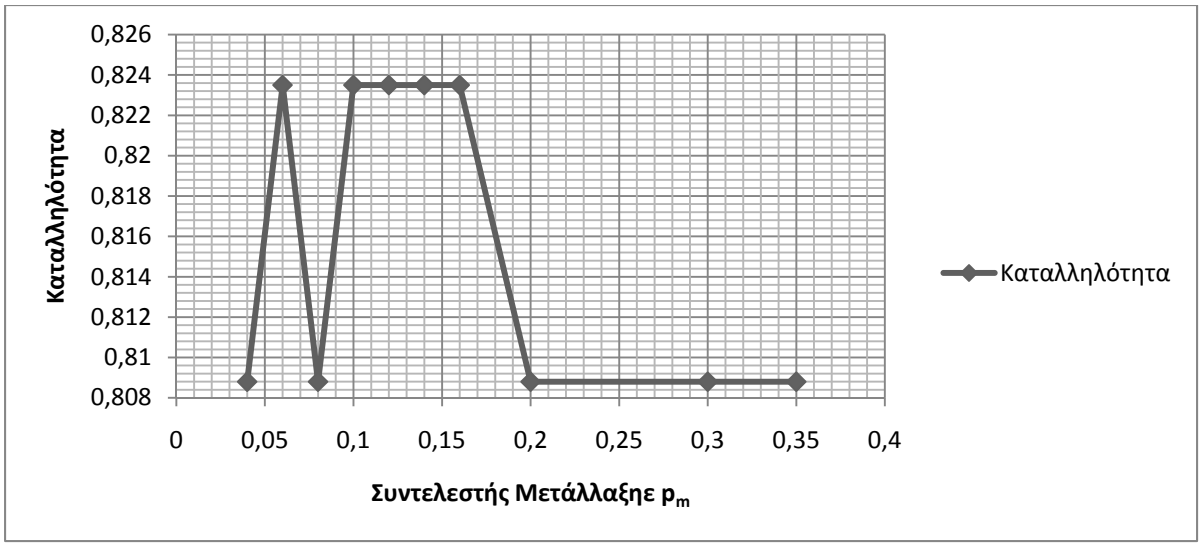
Ο συντελεστής μετάλλαξης είναι στην πραγματικότητα η μοντελοποίηση της τυχαιότητας στο σύστημά μας και σκοπό έχει να βοηθήσει όσο μπορεί τον πληθυσμό να ξεφεύγει από τοπικά μέγιστα. Παρεμβαίνει σε διάφορα σημεία του χρωμοσώματος απογόνου και αντιστρέφει την τιμή του γονίδιου, με μικρή συχνότητα πάντως, καθώς το αντίθετο θα προκαλούσε χάος στη διαδικασία.

Η προτεινόμενη τιμή της παραμέτρου p_m είναι το 0,1. Μεταβάλλαμε την τιμή της από το 0,04 έως το 0,16 με βήμα 0,02 και ύστερα λάβαμε μετρήσεις και σε κάποια πιο απομακρυσμένα σημεία ώστε να . Στον παρακάτω πίνακα συνοψίζονται οι μετρήσεις:

<u>Πίνακας 7.5</u>: Μετρήσεις συντελεστή μετάλλαξης-καταλληλότητας

pm	Καταλληλότητα
0,04	0,8088
0,06	0,8235
0,08	0,8088
0,1	0,8235
0,12	0,8235
0,14	0,8235
0,16	0,8235
0,2	0,8088
0,3	0,8088
0,35	0,8088

και ακολουθεί η οπτικοποίηση τους στην εξής γραφική παράσταση:



Σχήμα 7.5: Γραφική παράσταση συντελεστή μετάλλαξης-καταλληλότητας βέλτιστου χρωμοσώματος

Από τη γραφική παράσταση φαίνεται πως παρουσιάζεται ένα σταθερό μέγιστο στην περιοχή από 0,1 έως 0,16 και ένα μέγιστο στο 0,06. Η τιμή του είναι 0,8235. Για τις τιμές έξω από αυτά τα διαστήματα η τιμή είναι μικρότερη και σταθερή στο 0,8088. Μάλιστα όσο μεγαλώνει η τιμή της παραμέτρου δε φαίνεται να μεταβάλλεται κάτι, γεγονός που επιβεβαιώνει την ανάγκη ισορροπημένης εφαρμογής της μετάλλαξης, εάν θέλουμε να καρπωθούμε τα πλεονεκτήματά της. Επίσης να τονίσουμε ότι λόγω της ύπαρξης της τοπικής βελτιστοποίησης, η μετάλλαξη χάνει σε ένα βαθμό τη σημαντικότητά της, καθώς προηγείται αυτής και συνεπώς αν κάποια αλλαγή δεν είναι ωφέλιμη είναι πιθανό πως θα διορθωθεί στη συνέχεια.

Με βάση τα παραπάνω, η παράμετρος επελέχθη να λάβει την τιμή 0,12 η οποία βρίσκεται περίπου στη μέση του διαστήματος όπου παρατηρήθηκε σταθερά το μέγιστο.

7.5.5 Πίεση Επιλογής

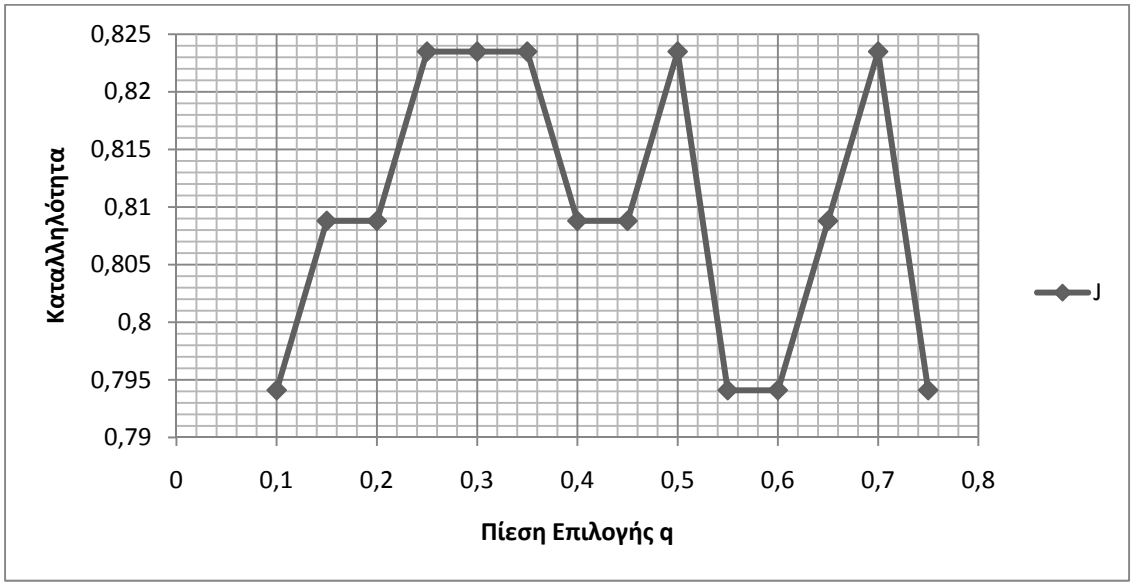
Η παράμετρος της πίεσης επιλογής καθορίζει το βαθμό κατά τον οποίο τα καλύτερα χρωμοσώματα υπερισχύουν στο κυνήγι του ζευγαρώματος των λιγότερο κατάλληλων χρωμοσωμάτων. Κυμαίνεται στο διάστημα [0,1], με τις μεγαλύτερες τιμές να ασκούν μεγαλύτερη πίεση και να ενισχύουν την κυριαρχία στον πληθυσμό των καταλληλότερων χρωμοσωμάτων. Η παράμετρος αυτή έχει και μια θεωρητική σχέση με τον αριθμό των χρωμοσωμάτων, καθώς για μεγάλες τιμές μόνο τα καταλληλότερα χρωμοσώματα πρωταγωνιστούν σε κάθε γενεά και συνεπώς ένας μεγάλος αριθμός πληθυσμού δε θα έχει νόημα.

