



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΤΟΜΕΑΣ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΑΝΑΠΤΥΞΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΑΛΓΟΡΙΘΜΩΝ ΤΕΧΝΗΤΩΝ
ΜΕΛΙΣΣΩΝ ΣΕ ΠΡΟΒΛΗΜΑΤΑ ΤΜΗΜΑΤΟΠΟΙΗΣΗΣ
ΚΑΤΑΝΑΛΩΤΙΚΗΣ ΑΓΟΡΑΣ

Διατριβή που υπεβλήθη για την μερική ικανοποίηση των απαιτήσεων για
την απόκτηση του μεταπτυχιακού διπλώματος ειδίκευσης

Ιωάννης - Αντώνιος Α. Ντουσάκης

Χανιά, Ιούνιος 2015

Ευχαριστίες

Η παρούσα μεταπτυχιακή διατριβή, εκπονήθηκε στον τομέα Οργάνωσης και Διοίκησης και συγκεκριμένα στο εργαστήριο υποστήριξης αποφάσεων του Πολυτεχνείου Κρήτης.

Υπεύθυνος κατά την εκπόνηση της μεταπτυχιακής μου διατριβής ήταν ο λέκτορας κ. Στέλιος Τσαφαράκης στον οποίο οφείλω ιδιαίτερες ευχαριστίες για την ανάθεσή της καθώς μέσα από το συγκεκριμένο θέμα μου δόθηκε η ευκαιρία να ασχοληθώ με έναν πραγματικά ενδιαφέροντα ερευνητικό τομέα. Με την ευκαιρία θα ήθελα να εκφράσω τις ευχαριστίες μου και στον επίκουρο καθηγητή κ. Ιωάννη Μαρινάκη ο οποίος με τις τεχνικές του γνώσεις με βοήθησε στην υλοποίηση συγκεκριμένων μερών της εργασίας καθώς και στον υποψήφιο διδάκτορα Κώστα Μακαντάση ο οποίος προσέφερε την βοήθειά του στην ανάλυση των σετ δεδομένων.

Κλείνοντας θα ήθελα να εκφράσω την ευγνωμοσύνη μου στην οικογένειά μου και την σύντροφό μου για την αμέριστη εμπιστοσύνη και συμπαράσταση που μου έδειξαν αυτά τα δύο χρόνια του μεταπτυχιακού προγράμματος.

Η ολοκλήρωση της διπλωματικής εργασίας συγχρηματοδοτήθηκε μέσω του Έργου «Υποτροφίες ΙΚΥ» από πόρους του ΕΠ «Εκπαίδευση και Δια Βίου Μάθηση», του Ευρωπαϊκού Κοινωνικού Ταμείου (ΕΚΤ) του ΕΣΠΑ, 2007-2013.

Περίληψη

Στην παρούσα εργασία θα αναπτυχθούν τρεις διαφορετικοί αλγόριθμοι οι οποίοι βασίζονται στη συμπεριφορά των μελισσών στη φύση και οι οποίοι χωρίζονται σε δύο κύριες κατηγορίες. Στην μίμηση της συμπεριφοράς αναζήτησης τροφής (foraging behaviour) και στην μίμηση της συμπεριφοράς ζευγαρώματος (mating behavior). Αφού γίνει η απαραίτητη προσαρμογή των αλγορίθμων και η βελτιστοποίηση των παραμέτρων τους, θα εφαρμοστούν σε προβλήματα τμηματοποίησης καταναλωτικής αγοράς τα οποία ανήκουν στην κατηγορία των NP-hard προβλημάτων clustering. Θα πραγματοποιηθούν προσομοιώσεις σε τεχνητά και πραγματικά δεδομένα και θα εξεταστεί η πρακτική εφαρμογή τους. Ως μέτρα απόδοσης των αλγορίθμων θα χρησιμοποιηθούν αφενός η ποιότητα της λύσης (δηλαδή το κατά πόσο βελτιστοποιεί την αντικειμενική συνάρτηση) και αφετέρου ο χρόνος επίλυσης του προβλήματος και το υπολογιστικό φορτίο.

Λέξεις κλειδιά: Τμηματοποίηση αγοράς, βελτιστοποίηση, αλγόριθμοι μελισσών, K-means

Abstract

In this thesis three different optimization algorithms will be developed, which are mainly inspired by the behavior of bees in the nature and can be divided into two categories. The first one contains algorithms that are based on the foraging behavior of bees, while the second one contains algorithms that are based on the mating behavior of bees. Once the necessary adjustments are made and the parameters of the algorithms are correctly tuned, the algorithms will be tested on market segmentation problems, which belong to the NP-hard clustering problems. Simulations will be executed on artificial and real data in order to examine the practicality of the algorithms. As performance indexes we will use the value of the optimal solution as well as the time and computational cost.