Η προτεινόμενη τιμή της παραμέτρου q είναι το 0,5. Μεταβάλλαμε την τιμή της από το 0,1 έως το 0,75 με βήμα 0,05. Ακολουθεί ο πίνακας :που συνοψίζει τις μετρήσεις:

<u>Πίνακας 7.6</u>: Μετρήσεις συντελεστή πίεσης επιλογής-καταλληλότητας

q	Καταλληλότητα
0,1	0,7941
0,15	0,8088
0,2	0,8088
0,25	0,8235
0,3	0,8235
0,35	0,8235
0,4	0,8088
0,45	0,8088
0,5	0,8235
0,55	0,7941
0,6	0,7941
0,65	0,8088
0,7	0,8235
0,75	0,7941

και ακολουθεί η οπτικοποίηση τους στην εξής γραφική παράσταση:



Σχήμα 7.6: Γραφική παράσταση πίεσης επιλογής-καταλληλότητας βέλτιστου χρωμοσώματος

Παρατηρούμε πως εμφανίζεται ένα ολικό μέγιστο, το 0,8235, στην περιοχή από 0,25 έως 0,35 και στα σημεία 0,5 και 0,7. Το ολικό ελάχιστο ήταν το 0,7941. Η μεταβολή αυτής της παραμέτρου δεν επηρέασε τον χρόνο τρεξίματος.

Επιλέξαμε να δώσουμε σε αυτήν την παράμετρο την τιμή 0,3, η οποία βρίσκεται στη μέση του διαστήματος όπου παρατηρήθηκε σταθερά η μέγιστη τιμή.

7.5.6 Ποινή w

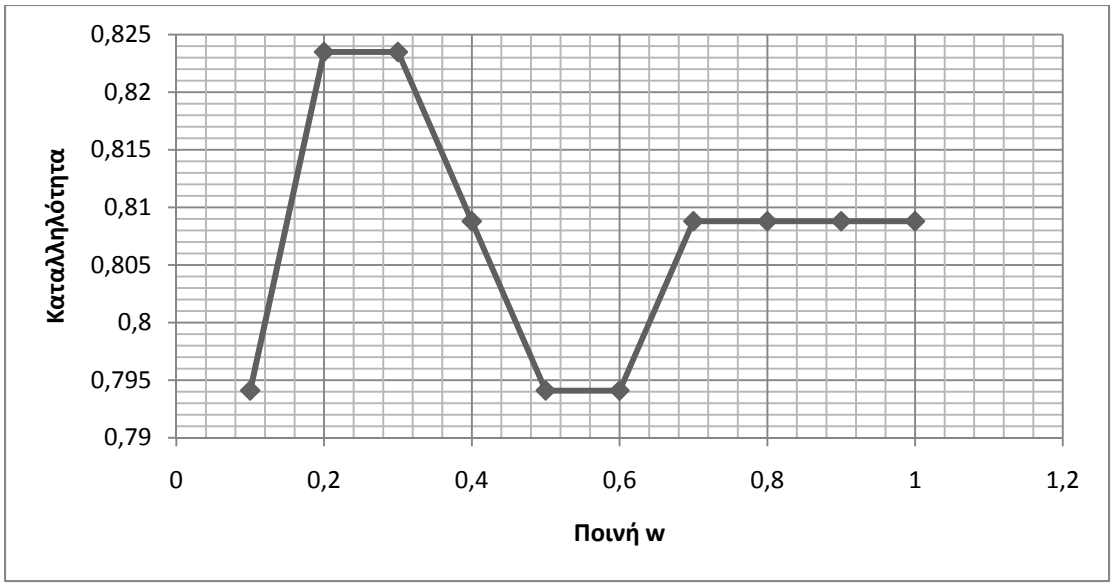
Η ποινή w είναι ένας ποσό που αφαιρείται από την καταλληλότητα του χρωμοσώματος, ανάλογα με το πόσο αποκλίνει το μέγεθος του από το προκαθορισμένο d. Αναλυτική περιγραφή έχει γίνει στο αντίστοιχο κεφάλαιο της θεωρίας. Οι τιμές της κυμαίνονται στο διάστημα [0,1], με υψηλότερες τιμές να επιφέρουν μεγαλύτερη ποινή για την ίδια παράβαση.

Η προτεινόμενη τιμή της παραμέτρου w είναι το 0,25. Μεταβάλλαμε την τιμή της από το 0,1 έως το 1 με βήμα 0,1. Ακολουθεί ο πίνακας :που συνοψίζει τις μετρήσεις:

<u>Πίνακας 7.7</u>: Μετρήσεις ποινής-καταλληλότητας

w	Καταλληλότητα
0,1	0,7941
0,2	0,8235
0,3	0,8235
0,4	0,8088
0,5	0,7941
0,6	0,7941
0,7	0,8088
0,8	0,8088
0,9	0,8088
1	0,8088

και ακολουθεί η οπτικοποίηση τους στην εξής γραφική παράσταση:



Σχήμα 7.7: Γραφική παράσταση ποινής-καταλληλότητας βέλτιστου χρωμοσώματος

Παρατηρούμε πως το μέγιστο εμφανίζεται στην περιοχή από 0,2 έως 0,3 και έχει την τιμή 0,8235. Στη συνέχεια μειώνεται η τιμή σταθερά, μέχρι να αυξηθεί εκ νέου και να μονιμοποιηθεί σε ένα τοπικό μέγιστο στο 0,7 με τιμή 0,8088.

Η τιμή που επελέχθει για την παράμετρο είναι η 0,25 η οποία βρίσκεται στη μέση του διαστήματος όπου παρατηρήθηκε σταθερά η βέλτιστη τιμή.

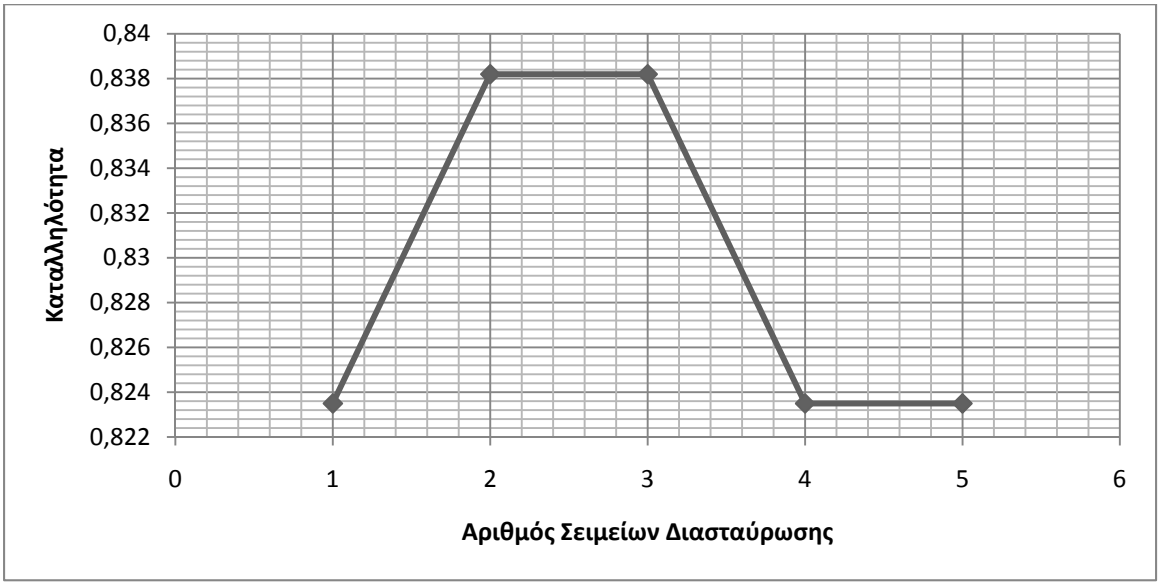
7.5.7 Αριθμός Σημείων Διασταύρωσης

Ο αριθμός των σημείων διασταύρωσης καθορίζει τον αριθμό των σημείων στα οποία θα κοπούν τα χρωμοσώματα γονείς και θα ανταλλάξουν τα κομμάτια τους ώστε να δημιουργήσουν τους απογόνους. Δεν υπήρχε κάποια προτεινόμενη τιμή για αυτή την παράμετρο αλλά γενικά λαμβάνει μικρές τιμές. Λάβαμε μετρήσεις από το 1 έως το 5 με βήμα 1. Ακολουθεί ο συγκεντρωτικός πίνακας με τις μετρήσεις:

<u>Πίνακας 7.8</u>: Μετρήσεις σημείων διασταύρωσης-καταλληλότητας

σημεία διασταύρωσης	καταλληλότητα
1	0,8235
2	0,8382
3	0,8382
4	0,8235
5	0,8235

και η οπτικοποίηση του στην παρακάτω γραφική παράσταση:



Σχήμα 7.8: Γραφική παράσταση αριθμού σημείων διασταύρωσης-καταλληλότητας βέλτιστου χρωμοσώματος

Παρατηρούμε πως παίρνει τη μέγιστη τιμή της για 2 και 3 σημεία διασταύρωσης. Επιλέξαμε την τιμή 3 για την παράμετρο.

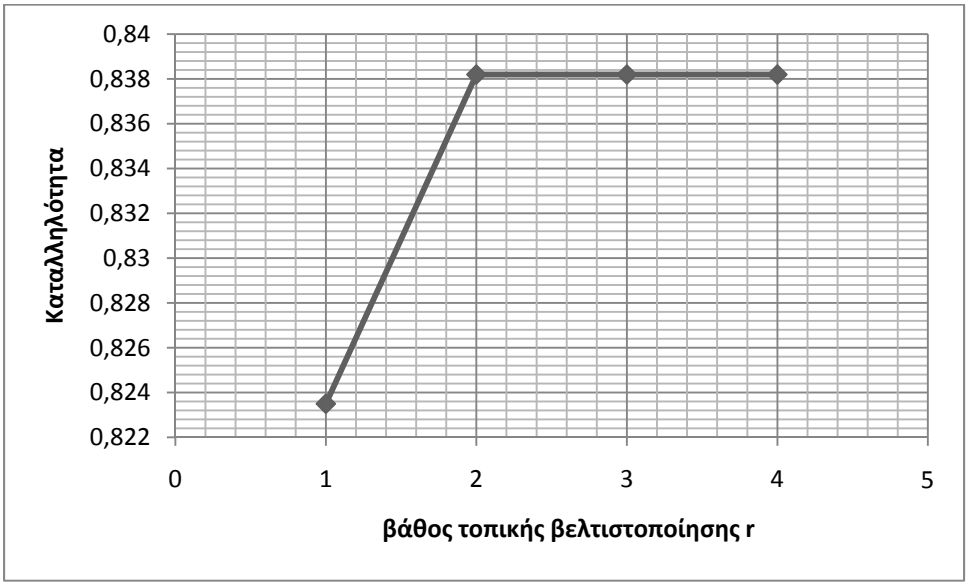
7.5.8 Βάθος Τοπικής Βελτιστοποίησης

Το βάθος της τοπικής βελτιστοποίησης r μπορεί να λάβει ακέραιες τιμές μεγαλύτερες του 1. Η προτεινόμενη τιμή είναι το 2. Κάναμε μετρήσεις από το 1 έως και το 4, χρησιμοποιώντας αυτήν τη φορά τις βέλτιστες τιμές όπως προέκυψαν από τα παραπάνω. Ακολουθεί ο πίνακας που συγκεντρώνει τις μετρήσεις:

<u>Πίνακας 7.9</u>: Μετρήσεις βάθους τοπικής βελτιστοποίησης-καταλληλότητας

r	καταλληλότητα
1	0,8235
2	0,8382
3	0,8382
4	0,8382

και η οπτικοποίησή τους στην παρακάτω γραφική παράσταση:



Σχήμα 7.9: Γραφική παράσταση βάθους τοπικής βελτιστοποίησης-καταλληλότητας βέλτιστου χρωμοσώματος

Αύξηση της τιμής του r αυξάνει κατά πολύ το χρόνο τρεξίματος. Παρατηρούμε όμως πως αυτό δε συνάγει με βελτίωση της μέγιστης καταλληλότητας, η οποία παραμένει σταθερή στην τιμή J=0,8382. Οπότε επιλέγουμε να δώσουμε στην παράμετρο την τιμή r=2, καθώς με αυτή έχουμε το μικρότερο χρόνο τρεξίματος.

Συνοψίζοντας τα παραπάνω συμπεράσματα για τις βέλτιστες τιμές των παραμέτρων του υλοποιημένου αλγορίθμου HGA έχουμε:

- P (αριθμός χρωμοσωμάτων)=30
- d (αριθμός γονιδίων)=5
- p_m (παράμετρος μετάλλαξης)=0,12
- T (Αριθμός Γενεών)= 30
- o αριθμός των σημείων διασταύρωσης=3
- r (βάθος τοπικής βελτιστοποίησης)=2
- w (η ποινή)=0,25
- q (η πίεση επιλογής)=0,3

Επίσης φάνηκε πως η πιο σημαντική παράμετρος για την απόδοση του συστήματος είναι ο αριθμός των γονιδίων και κατ’ επέκταση ο αριθμός των γενεών. Επίσης ευαισθησία

εμφανίστηκε και στις αλλαγές στην πίεση επιλογής αλλά και στο βάθος της τοπικής βελτιστοποίησης.

Όμως σε αυτό το σημείο πρέπει να θυμηθούμε πως στην προεπεξεργασία είχαμε χρησιμοποιήσει απλή κανονικοποίηση των δεδομένων. Συνεπώς όλα τα παραπάνω μεγέθη είναι βασισμένα σε αυτό. Διατηρώντας τις βέλτιστες τιμές για τις παραμέτρους αλλάξαμε το κομμάτι της προεπεξεργασίας, ώστε να κανονικοποιεί αλλά και να παραγωγίζει το δεδομένα με Γκαουσιανό πυρήνα, όπως εξηγήθηκε στη θεωρία. Πραγματοποιήσαμε 5 συνεχόμενες μετρήσεις οι οποίες συνοψίζονται στον παρακάτω πίνακα:

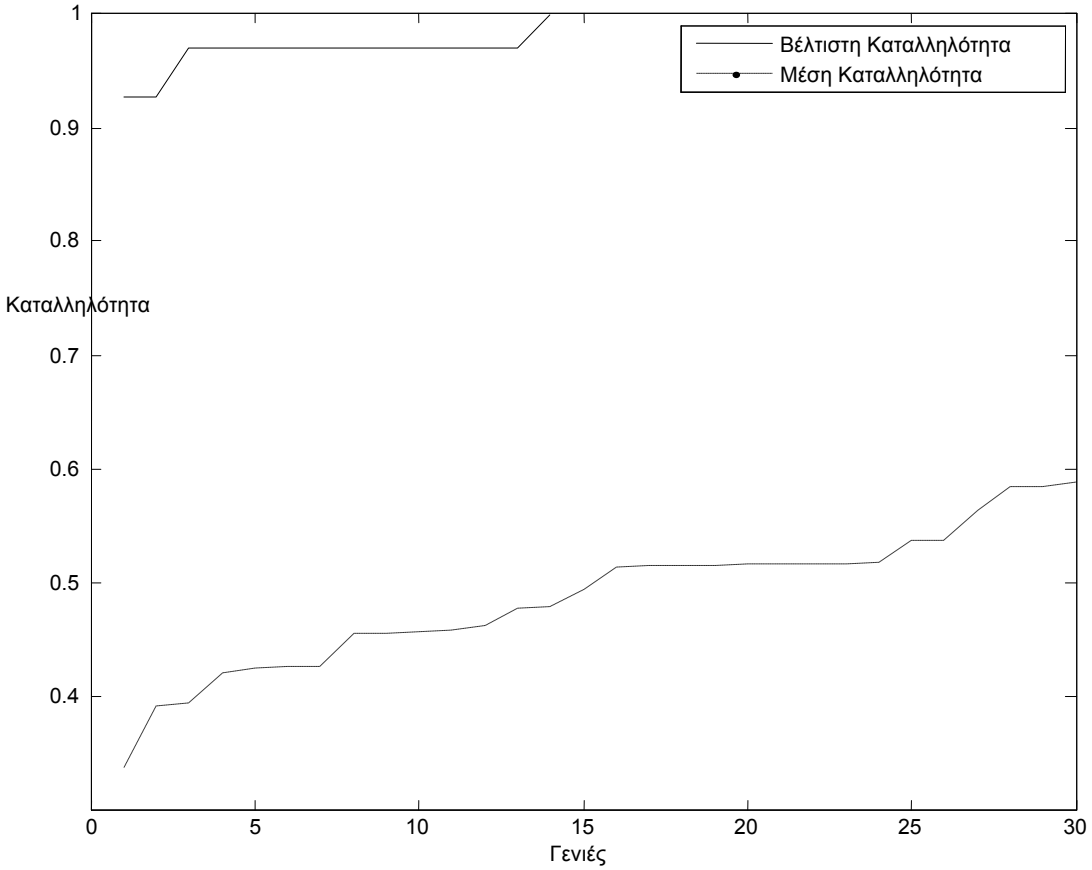
<u>Πίνακας 7.10</u>: Μέτρηση καταλληλότητας παραγωγιζόμενων στοιχείων με τις βέλτιστες παραμέτρους για 5 επαναλήψεις

Αριθμός Επανάληψης	Τιμή Βέλτιστης Καταλληλότητας
1	1
2	0,9706
3	0,9853
4	1
5	1
Μέση Τιμή	0,99118

Στο σχήμα 7.10 παρουσιάζεται μια ενδεικτική γραφική παράσταση της μέσης και βέλτιστης καταλληλότητας για τις βελτιστοποιημένες παραμέτρους.