Keywords: Market segmentation, optimization, bees algorithms, K-means

Περιεχόμενα

Ευχαριστίες.....	3
Περίληψη	5
Abstract	6
1. Εισαγωγή.....	11
1.1. Αντικείμενο της μεταπτυχιακής διατριβής.....	11
1.2. Στόχος.....	11
1.3. Οργάνωση του τόμου.....	12
2. Η τμηματοποίηση της αγοράς.....	12
2.1. Η έννοια της αγοράς και οι κατηγορίες της.....	12
2.2. Τμηματοποίηση.....	13
2.3. Επίπεδα τμηματοποίησης της αγοράς.....	17
2.4. Διαδικασία τμηματοποίησης της αγοράς.....	18
2.5. Κριτήρια τμηματοποίησης της αγοράς καταναλωτών.....	19
2.6. Οφέλη τμηματοποίησης.....	23
2.7. Η βιομηχανική αγορά.....	24
2.8. Κριτήρια τμηματοποίησης της βιομηχανικής αγοράς.....	25
3. Κριτήρια τμηματοποίησης.....	28
3.1. Μέτρα Απόστασης.....	28
3.2. Αξιολόγηση της ποιότητας της τμηματοποίησης.....	31
4. Μέθοδοι τμηματοποίησης.....	35
4.1. Ιεραρχικές Μέθοδοι.....	35
4.2. Διαχωριστικές (Partitioning) Μέθοδοι.....	38
4.3. Αλγόριθμοι που ελαχιστοποιούν κάποιο σφάλμα.....	39
4.4. Μέθοδοι που βασίζονται σε θεωρία γράφων.....	40
4.5. Μέθοδοι που βασίζονται στην πυκνότητα.....	40
5. Προσδιορισμός του αριθμού των τμημάτων.....	41
5.1. Μέθοδοι που βασίζονται στην διασπορά εντός των τμημάτων.....	42

5.2.	Μέθοδοι που βασίζονται στην διασπορά εντός και μεταξύ των τμημάτων.....	43
6.	Αλγόριθμοι εμπνευσμένοι από τη φύση	45
6.1.	Αλγόριθμος Βελτιστοποίησης Ζευγαρώματος Μελισσών (Honey Bees Mating Optimization – HBMO).....	46
6.2.	Αλγόριθμος Βελτιστοποίησης Ζευγαρώματος Μπάμπουρων (Bumble Bees Mating Optimization – BBMO)	50
6.3.	Αλγόριθμος Τεχνητής Αποικίας Μελισσών (Artificial Bee Colony Optimization – ABC)	53
7.	Προσαρμογή των αλγορίθμων στο πρόβλημα.....	55
7.1.	Συνάρτηση ποιότητας	55
7.2.	Αναπαράσταση λύσεων	55
7.3.	Ειδικά θέματα στη διακριτή αναπαράσταση της λύσης	58
7.4.	Τελεστές αλγορίθμων	61
8.	Ρύθμιση των παραμέτρων των αλγορίθμων	66
8.1.	Σετ δεδομένων	66
8.2.	Συνεχής Αναπαράσταση	67
8.3.	Διακριτή αναπαράσταση	87
8.4.	Σύνοψη αποτελεσμάτων διαδικασίας ρύθμισης.....	94
9.	Αξιολόγηση σε γνωστά σετ δεδομένων	95
9.1.	Αποτελέσματα στο Iris Dataset.....	96
9.2.	Αποτελέσματα στο Wine Dataset	97
9.3.	Αποτελέσματα στο Wholesale Customer Dataset.....	97
10.	Συμπεράσματα	101
11.	Προτάσεις για μελλοντική έρευνα	103
	Βιβλιογραφία	104
	Παράρτημα	107
	Σετ δεδομένων Aggregator.....	107