Παρατηρούμε λοιπόν ότι η αλλαγή αυτή επέφερε δραματική βελτίωση στην απόδοση του συστήματος επιλογής χαρακτηριστικών, η οποία πλέον αγγίζει το τέλειο, με αξιοσημείωτη σταθερότητα. Με απλή κανονικοποίηση παρατηρήσαμε πως η καταλληλότητα αδυνατούσε να ξεπεράσει την τιμή 0,8382. Επίσης, αποδεικνύει πως η επιλογή των παραμέτρων με βάση τα προαναφερθέντα είναι εύστοχη και επιτυχημένη. Μάλιστα από τις γραφικές παραστάσεις των υπολοίπων μετρήσεων φάνηκε πως ο αλγόριθμος κάθε φορά συγκλίνει πριν την 20^{η} γενεά(και μερικές φορές και πριν τη 10^{η}). Για το λόγο αυτό αποφασίστηκε να μειωθεί ο αριθμός των γενεών από 30 σε 25. Με αυτόν τον τρόπο γλυτώνουμε χρόνο χωρίς να θυσιάζουμε αποτελεσματικότητα, απλά εισαγάγουμε το ρίσκο αν ο αλγόριθμος δε συγκλίνει μέσα σε αυτό το διάστημα να μην έχουμε τέλεια καταλληλότητα αλλά κάτι πολύ κοντά σε αυτή. Η πιθανότητα αυτή πάντως χαρακτηρίζεται μικρή. Συνεπώς για χρήση του HGA προτείνεται ανεπιφύλακτα η χρήση παραγώγου με Γκαουσιανό πυρήνα. Τέλος καταδεικνύεται η σημασία του υποσυστήματος προεπεξεργασίας για την απόδοση του συνολικού συστήματος. Άρα τελικά έχουμε την εξής αλλαγή:

- Τ (Αριθμός Γενεών)= 25



Σχήμα 7.10: Γραφική παράσταση καταλληλότητας βέλτιστου χρωμοσώματος και μέσης καταλληλότητας του πληθυσμού σε σχέση με τον αριθμό των γενεών για τα παραγωγιζόμενα δεδομένα. Παρατηρούμε πως η καταλληλότητα του βέλτιστου χρωμοσώματος γίνεται τέλεια μετά τη 14^{η} γενιά.

7.6 Υποσύστημα Κατηγοριοποίησης

Οι μετρήσεις πραγματοποιήθηκαν με χρήση ενός συνόλου που είχε προέλθει από το υποσύστημα επιλογής χαρακτηριστικών και είχε πετύχει καταλληλότητα 0,8382 για τα κανονικοποιημένα δεδομένα και 1 για τα παραγωγιζόμενα δεδομένα. Η μέθοδος που ακολουθήθηκε για την επιλογή χαρακτηριστικών ήταν η SFFS, και χρησιμοποιήθηκε το πρώτο βέλτιστο υποσύνολο λύσης. Αυτό διατηρήθηκε κοινό σε όλες τις μετρήσεις με σκοπό να έχουμε μια σταθερά με βάση την οποία να κάνουμε τις συγκρίσεις μας. Αυτό που θέλουμε

είναι περισσότερο να δούμε τη συμπεριφορά των συστημάτων και να την αξιολογήσουμε, παρά να πετύχουμε αντικειμενικά μεγάλες τιμές (χωρίς αυτό να σημαίνει πως αυτό δε είναι επιθυμητό!).

Οι κατηγοριοποιητές SVM και τα Νευρωνικά χρειάζονται κάποια πρότυπα εκπαίδευσης και κάποια πρότυπα ελέγχου ώστε να λειτουργήσουν. Αν και υπάρχουν διάφορες μέθοδοι για να γίνει αυτός ο διαχωρισμός των προτύπων, στην εργασία χρησιμοποιήσαμε την απλούστερη· τα 3/4 των προτύπων από κάθε κατηγορία απετέλεσαν τα σύνολο εκπαίδευσης και το υπόλοιπο 1/4 το υποσύνολο ελέγχου, με βάση τη σειρά εμφάνισής τους.

7.7 Παράμετροι SVM

Τα SVM περιγράφησαν αναλυτικά στο αντίστοιχο θεωρητικό κεφάλαιο. Από εκεί προκύπτει ότι δε χαρακτηρίζονται από πολλές παραμέτρους. Το μόνο που μπορούμε να μεταβάλλουμε είναι το είδος της συνάρτησης πυρήνα που χρησιμοποιείται καθώς και τη μέθοδο σύγκλισης. Οι επιλογές είναι οι εξής:

Συνάρτηση Πυρήνα:

- Linear (Γραμμική)
- Polynomial (Πολυωνυμική)
- Quadratic (Τετραγωνική)
- RBF

Μέθοδος Σύγκλισης:

- Σειριακή Ελαχιστοποίηση (SMO)
- Ελάχιστα Τετράγωνα (LS)

Ακολουθεί ο παρακάτω πίνακας που περιέχει τις διάφορες μετρήσεις ανάλογα με τις παραμέτρους που χρησιμοποιούνται κάθε φορά:

<u>Πίνακας 7.11</u>: Μετρήσεις απόδοσης SVM για μεταβολή των διαφόρων παραμέτρων του τόσο για κανονικοποιημένα όσο και για παραγωγιζόμενα δεδομένα

SVM				
Τύπος Δεδομένων	Παράγωγος	Κανονικοποίηση	Παράγωγος	Κανονικοποίηση
Μέθοδος Σύγκλισης	SMO	SMO	LS	LS
Συνάρτηση Πυρήνα				
Quadratic	0,7778	0,9444	0,7222	0,9444
Linear	0,8333	0,8889	0,8333	0,8333
Polynomial Order				
1	0,8333	0,8889	0,8333	0,8333
2	0,7778	0,9444	0,7222	0,9444
3	0,8333	0,8889	0,8333	0,8889
4	0,7778	0,8889	0,7888	0,7778
5	0,8333	0,8333	0,8889	0,444
6	0,7222	0,8889	0,7222	0,7778
7	0,7778	0,7778	0,6111	0,7778
8	0,6667	0,8333	0,6111	0,5
RBF	0,8333	0,7778	0,8333	0,7778

Παρατηρούμε πως η μέγιστη τιμή για τα κανονικοποιημένα δεδομένα προκύπτει με χρήση Quadratic πυρήνα και είναι ανεξάρτητη της μεθόδου σύγκλισης. Η τιμή που δίνει είναι J=0,94 και αποτελεί την ολικά μέγιστη τιμή όλων των μετρήσεων. Συνεπώς επιλέγουμε αυτή τη συνάρτηση σε περίπτωση κανονικοποιημένων δεδομένων.

Για τα παραγωγιζόμενα δεδομένα το αντίστοιχο μέγιστο προκύπτει για τον Polynomial πυρήνα τάξης 5. Εδώ, η μέθοδος σύγκλισης παίζει ρόλο, καθώς για τα ελάχιστα τετράγωνα επιτυγχάνεται καλύτερη τιμή, J=0,89, ενώ για τη σειριακή μέθοδο έχουμε, J=0,83. Συνεπώς η Polynomial συνάρτηση τάξης 5 επιλέγεται σε αυτήν την περίπτωση.

7.8 Νευρωνικά Δίκτυα

Οι παράμετροι του νευρωνικού είναι:

- Συνάρτηση Μεταφοράς
    - tansig (φαινομενική)
    - logsig (λογαριθμική)
    - purelin(γραμμική)

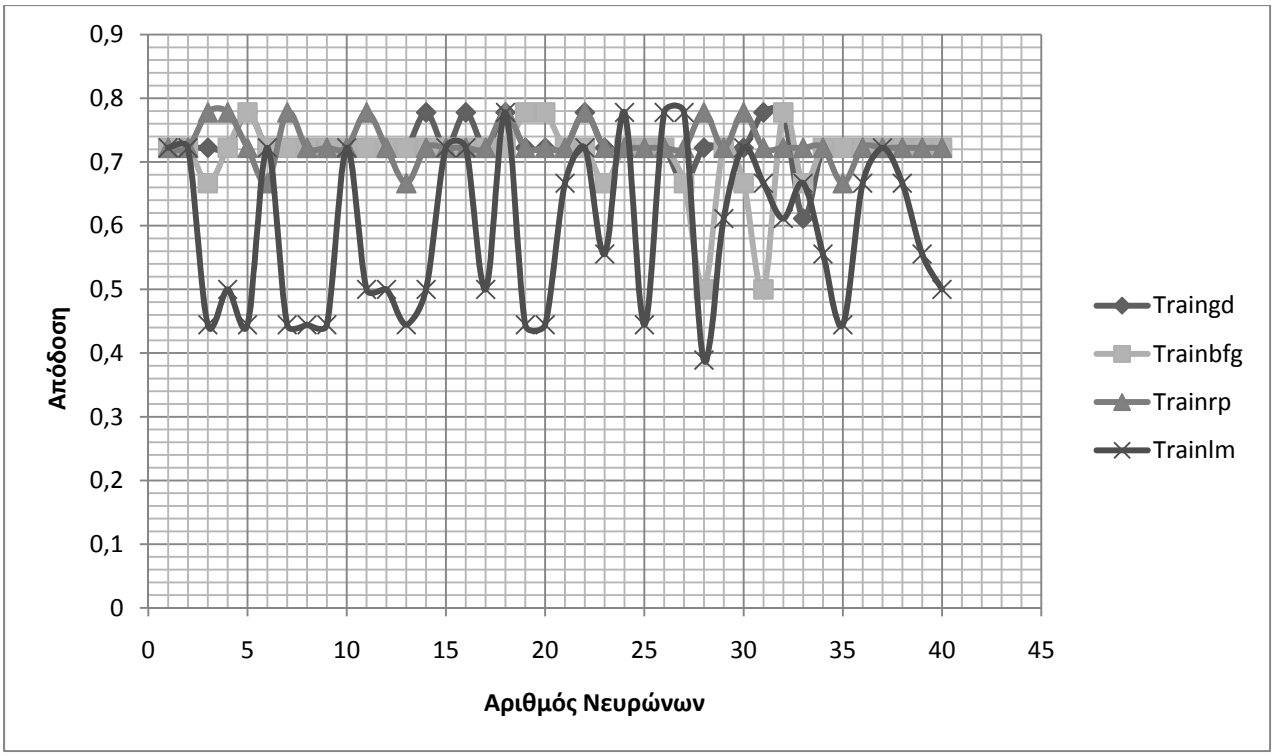
- Συνάρτηση Μάθησης (αλγόριθμος εκπαίδευσης)
    - trainlm (Levenberg-Marquardt backpropagation)
    - trainbfg (BFGS quasi-Newton backpropagation)
    - trainrp (Αλγόριθμος ελαστικού backpropagation)
    - traingd (Κατάβαση δυναμικού)

- Αριθμός Κρυφών Σταδίων
- Αριθμός Νευρώνων/Στάδιο

Ο αριθμός των κρυφών σταδίων κρατήθηκε μικρός και συγκεκριμένα στο 1. Δοκιμάστηκαν και αρχιτεκτονικές 2 σταδίων αλλά δεν επελέχθησαν καθότι αύξαναν την πολυπλοκότητα χωρίς να βελτιώνουν την απόδοση. Συνεπώς σκοπός μας είναι να βρούμε κάποιο καλό συνδυασμό για τις υπόλοιπες παραμέτρους, ώστε να έχουμε ένα σύστημα κατηγοριοποίησης με υψηλές αποδόσεις και ως ένα λογικό βαθμό ανεξάρτητο από το τι συμβαίνει στα προηγούμενα(σταθερότητα).

Κανονικοποιημένα

Σχετικά με τις συναρτήσεις μεταφοράς μπορούμε να πούμε πως η logsig είχε τη χειρότερη απόδοση, ενώ η tansig και purelin κυμαίνονταν στα ίδια επίπεδα, με την purelin να εμφανίζει μεγαλύτερη σταθερότητα. Για το λόγο αυτό επελέχθει η δεύτερη για το σύστημά μας. Στη συνέχεια ακολουθεί η γραφική παράσταση του αριθμού των νευρώνων του κρυφού σταδίου σε σχέση με την τιμή της απόδοσης για τις διάφορες συναρτήσεις μάθησης:



Σχήμα 7.11: Γραφική παράσταση απόδοσης-αριθμού νευρώνων για τα διάφορα είδη συναρτήσεων μάθησης για κανονικοποιημένα δεδομένα

Παρατηρούμε ότι από τις 4 συναρτήσεις, η traingd εμφανίζει τη μεγαλύτερη σταθερότητα, ενώ η trainlm τη μικρότερη. Αυτό σημαίνει πως η traingd εμφανίζει τη μικρότερη ευαισθησία στον αριθμό των νευρώνων, γεγονός επιθυμητό. Παρόλα αυτά, το μέγιστο είναι για όλες το ίδιο στο 0,78. Για το λόγο αυτό επιλέξαμε να χρησιμοποιήσουμε την traingd με 20 νευρώνες, καθώς εκτός του ότι είναι η σταθερότερη σα συνάρτηση, σε εκείνη την περιοχή αριθμού νευρώνων εμφανίζεται το μέγιστο για 2 συνεχόμενα σημεία, οπότε είναι πιθανότερο να έχει υψηλή απόδοση και για άλλα προτεινόμενα υποσύνολα λύσης, πέραν του υποσυνόλου δοκιμής.

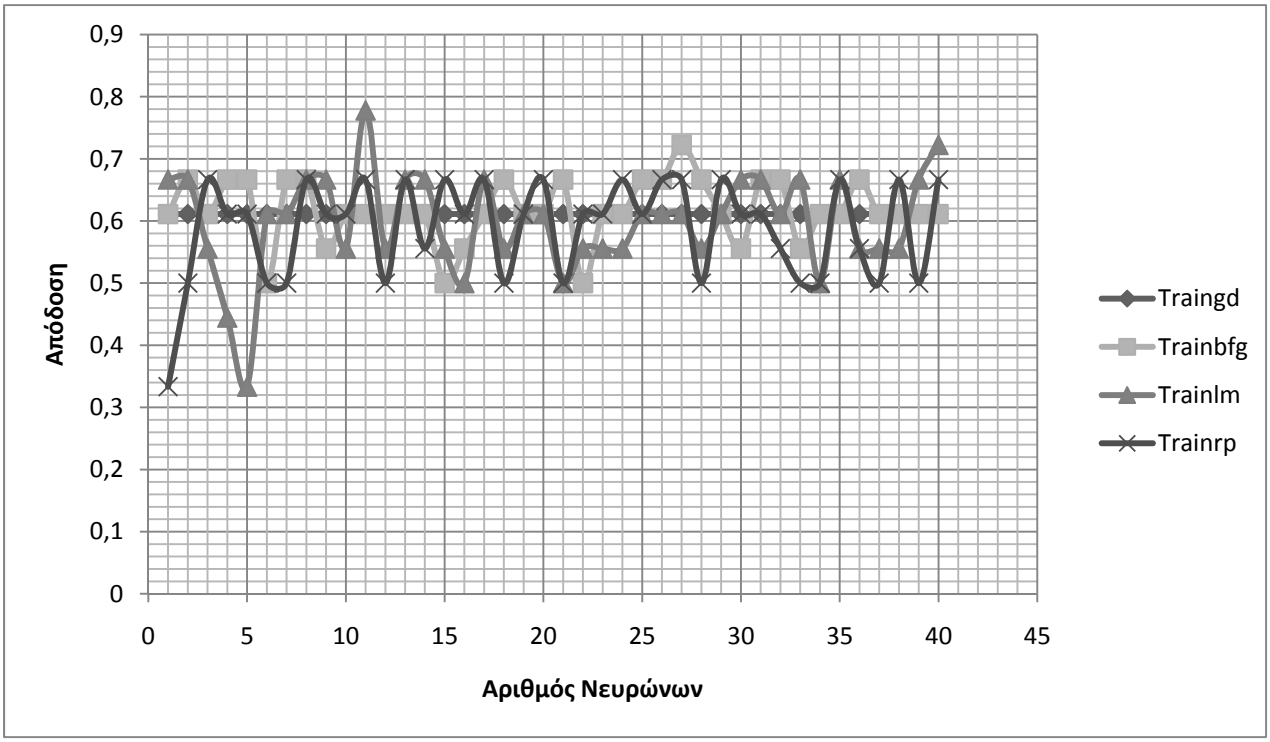
Παραγωγιζόμενα

Μεταβολή της συνάρτησης μεταφοράς από την tansig στη logsig έδωσε τη ζητούμενη σταθερότητα στην απόδοση, αλλά τη μείωσε αρκετά, οπότε απορρίπτεται. Από την άλλη πλευρά η μετάβαση στην purelin έδωσε λίγο καλύτερη σταθερότητα αλλά και καλύτερες τιμές από την tansig. Για το λόγο αυτό επιλέξαμε αυτή.