Κατάλογος εικόνων, πινάκων και γραφημάτων

Εικόνα 1 – Agglomerative και divisive τμηματοποίηση.....	36
Εικόνα 2 – Δενδρογράμμα για μία τμηματοποίηση. Η κόκκινη γραμμή αντιστοιχεί στο ανεκτό επίπεδο αποστάσεων.....	36
Εικόνα 3 – Γραφική αναπαράσταση των μεθόδων μονής, πλήρους και μέσου όρου ακμής.....	37
Εικόνα 4 - Αναπαράσταση Λύσης μέσω των συντεταγμένων του κέντρου του τμήματος (α' τρόπος)	56
Εικόνα 5 - Αναπαράσταση Λύσης μέσω των συντεταγμένων του κέντρου του τμήματος (β' τρόπος).....	56
Εικόνα 6 - Αναπαράσταση λύσης μέσω των τμημάτων που ανήκει το κάθε αντικείμενο	57
Εικόνα 7 - Αναπαράσταση λύσης μέσω της αναπαράστασης των αντικειμένων-προτύπων	57
Εικόνα 8 - Το πρόβλημα της πολλαπλής αναπαράστασης λύσεων	60
Εικόνα 9 – Γραφική αναπαράσταση μίας επανάληψης του αλγορίθμου K-means	65
Εικόνα 10 - Το σετ δεδομένων Aggregation ταξινομημένο σωστά.....	66
Εικόνα 11 – Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος HBMO περ.Α-Α.....	69
Εικόνα 12 - Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος HBMO περ.Β-Α.....	71
Εικόνα 13 - Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος ABC.....	77
Εικόνα 14 - Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος BBMO περ.Β-Α.....	82
Εικόνα 15 -Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος BBMO περ.Β-Β.....	84
Εικόνα 16 – Διάγραμμα Silhouette για 4 τμήματα	98
Εικόνα 17 - Διάγραμμα Silhouette για 5 τμήματα.....	98
Εικόνα 18 - Διάγραμμα Silhouette για 6 τμήματα.....	99
Πίνακας 1 – Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Α-Α.....	67
Πίνακας 2 – Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Β-Α	70
Πίνακας 3 - Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Γ-Α	72
Πίνακας 4 - Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Β-Β	73
Πίνακας 5 Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Γ-Β.....	75
Πίνακας 6 - Αποτελέσματα διαδικασίας ρύθμισης ABC.....	76
Πίνακας 7 - Αποτελέσματα διαδικασίας ρύθμισης BBMO περ.Α-Α	78
Πίνακας 8 - Αποτελέσματα διαδικασίας ρύθμισης BBMO περ.Α-Β.....	79

Πίνακας 9 - Αποτελέσματα διαδικασίας ρύθμισης BBMO περ.Β-Α.....	81
Πίνακας 10 - Αποτελέσματα διαδικασίας ρύθμισης BBMO περ.Β-Β.....	83
Πίνακας 11 - Αποτελέσματα διαδικασίας ρύθμισης BBMO περ.Γ-Α.....	85
Πίνακας 12 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Γ-Β.....	87
Πίνακας 13 - Αποτελέσματα διαδικασίας ρύθμισης διακριτού HBMO περ.Α-Δ.....	88
Πίνακας 14 – Αποτελέσματα διαδικασίας ρύθμισης διακριτού HBMO με χρήση του τελεστή mutation.....	90
Πίνακας 15 - Αποτελέσματα διαδικασίας ρύθμισης διακριτού HBMO με χρήση του τελεστή mutation και KMO.....	91
Πίνακας 16 - Αποτελέσματα διαδικασίας ρύθμισης HBMO medoid περ. Α-Δ.....	93
Πίνακας 17 – Αποτελέσματα στο Iris Dataset.....	96
Πίνακας 18 - Αποτελέσματα στο Iris dataset για την διακριτή αναπαράσταση λύσης.....	96
Πίνακας 19 – Αποτελέσματα στο Wine Dataset.....	97
Πίνακας 20-Αποτελέσματα στο Wine dataset για την διακριτή αναπαράσταση λύσης.....	97
Πίνακας 21 – Αποτελέσματα στο WC dataset για 4 clusters.....	99
Πίνακας 22 - Αποτελέσματα στο WC dataset για την διακριτή αναπαράσταση λύσης για 4 clusters.....	100
Πίνακας 23 - Αποτελέσματα στο WC dataset για 5 clusters.....	100
Πίνακας 24-Αποτελέσματα στο WC dataset για την διακριτή αναπαράσταση λύσης για 5 clusters.....	100
Πίνακας 25 - Αποτελέσματα στο WC dataset για 6 clusters.....	100
Πίνακας 26 - Αποτελέσματα στο WC dataset για την διακριτή αναπαράσταση λύσης για 6 clusters.....	100
Γράφημα 1 -Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Α-Α.....	68
Γράφημα 2 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Β-Α.....	70
Γράφημα 3 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Γ-Α.....	72
Γράφημα 4 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Β-Β.....	74
Γράφημα 5 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Γ-Β.....	75
Γράφημα 6 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Α-Β.....	80
Γράφημα 7 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Β-Α.....	81
Γράφημα 8 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Β-Β.....	84
Γράφημα 9 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Γ-Α.....	86
Γράφημα 10 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ. Γ-Β.....	87
Γράφημα 11 - Ποιότητα λύσης με το χρόνο για τον διακριτό αλγόριθμο HBMO περ.Α-Δ.....	89
Γράφημα 12 – Ποιότητα λύσης με το χρόνο για τον διακριτό αλγόριθμο HBMO με χρήση του τελεστή mutation.....	90
Γράφημα 13 - Ποιότητα λύσης με το χρόνο για τον διακριτό αλγόριθμο HBMO με χρήση του τελεστή mutation και KMO.....	92
Γράφημα 14 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO medoid περ.Α-Δ.....	93

1. Εισαγωγή

1.1. Αντικείμενο της μεταπτυχιακής διατριβής

Τα προϊόντα ή οι υπηρεσίες που προσφέρει μια επιχείρηση δεν είναι εφικτό να ικανοποιούν το σύνολο των αγοραστών, ιδιαίτερα σήμερα με την βελτίωση της τεχνολογίας και συνεπώς την ενίσχυση του ανταγωνισμού όσο αφορά τα προσφερόμενα αγαθά/υπηρεσίες. Για να μπορέσει μια επιχείρηση να αποκτήσει συγκριτικό πλεονέκτημα, λοιπόν, σε ένα τόσο ανταγωνιστικό πλαίσιο θα πρέπει να προσδιορίσει με ακρίβεια τα τμήματα εκείνα της αγοράς που την ενδιαφέρουν και στα οποία θέλει να δώσει μεγαλύτερη βαρύτητα. Η διαδικασία της τμηματοποίησης πετυχαίνει αυτό ακριβώς· συμβάλλει ώστε να εντοπιστούν οι αγορές-στόχοι και να ικανοποιηθούν με τον καλύτερο δυνατό τρόπο οι ανάγκες τους.

Πιο συγκεκριμένα, η τμηματοποίηση της αγοράς, αφορά στον διαχωρισμό της αγοράς σε πιο μικρά τμήματα, με τέτοιο τρόπο ώστε το καθένα να μπορεί να αντιμετωπιστεί σαν μια μικρή ομάδα. Η τμηματοποίηση της αγοράς στηρίζεται στη συμπεριφορά των καταναλωτών, αλλά και στα ιδιαίτερα χαρακτηριστικά τους. Ενδιαφέρεται συνεπώς για την αξία και τον όγκο των αγορών, την τιμή, την συχνότητα με την οποία αγοράζουν οι καταναλωτές, από πού προτιμάνε να αγοράζουν κ.ά.

Αντικείμενο της παρούσας μεταπτυχιακής διατριβής είναι η μελέτη της επίδοσης των αλγορίθμων που βασίζονται στην συμπεριφορά των μελισσών όσον αφορά την επίλυση του προβλήματος της τμηματοποίησης δεδομένων. Αφού τροποποιηθούν με κατάλληλο τρόπο οι αλγόριθμοι και ρυθμιστούν οι παράμετροι, θα γίνει εφαρμογή και σε δεδομένα που αφορούν την καταναλωτική αγορά.

1.2. Στόχος

Στόχος της εργασίας είναι να εξεταστεί η επίδοση των αλγορίθμων μελισσών στο πρόβλημα της τμηματοποίησης. Επιμέρους στόχοι της εργασίας είναι να γίνει βιβλιογραφική ανασκόπηση των υπάρχουσών μεθοδολογιών τόσο όσον αφορά τους αλγορίθμους αλλά και στις αντικειμενικές συναρτήσεις που χρησιμοποιούνται από τις διάφορες ερευνητικές ομάδες. Παράλληλα, εξετάζονται διάφορες παραλλαγές των αλγορίθμων και συγκρίνονται με ίσους όρους όσον αφορά την επίδοση τους καθώς και το υπολογιστικό τους φόρτο. Αξιοσημείωτη είναι βέβαια και η συμβολή της εργασίας στην μετέπειτα έρευνα, μιας και οι κώδικες που δημιουργήθηκαν είναι εύκολα επεκτάσιμοι και προσαρμόσιμοι σε διάφορες αντικειμενικές συναρτήσεις.