Σχετικά με τη συνάρτηση μάθησης μπορούμε να πούμε ότι τόσο η trainlm όσο και η trainbfg έδωσαν σταθερότητα στο σύστημα. Η traingd επίσης συνδύασε σταθερότητα με γενικά

καλύτερα αποτελέσματα, ενώ από την άλλη πλευρά η trainrp είχε τη χειρότερη απόδοση. Για το λόγο αυτό επιλέξαμε την traingd.

Ακολουθεί η γραφική παράσταση της μέσης απόδοσης του νευρωνικού σε σχέση με τον αριθμό των νευρώνων, έχοντας τις επιλεγμένες από τα προηγούμενα παραμέτρους :



Σχήμα 7.12: Γραφική παράσταση απόδοσης-αριθμού νευρώνων για τα διάφορα είδη συναρτήσεων μάθησης για παραγωγιζόμενα δεδομένα

Παρατηρούμε πως η πιο σταθερή συνάρτηση είναι η traingd, η οποία πραγματικά δίνει αποτελέσματα ανεξάρτητα εντελώς από αριθμό των νευρώνων. Παρόλα αυτά η τιμή που επιτυγχάνει δεν είναι η μεγαλύτερη δυνατή. Για τις υπόλοιπες συναρτήσεις βλέπουμε ότι η πιο ασταθής αλλά και αυτή που δίνει τη μέγιστη τιμή στο σύστημά μας είναι η trainlm για 11 νευρώνες. Μια επίσης καλή επιλογή είναι και η trainbfg για 27 νευρώνες. Θα επιλέξουμε την trainlm καθώς έχει καλύτερη απόδοση και λιγότερους νευρώνες, δηλαδή μικρότερη πολυπλοκότητα, λαμβάνοντας πάντα υπ' όψιν το ελαφρώς μεγαλύτερη πιθανότητα της ύπαρξης αστάθειας στα αποτελέσματα.

Από τα παραπάνω βλέπουμε επίσης πως η απόδοση των νευρωνικών είναι κατά μέσο όρο καλύτερη όταν χρησιμοποιούνται τα απλά κανονικοποιημένα δεδομένα, σε αντίθεση με τα παραγωγιζόμενα. Κάτι τέτοιο συνέβαινε και στα SVM. Αυτό έρχεται σε αντίθεση με την κατάσταση που επικρατούσε στο υποσύστημα επιλογής χαρακτηριστικών, όπου τα δεύτερα έδιναν πολύ υψηλότερες αποδόσεις, για την ακρίβεια τις τέλειες αποδόσεις.

7.9 K-means

Η κατηγοριοποίηση με τον αλγόριθμο K-means χαρακτηρίζεται από σταθερότητα, δηλαδή για συγκεκριμένο υποσύνολο εισόδου και συγκεκριμένες παραμέτρους θα δίνει πάντα την ίδια τιμή στην απόδοση. Στον αλγόριθμο αυτό υπάρχουν 2 παράμετροι της οποίες μπορούμε να μεταβάλλουμε στην προσπάθειά μας να βελτιστοποιήσουμε την απόδοσή του:

- Ο τρόπος υπολογισμού των αποστάσεων:
    - Ευκλείδεια Απόσταση
    - Άθροισμα απολύτων διαφορών (cityblock)
    - Συνημίτονο (cosine)
    - Συσχέτιση (correlation)
    - Κώδικα Hamming
- Η μέθοδος επιλογής των αρχικών θέσεων:
    - Τυχαίος από τα σημεία του πίνακα εισόδου
    - Τυχαίος από τα σημεία του πίνακα εισόδου με ομοιόμορφη κατανομή
    - Τυχαίος από ένα 10% των σημείων του πίνακα εισόδου

Οι προεπιλεγμένες παράμετροι ήταν η Ευκλείδεια απόσταση και ο τυχαίος από τα σημεία του πίνακα εισόδου. Συνεπώς αυτές ήταν οι επιλογές όταν μεταβάλαμε την έτερη παράμετρο.

Σχετικά με τον τρόπο υπολογισμού των αποστάσεων, μόνο η Ευκλείδεια απόσταση και η cityblock μπορούσαν να εφαρμοστούν, ενώ οι υπόλοιπες αδυνατούσαν λόγω της φύσης των δεδομένων, είτε αυτά είχαν προέλθει από απλή κανονικοποίηση είτε από παραγώγιση.

Κανονικοποιημένα

Στον παρακάτω πίνακα ακολουθούν συγκεντρωμένες οι μετρήσεις:

<u>Πίνακας 7.12</u>: Μετρήσεις απόδοσης Kmeans για τις διάφορες παραμέτρους του για κανονικοποιημένα δεδομένα

Απόσταση	Απόδοση	Αρχικοποίηση	Απόδοση
Ευκλείδεια	0,8235	sample	0,8235
Cityblock	0,8235	cluster	0,8235
		uniform	0,8235

Παρατηρούμε ότι η απόδοση παραμένει σταθερή και ανεξάρτητη όλων των παραμέτρων για τα κανονικοποιημένα δεδομένα.

Παραγωγιζόμενα

Στον παρακάτω πίνακα ακολουθούν συγκεντρωμένες οι μετρήσεις:

<u>Πίνακας 7.13</u>: Μετρήσεις απόδοσης Kmeans για τις διάφορες παραμέτρους του για παραγωγιζόμενα δεδομένα

Απόσταση	Απόδοση	Αρχικοποίηση	Απόδοση
Ευκλείδεια	0,5882	sample	0,5882
Cityblock	0,5882	cluster	0,5882
		uniform	0,5882

Παρατηρούμε πως και σε αυτήν την περίπτωση, η απόδοση παραμένει σταθερή και ανεξάρτητη όλων των παραμέτρων. Επίσης παρατηρούμε πως λαμβάνει πολύ χαμηλότερες τιμές σε σχέση με το άλλος είδος δεδομένων.

7.10 FCM

Η κατηγοριοποίηση με τη βοήθεια του FCM δεν έχει κάποιες παραμέτρους προς βελτιστοποίηση. Να υπενθυμίσουμε ότι αυτός ο αλγόριθμος κατηγοριοποίησης απετέλεσε τον πυρήνα του συστήματος επιλογής χαρακτηριστικών και τα αποτελέσματα του ήταν υπεύθυνα για την τιμή της συνάρτησης αξιολόγησης J. Άρα, αφού απετέλεσε τη βάση επιλογής των υποσυνόλων λύσης, η απόδοση του θα ταυτίζεται με την τιμή της συνάρτησης αξιολόγησης. Αυτό επιβεβαιώθηκε και πειραματικά.

7.11 K-NN

Ο αλγόριθμος KNN δεν έχει άλλες παραμέτρους εκτός από τον αριθμό των γειτόνων για τους οποίους θα υπολογίσει την απόσταση. Στην υλοποίηση της εργασίας τρέχει ο αλγόριθμος 15 φορές και σε κάθε μία αυξάνει τον αριθμό των γειτόνων που υπολογίζει, ξεκινώντας από το 1. Ύστερα βρίσκει για ποιόν αριθμό προκύπτει η καλύτερη τιμή και δίνει αυτόν ως απάντηση. Η κίνηση αυτή έγινε διότι παρατηρήθηκε κατά το τρέξιμο του αλγορίθμου πως ήταν ευαίσθητος στο υποσύνολο εισόδου, δηλαδή για διαφορετικά υποσύνολα έδινε διαφορετικές τιμές, ακόμα και αν αυτά ήταν ισοδύναμης αξίας για το σύστημα επιλογής χαρακτηριστικών. Με αυτό τον τρόπο περιορίζεται η ευαισθησία του συστήματος και αυξάνεται κατά πολύ η σταθερότητα του σε μεγάλες τιμές. Για το λόγο αυτό δεν έχει νόημα να πραγματοποιηθούν μετρήσεις με μεταβολή της παραμέτρου.

7.12 Επιτροπή

Στην εργασία λοιπόν χρησιμοποιήθηκαν αρκετοί διαφορετικοί αλγόριθμοι κατηγοριοποίησης με σκοπό την ύπαρξη ποικιλομορφίας και τη δυνατότητα εξαγωγής πληρέστερων συμπερασμάτων για τα αποτελέσματα. Τα αποτελέσματά τους όμως δεν είναι αναγκαστικό να χρησιμοποιούνται ανεξάρτητα μεταξύ τους. Μπορούν να συνδυαστούν ώστε να μας δώσουν έναν σύνθετο κατηγοριοποιητή ο οποίος συνδυάζοντας τις γνώμες όλων των κατηγοριοποιητών θα μας δίνει τη δικιά του άποψη για το που ανήκει το κάθε στοιχείο.

Αυτό που κάνει λοιπόν είναι να παίρνει την άποψη κάθε επιμέρους κατηγοριοποιητή για το κάθε στοιχείο και να την πολλαπλασιάζει με ένα συντελεστή βάρους $w_i \epsilon [0,1]$ $i = 1, \dots, N$ όπου N ο αριθμός των επιμέρους κατηγοριοποιητών. Σε κάθε περίπτωση θα πρέπει να ισχύει:

$$\sum_{i=1}^{N} w_i = 1$$

Η βαρύτητα που αποδίδεται στη γνώμη κάθε κατηγοριοποιητή είναι άμεση συνάρτηση της απόδοσής του αλλά και της σταθερότητάς του. Ακολουθούν συνοπτικά οι κατηγοριοποιητές που χρησιμοποιήθηκαν καθώς και το βάρος του καθενός:

- Νευρωνικά, w=0,2
- SVM, w=0,35
- K-means, w=0,1
- K-NN, w=0.35

Τα SVM και K-NN έχουν τη μεγαλύτερη τιμή, καθώς και τα δύο χαρακτηρίζονται από μεγάλη σταθερότητα σε υψηλές τιμές. Τα νευρωνικά επιτυγχάνουν καλές τιμές αλλά έχουν μέτρια σταθερότητα γι’ αυτό και έχουν μια ενδιάμεση τιμή και τέλος ο K-means έχει χαμηλότερη απόδοση από όλους γι’ αυτό και η γνώμη του είναι συμπληρωματική και δεν έχει μεγάλη βαρύτητα. Να πούμε επίσης ότι δε χρησιμοποιήθηκε ο FCM καθώς αυτός απλά αποτελεί τη μαλακή εκδοχή του K-means, οπότε μπορούμε να πούμε πως αυτή η οικογένεια αλγορίθμων εκπροσωπείται. Εξάλλου επειδή αυτός ο αλγόριθμος ήδη χρησιμοποιείται στο σύστημα επιλογής χαρακτηριστικών θα είχε πιο ενδιαφέρον να μη συμπεριλάβουμε τη γνώμη του. Πάντως γενικά η ανάθεση τιμών έχει μια ασάφεια και βασίζεται στην υποκειμενικότητα του δημιουργού.

Για το λόγο αυτό δοκιμάστηκε και μια άλλη μέθοδος ανάθεσης βαρών η οποία βασίζεται στην απόδοση του κάθε κατηγοριοποιητή για κάθε περίπτωση (τρέξιμο). Αρχικά αθροίζονταν όλες οι αποδόσεις και στη συνέχεια κάθε επιμέρους απόδοση διαιρούνταν με αυτόν τον αριθμό,

ώστε να μας δώσει το βάρος της συμμετοχής του κάθε κατηγοριοποιητή στην επιτροπή. Συνεπώς τα βάρη αυτή τη φορά ήταν μεταβλητά. Η μέθοδος αυτή πάντως, αν και πιο δίκαια, δεν είχε καλύτερη απόδοση από την προηγούμενη καθώς κατά κύριο λόγο έδινε ένα μέσο όρο τον κατηγοριοποιητών. Αυτό οφείλεται στο ότι οι τιμές της απόδοσης κινούνται σε ένα περιορισμένο εύρος και η διαφοράς της καλής από την κακή είναι της τάξης των δεκαδικών. Μια τέτοια ευαισθησία δεν είναι δυνατόν να ληφθεί υπ' όψιν με αυτόν τον τρόπο, καθώς τα βάρη απείχαν μεταξύ τους κατά συνέπεια λίγο. Για το λόγο αυτό δεν διερευνήθηκε περαιτέρω σα μέθοδος.

Επίσης να προσθέσουμε ότι επειδή τα νευρωνικά και τα SVM όπως έχει εξηγηθεί και προηγουμένως χωρίζουν τα δεδομένα εισόδου σε ένα υποσύνολο εκπαίδευσης και ένα υποσύνολο ελέγχου, και με βάση την απόδοσή τους στο δεύτερο αξιολογούνται, κρίθηκε σωστό να λαμβάνεται υπ' όψιν η απόδοση και των υπολοίπων αλγορίθμων μόνο στο υποσύνολο ελέγχου της εργασίας, κατά την χρήση τους από την επιτροπή.

8. ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

Αφού ολοκληρώθηκε και η πειραματική ανάλυση των μεθόδων, ήρθε η στιγμή να δούμε για τους βέλτιστους συνδυασμούς παραμέτρων των επιμέρους υποσυστημάτων, πως αποδίδει το συνολικό σύστημα κατηγοριοποίησης. Πραγματοποιήθηκαν όλοι οι πιθανοί συνδυασμοί και έγιναν 3 μετρήσεις για κάθε έναν από αυτούς, με σκοπό να έχουμε μια ακριβέστερη εικόνα της απόδοσης του συστήματος.

Ακολουθεί ο πίνακας με τις μέσες τιμές των μετρήσεων:

<u>Πίνακας 8.1</u>: Μετρήσεις απόδοσης των διάφορων υλοποιήσεων του συνολικού συστήματος

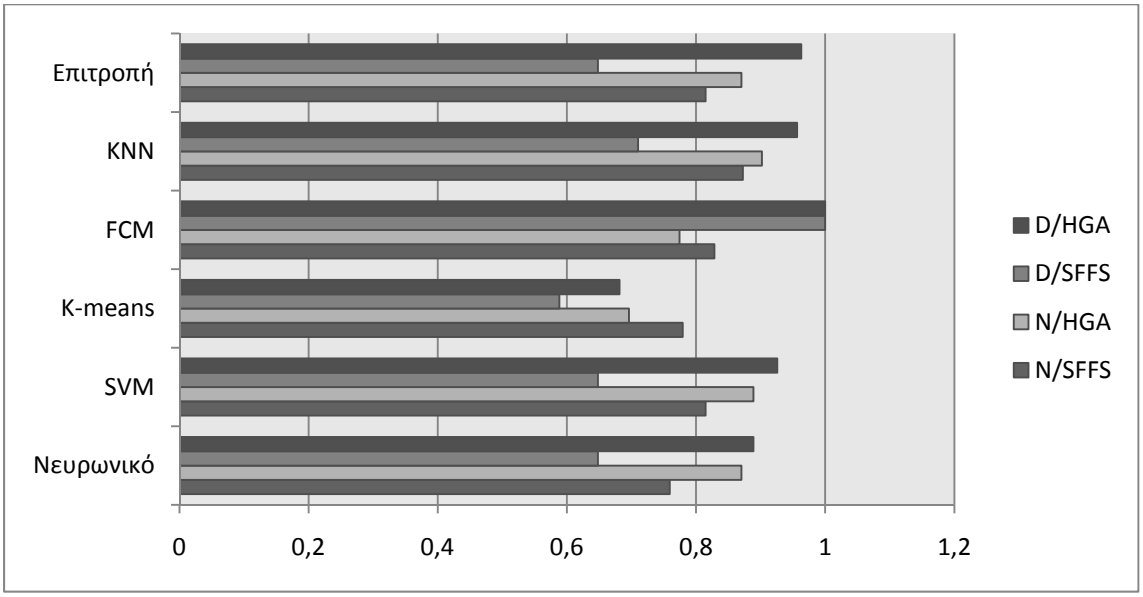
Κατηγοριοποιητής	N/SFFS	N/HGA	D/SFFS	D/HGA
Νευρωνικό	0,759259	0,87037	0,648148	0,888889
SVM	0,814815	0,888889	0,648148	0,925926
K-means	0,779412	0,696078	0,588235	0,681373
FCM	0,828431	0,77451	1	1
KNN	0,872549	0,901961	0,710145	0,956522
Επιτροπή	0,814815	0,87037	0,648148	0,962963
Μέσος Όρος	0,811547	0,833696	0,707137	0,902612

όπου N αντιστοιχεί στα κανονικοποιημένα και D στα παραγωγιζόμενα δεδομένα. Στο σχήμα 8.1 φαίνονται και οπτικοποιημένες οι παραπάνω μετρήσεις.