1.3. Οργάνωση του τόμου

Η εργασία είναι οργανωμένη ως εξής. Στο κεφάλαιο 2 παρουσιάζουμε μία θεωρητική προσέγγιση στο πρόβλημα της τμηματοποίησης της αγοράς, τονίζοντας τα είδη τμηματοποίησης καθώς και τη σημασία της για τη σύγχρονη εταιρία. Στο κεφάλαιο 3 προχωρούμε στην παράθεση των κριτηρίων τμηματοποίησης, δηλαδή των διαφόρων αντικειμενικών συναρτήσεων που έχουν προταθεί στη βιβλιογραφία. Στο κεφάλαιο 4 παραθέτουμε τις κυριότερες κατηγορίες μεθόδων τμηματοποίησης. Στο κεφάλαιο 5 περιγράφουμε τις μεθόδους που έχουν προταθεί για την εκτίμηση του αριθμού των τμημάτων σε ένα σετ δεδομένων. Στο κεφάλαιο 6 περιγράφουμε αναλυτικά τους αλγορίθμους που θα εφαρμόσουμε και στο κεφάλαιο 7 αναφέρουμε τον τρόπο με τον οποίο προσαρμόζονται στο πρόβλημα. Στο κεφάλαιο 8 ρυθμίζουμε τις παραμέτρους τους για τις διάφορες παραλλαγές τους και τις διάφορες αναπαραστάσεις των λύσεων. Στο κεφάλαιο 9 εφαρμόζουμε τους αλγορίθμους και αξιολογούμε τα αποτελέσματά τους σε γνωστά σετ δεδομένων. Τέλος, στο κεφάλαιο 10, παρουσιάζουμε τα συμπεράσματα και στο κεφάλαιο 11 τις προτάσεις για μελλοντική έρευνα.

2. Η τμηματοποίηση της αγοράς

2.1. Η έννοια της αγοράς και οι κατηγορίες της

Η λέξη αγορά, ανάλογα με τη συγκεκριμένη περίπτωση στην οποία χρησιμοποιείται, έχει διαφορετικές έννοιες. Για παράδειγμα όταν αναφερόμαστε στην αγορά της οδού Αθηνάς, ή στην αγορά της Κύπρου, τότε η αγορά αναφέρεται σε συγκεκριμένη γεωγραφική περιοχή. Σε άλλες περιπτώσεις λέμε ότι η αγορά του πετρελαίου ή η αγορά του χρυσού είναι αδύνατη, όπου η αγορά αναφέρεται σε συγκεκριμένο προϊόν. Όταν ο καταναλωτής λέει ότι πήγε στην αγορά, εννοεί ότι πήγε εκεί όπου υπάρχουν τα καταστήματα, δηλαδή η προσφορά. Όταν πάλι λέει ότι η αγορά ενός προϊόντος ήταν γι' αυτόν δύσκολη, εννοεί ότι οι συνθήκες προσφοράς (π.χ. τιμή, τοποθεσία, διαθεσιμότητα προϊόντος, όροι πληρωμής, συμπεριφορά πωλητών, εξυπηρέτηση μέσα στο μαγαζί) δεν ήταν αυτές που θα επιθυμούσε να ήταν.

Για το Μάρκετινγκ, αγορά είναι ένα σύνολο αγοραστών με ανικανοποίητες ανάγκες, που έχουν τόσο τη δυνατότητα όσο και την αποφασιστικότητα να τις ικανοποιήσουν. Με βάση αυτή την περιγραφή της έννοιας, για να υπάρχει αγορά, θα πρέπει να πληρούνται και οι εξής τρεις προϋποθέσεις:

- Να υπάρχουν ανάγκες (σε ευρεία έννοια πάντοτε, για να συμπεριλαμβάνονται και οι επιθυμίες)
- Οι αγοραστές να έχουν τη δυνατότητα ικανοποίησης αυτών των αναγκών (δηλαδή τους οικονομικούς πόρους, τα εισοδήματα, τις αποταμιεύσεις, την πίστωση)
- Οι αγοραστές να θέλουν, να είναι αποφασισμένοι, να επιδιώκουν την ικανοποίηση των αναγκών τους

Όσο αυξάνονται οι ανάγκες, οι δυνατότητες και η αποφασιστικότητα, τόσο αυξάνεται το μέγεθος της αγοράς που για το Μάρκετινγκ σημαίνει ότι αυξάνονται οι δυνατότητες της αγοράς και οι δυνατότητες πωλήσεων. Οι ανάγκες αυξάνονται, όταν αυξάνεται ο αριθμός των αγοραστών, με αμετάβλητες τις ανάγκες του καθενός από αυτούς, όταν αυξάνονται αυτές οι ίδιες οι ανάγκες, με αμετάβλητο τον αριθμό των αγοραστών, ή (και αυτό είναι το πιο συνηθισμένο) όταν αυξάνονται και τα δύο.