Παρατηρούμε ότι ο καλύτερος συνδυασμός είναι ο D/HGA οποίος δίνει μέσα ποσοστά επιτυχίας 90%. Οι N/SFFS και N/HGA δίνουν 81% και 83% αντίστοιχα, που αποτελεί μια καλή απόδοση και τέλος ο D/SFFS δίνει 70%, το χειρότερο από όλους, αλλά γενικά αποδεκτό ποσοστό επιτυχίας. Γενικά βλέπουμε πως με τα κανονικοποιημένα δεδομένα έχουμε σταθερές αποδόσεις, και όχι τόσο εξαρτημένες από το σύστημα, ενώ αντίθετα με τα δεδομένα που προέρχονται από παράγωγο έχουμε μεγάλη απόκλιση στην απόδοση που εξαρτάται αρκετά από τις επιμέρους επιλογές υλοποίησης.

Σε ό,τι αφορά τα συστήματα κατηγοριοποίησης, παρατηρούμε ότι ο K-NN είναι ο καλύτερος κατηγοριοποιητής με σταθερά υψηλότερες αποδόσεις από τους υπόλοιπους. Επίσης, παρόμοια συμπεριφορά παρουσιάζουν ο SVM και ο FCM, αν και για το δεύτερο αυτό δεν εντυπωσιάζει τόσο καθώς, όπως προαναφέραμε αποτελεί και τον πυρήνα του συστήματος επιλογής χαρακτηριστικών, οπότε επί της ουσίας αυτός επιλέγει τα καλύτερα χαρακτηριστικά και συνεπώς είναι ευνόητο να πετυχαίνει υψηλές τιμές.

Από την άλλη πλευρά, ο Kmeans φάνηκε να είναι ο χειρότερος από όλους καθώς πάντα έδινε την χαμηλότερη τιμή στην απόδοση η οποία μάλιστα απέκλινε και αρκετά από τις υπόλοιπες. Τα νευρωνικά λόγω ίσως και των πολλών παραμέτρων τους είχαν ενδιάμεσες αποδόσεις που χαρακτηρίζονταν από αστάθεια (δηλαδή για διαφορετικές επαναλήψεις των ίδιων εισόδων έδιναν άλλες αποδόσεις).



Σχήμα 8.1: Γραφική παράσταση της απόδοσης των κατηγοριοποιητών για τις διάφορες υλοποιήσεις του συνολικού συστήματος

Σχετικά με την επιτροπή παρατηρούμε ότι καταφέρνει γενικά να είναι κοντά στις μέγιστες τιμές των επιμέρους κατηγοριοποιητών του κάθε συνδυασμού και μάλιστα για τον καλύτερο συνδυασμό (D/SFFS) την ξεπερνά και θέτει νέο μέγιστο, με ποσοστό επιτυχίας 96% (επαναλαμβάνουμε πως ο FCM δε λαμβάνεται υπ' όψιν για τις συγκρίσεις και απλά έχει προστεθεί τυπικά για λόγους πληρότητας). Αυτό σημαίνει ότι η χρήση της είναι ωφέλιμη για το σύστημα.

Για το υποσύστημα επιλογής χαρακτηριστικών να πούμε πως για τα κανονικοποιημένα δεδομένα η τιμή της συνάρτησης αξιολόγησης J δεν κατάφερε ποτέ να ξεπεράσει την τιμή 0,8382 η οποία απετέλεσε και το ολικό μέγιστο για αυτό το είδος δεδομένων. Από την άλλη πλευρά τα δεδομένα που είχαν προέλθει από παραγώγιση διαμόρφωσαν ένα σύνολο δεδομένων το οποίο μπορούσε να δώσει υποσύνολα χαρακτηριστικών που να είναι τέλεια (J=1) ή πολύ κοντά σε αυτά, με αρκετή συνέπεια. Οπότε από εκεί και έπειτα εξαρτώνταν μόνο από τον κατηγοριοποιητή η απόδοση του συστήματος.

Επίσης φάνηκε πως ο HGA είχε πολύ καλύτερη συμπεριφορά από τον SFFS, καθώς συνεργάζονταν καλύτερα με το υπόλοιπο σύστημα και έδινε σταθερά υψηλές τιμές.

Συνοψίζοντας λοιπόν προτείνεται η χρήση των παραγωγιζόμενων δεδομένων με τον HGA και τον κατηγοριοποιητή K-NN σε περίπτωση που θέλουμε να χρησιμοποιήσουμε απλό κατηγοριοποιητή (με το SVM να αποτελεί μια καλή εναλλακτική) ή της επιτροπής αν θέλουμε σύνθετο κατηγοριοποιητή. Με αυτόν τον τρόπο φαίνεται να επιτυγχάνουμε σταθερά υψηλότατα ποσοστά επιτυχίας. Χρησιμοποιούμε το ρήμα φαίνεται για να υπενθυμισθεί επ' ευκαιρία πως το αρχικό σύνολο δεδομένων ήταν σχετικά φτωχό ώστε με σιγουριά να δημιουργηθεί ένα ευνοϊκό και υγιές περιβάλλον εκπαίδευσης για τους αλγορίθμους. Έτσι αν και η πιθανότητα είναι πολύ μικρή, πάντα μπορεί να υπάρχει ο κίνδυνος του overfitting σε άγνωστα δεδομένα, παρόλο που έγιναν όλες οι απαραίτητες ενέργειες να περιοριστεί (μείωση χαρακτηριστικών, πολλαπλοί κατηγοριοποιητές κτλ.).

Για περαιτέρω έρευνα πάνω στο θέμα προτείνεται η δοκιμή μεθόδων επιλογής χαρακτηριστικών βασισμένων στην αμοιβαία πληροφορία (mutual Πληροφορία), καθώς φαίνεται να υπήρχαν ομάδες χαρακτηριστικών με παρόμοια συνεισφορά στη συνάρτηση αξιολόγησης του αντίστοιχου υποσυστήματος. Αυτό σίγουρα προκαλούσε πρόβλημα στον αλγόριθμο SFFS, το οποίο αρκετές φορές αδυνατούσε να το λύσει (δηλαδή προσέθετε μόνο τα παρόμοια χαρακτηριστικά δημιουργώντας υποσύνολα στα οποία αυτά κυριαρχούσαν, με αποτέλεσμα να κολλά σε τοπικά μέγιστα). Επίσης στο κομμάτι των κατηγοριοποιητών, προτείνεται η περαιτέρω εμβάθυνση στο θέμα του συνδυασμού των αποτελεσμάτων των επιμέρους κατηγοριοποιητών για τη σύσταση υπερκείμενων (γενικότερων/σύνθετων) κατηγοριοποιητών με τελικό στόχο την τέλεια απόδοση, ανεξάρτητα με το τι προηγείται. Αυτό ενισχύεται και από τα αποτελέσματα της επιτροπής που χρησιμοποιήθηκε στην εργασία, τα οποία φανερώνουν τις ευεργετικές επιπτώσεις της στο σύστημα. Καλό θα ήταν να δοθεί έμφαση στη χρήση κατηγοριοποιητών από διαφορετικές οικογένειες, ώστε ακόμα και αν τα προηγούμενα υποσυστήματα δεν αποδίδουν τόσο καλά, να μπορούμε να εκμεταλλευτούμε το ότι η ποικιλία των ειδών συνεπάγεται και διαφορετική επιρροή του οποιουδήποτε σφάλματος σε αυτούς και έτσι να αυξήσουμε την πιθανότητα εμφάνισης της σωστής άποψης (ταμπέλας).

Βιβλιογραφία

- Pattern Recognition 3rd Edition,Teodoridis S./ Koutroumbas K., Academic Press
- Floating Search Methods in Feature Selection, P.Pudil/ J. Novovicova,/J. Kittler
- Hybrid Genetic Algorithms for Feature Selection, Il-Seok Oh/ Jin-Seon Lee/ Byung-Ro Moon
- Feature Selection Algorithms: A Survey and Experimental Evaluation, Luis Carlos Molina/ Lluis Belanche/Angela Nebot
- Chapter 15, Infrared Spectroscopy: Theory, Online edition for students of organic chemistry lab courses at the University of Colorado, Boulder, Dept of Chem. and Biochem.
- Τεχνητά Νευρωνικά Δίκτυα, Κ. Διαμαντάρας, εκδ. Κλειδάριθμος
- Disease recognition by infrared and Raman spectroscopy, Christoph Krafft/ Gerald Steiner/Claudia Beleites/ Reiner Salzer
- M. Negnevitsky, Artificial Intelligence, A Guide to Intelligent Systems, Pearson Education (Addison-Wesley), 2002
- Pattern Classification, Duda R./Hart P./Stork D.