Οι δυνατότητες ικανοποίησης των αναγκών αυξάνονται, όταν αυξάνεται το εισόδημα και η περιουσία των αγοραστών. Τέλος, η αποφασιστικότητα αυξάνεται, όταν αυξάνεται η ροπή για δαπάνη του εισοδήματος. Το μέγεθος της αγοράς επηρεάζει όχι μόνο το μέγεθος της παραγωγής (άρα και το Μάρκετινγκ) αλλά και το κόστος παραγωγής.

2.2. Τμηματοποίηση

Είναι γεγονός ότι διαφορετικοί καταναλωτές έχουν και διαφορετικές προτιμήσεις, ανάγκες, ενδιαφέροντα καθώς επίσης και διαφορετικούς πόρους με τους οποίους μπορούν να καλύψουν αυτές τις ανάγκες.

Αυτό γίνεται εύκολα κατανοητό από την πληθώρα των προϊόντων που μπορεί να βρει κάποιος στα καταστήματα, όπως για παράδειγμα σε ένα σουπερμάρκετ, η οποία αυτή ποικιλία είναι αποτέλεσμα των διαφορετικών αγοραστικών συνηθειών των καταναλωτών, αλλά και της διαφορετικής ωφέλειας που απολαμβάνουν καταναλώνοντας διαφορετικά προϊόντα. Είναι σχεδόν ακατόρθωτο να ικανοποιήσει η αγορά όλους τους καταναλωτές με ένα μόνο προϊόν ή υπηρεσία [1]. Η βασική αυτή διαπίστωση λοιπόν ότι όλοι οι αγοραστές δεν είναι ίδιοι, οδήγησε στην έννοια της τμηματοποίησης της αγοράς (Market Segmentation).

Τμηματοποιώ μια αγορά, σε ευρύτερα πλαίσια, σημαίνει ότι την μελετώ με σκοπό να εντοπίσω τα τμήματά της (υποσύνολα αγοραστών), να υπολογίσω το μέγεθος καθενός από αυτά, να βρω τις ανάγκες ενός ή μερικών τμημάτων, να βρω πώς ικανοποιούνται και να προσπαθήσω να ικανοποιήσω αυτές τις ανάγκες με το κατάλληλο μίγμα μάρκετινγκ (MKT) . Σε στενότερα πλαίσια, τμηματοποιώ μια αγορά σημαίνει ψάχνω να βρω υποσύνολα αγοραστών που είτε:

- Δεν αγοράζουν καθόλου το προϊόν, για παράδειγμα, δεν αγοράζουν κινητό τηλέφωνο, επειδή και οι τρεις μάρκες προϋποθέτουν διαδικασίες τεχνικής φύσης που τους δημιουργούν δυσφορία.
- Αγοράζουν το προϊόν, αλλά σε μικρότερη σχετικά ποσότητα από εκείνη που δικαιολογεί το εισόδημά τους, επειδή δεν ικανοποιούνται πλήρως οι ανάγκες τους από την χρήση του. Για παράδειγμα, δεν αγοράζουν πολύ φρέσκο γάλα, γιατί το πίνουν λίγο - λίγο και χαλαί.
- Αγοράζουν το προϊόν σε κανονική ποσότητα, αλλά διατυπώνουν πολλά παράπονα. Παραδειγματος χάρη, συναλλάσσονται συχνά με τις τράπεζες αλλά δεν είναι ευχαριστημένοι από το επίπεδο εξυπηρέτησης.

Για παράδειγμα, μια εκδοτική εταιρία που εκδίδει εγκυκλοπαίδειες, τα τμήματα της αγοράς με κριτήρια ποιοι και γιατί αγοράζουν, μπορεί να είναι τα εξής:

1. Τμήμα 1. Μορφωμένοι ενήλικες με μέσο ή υψηλό εισόδημα οι οποίοι ενδιαφέρονται κυρίως για την ποιότητα. (δηλαδή έγκυρη πληροφόρηση).
2. Τμήμα 2. Άτομα που σπουδάζουν και εξαρτώνται οικονομικά από τους γονείς τους οι οποίοι ενδιαφέρονται κυρίως για απλή και φθηνή πληροφόρηση.
3. Τμήμα 3. Άτομα χαμηλής κυρίως μόρφωσης που θέλοντας να αποκτήσουν αυτό το μειονέκτημά τους, ενδιαφέρονται για όγιο που εντυπωσιάζει (πολλοί τόμοι).
4. Τμήμα 4. Άτομα ιδιαίτερα ευαίσθητα στη διακρίση του σπιτιού τους που ενδιαφέρονται βασικά για την αισθητική πλευρά της εγκυκλοπαίδειας.

Στο παράδειγμα που αναφέρθηκε, έγινε μια διαίρεση της αγοράς σε διάφορα τμήματα. Το βασικό ερώτημα που ανακύπτει είναι το κατά πόσο αυτή ήταν επιτυχής ή όχι. Η τμηματοποίηση θεωρείται επιτυχής για όταν:

1. Οι αγοραστές κάθε τμήματος αποτελούν μεταξύ τους μια ομοιογενή ομάδα, δηλαδή συμπεριφέρονται κατά τον ίδιο περίπου τρόπο.
2. Οι αγοραστές καθενός από τα τμήματα διαφέρουν από τους αγοραστές των άλλων τμημάτων, δηλαδή οι αγοραστές του συνόλου της αγοράς αποτελούν μεταξύ τους μια ετερογενή ομάδα.
3. Τουλάχιστον ένα από τα τμήματα είναι αρκετά μεγάλο, ώστε να έχει ενδιαφέρον από οικονομικής άποψης, δηλαδή να μπορεί να καθορισθεί σαν αγορά στόχος.
4. Το τμήμα που θα καθορισθεί σαν αγορά στόχος, να μπορεί να προσεγγισθεί από το μίγμα ΜΚΤ, δηλαδή δια μέσου της διανομής και της προβολής.

Σήμερα οι επιχειρήσεις δεν συμμετέχουν στο μαζικό μάρκετινγκ καθώς εστίασαν περισσότερο σε μια στρατηγική μάρκετινγκ, όπου ο κάθε στόχος είναι μια συγκεκριμένη ομάδα καταναλωτών. Αυτός ο προσδιορισμός των ομάδων πελατών - στόχος αποτελεί τον κατακερματισμό της αγοράς, όπου οι πελάτες συγκεντρώνονται σε ομάδες με παρόμοια αγοραστικά χαρακτηριστικά και ανάγκες από την αγορά [2].

Η λογική αυτή της τμηματοποίησης της αγοράς έχει παίξει πολύ σημαντικό ρόλο στον επιχειρηματικό κόσμο διότι βασίζεται στην παραδοχή ότι οι καταναλωτές παρουσιάζουν ανομοιογενή χαρακτηριστικά όσον αφορά τις προτιμήσεις τους σε προϊόντα αλλά και γενικότερα την αγοραστική τους συμπεριφορά. Αυτή η μεταβλητότητα αποδίδεται στις διαφορές που υπάρχουν στα προϊόντα αλλά και στις διαφορές στα χαρακτηριστικά των καταναλωτών.

Έτσι, η τμηματοποίηση της αγοράς βοηθάει τις επιχειρήσεις να αντιμετωπίζουν αυτήν την ανομοιογένεια αποτελεσματικά, εξισορροπώντας την μεταβλητότητα των αναγκών των πελατών με τους περιορισμένους διαθέσιμους πόρους.

Θεωρητικά, η τμηματοποίηση της αγοράς είναι κάτι το σχετικά εύκολο. Στην πράξη όμως, για να είναι επιτυχής αυτή η τμηματοποίηση, θα πρέπει να ικανοποιούνται και οι τέσσερις προϋποθέσεις που αναφέρθηκαν πιο πάνω. Το ποια είναι η πιο δύσκολη προϋπόθεση εξαρτάται από το περιβάλλον, την επιχείρηση και το προϊόν. Το περιβάλλον και το προϊόν συνδέονται και με τις τέσσερις προϋποθέσεις, ενώ η επιχείρηση με τις δύο τελευταίες.

Το περιβάλλον επηρεάζει την ομοιογένεια της ομάδας, την ετερογένεια του συνόλου της αγοράς, τη δυνατότητα προσέγγισης των τμημάτων αλλά και η φύση του προϊόντος επηρεάζει τα ίδια όπως και το περιβάλλον. Επιπλέον, αποτελεί πραγματικότητα το γεγονός ότι όσα περισσότερα γνωρίζουν οι επιχειρήσεις για τους πελάτες τους, τόσο καλύτερα μπορούν να τους εξυπηρετήσουν και να καλύψουν τις ανάγκες τους και σε αυτό συμβάλλει αρκετά ανάπτυξη της τεχνολογίας [3] μέσω του big data κ.ά..

Τέλος, τα χαρακτηριστικά της επιχείρησης (π.χ. πόροι, φήμη) επηρεάζουν το κατά πόσο ένα τμήμα κρίνεται σαν αρκετά μεγάλο και το κατά πόσο αυτό το τμήμα, σαν αγορά στόχος, μπορεί να προσεγγισθεί από την επιχείρηση αυτή.

Οι οικονομικά αναπτυγμένες χώρες έχουν περιβάλλον τέτοιο όπου η επιτυχής τμηματοποίηση μιας αγοράς τους γίνεται πιο εύκολα κάτι το οποίο δεν συμβαίνει στις οικονομικά υποανάπτυκτες χώρες. Σε ότι αφορά τη φύση του προϊόντος μπορεί να υποστηριχθεί ότι μόνο οι αγορές μερικών προϊόντων δεν προσφέρονται για τμηματοποίηση.

Τέλος, όσο αυξάνεται το μέγεθος μιας επιχείρησης, τόσο γίνεται πιο δύσκολη η τρίτη προϋπόθεση και τόσο γίνεται εύκολη η τέταρτη. Αντίθετα, όσο μειώνεται το μέγεθος, τόσο γίνεται εύκολη η τρίτη και τόσο δυσκολεύει η τέταρτη.

Για να διευκρινισθεί επαρκώς η τμηματοποίηση σαν έννοια αλλά κυρίως σαν στρατηγική, θα πρέπει να ξεχωρίσουμε ότι άλλο είναι επιτυχής τμηματοποίηση (που προηγείται) και άλλο επιτυχής επιλογή τμήματος (που έπεται). Επιτυχής τμηματοποίηση είναι αυτή που πληροί και τις τέσσερις προϋποθέσεις.

Τότε και μόνο τότε τα τμήματα που προκύπτουν είναι ξεχωρισμένα, με αγοραστές μέσα σε κάθε τμήμα με ίδιες ανάγκες, με ανάγκες που σαφώς διαφοροποιούνται από τμήμα σε τμήμα, επαρκή ή ελκυστικά σε μέγεθος, δηλαδή σε πιθανολογούμενη ποσότητα που θα ζητήσουν, και τέλος προσεγγίσιμα - προσπελάσιμα με τη διανομή - προβολή.

Η επιτυχής τμηματοποίηση καταλήγει με την καταγραφή και κατονομασία των τμημάτων της αγοράς, όπως αυτά προέκυψαν με βάση ένα ή περισσότερα κριτήρια και αφού βέβαια ελέγχθηκε ότι πληρούν τις τέσσερις προϋποθέσεις. Το επόμενο βήμα είναι η επιλογή ενός ή περισσότερων τμημάτων από την επιχείρηση, ώστε να αποτελέσουν το μίγμα αγορών στόχων.

Επιτυχή επιλογή τμήματος έχουμε, όταν το τμήμα:

1. Έχει προοπτική μεγέθυνσης.
2. Δεν έλκει ή αναμένεται ότι δεν θα έλξει στο μέλλον την προσοχή ανταγωνιστών καθώς όσο λιγότεροι ανταγωνιστές σε ένα τμήμα τόσο το καλύτερο.
3. Δεν απαιτεί δυσανάλογο κόστος ικανοποίησης των ιδιαίτερων αναγκών του διότι όσο μικρότερες αλλαγές στο Μίγμα Μάρκετινγκ, τόσο μειώνεται το κόστος.
4. Δημιουργεί έσοδα τα οποία συγκρινόμενα με τα έξοδα, προσθέτουν κέρδη που μεγιστοποιούν την αποδοτικότητα της επιχείρησης.
5. Η ικανοποίηση των αναγκών του ταιριάζει με την αποστολή, την εικόνα και γενικά τους αντικειμενικούς στόχους της επιχείρησης.

Για να είναι επιτυχημένη λοιπόν μια τμηματοποίηση καθίσταται απαραίτητο τα τμήματα που θα προκύψουν από αυτή να εντοπίζονται σύμφωνα με τα κύρια χαρακτηριστικά τους, έπειτα να είναι ικανοποιητικού μεγέθους ώστε να αξίζει η επένδυση και να έχουν οικονομικό ενδιαφέρον και τέλος να είναι σταθερά [4].

Η τμηματοποίηση αγοράς και η επιλογή τμήματος, επειδή είναι στρατηγικής φύσης, θα πρέπει να προσεγγίζονται με βάση την ενδελεχή μελέτη κόστους/ωφέλειας και με ζητούμενο την ελαχιστοποίηση του κόστους παράλληλα με τη μεγιστοποίηση της ωφέλειας [5].

Πριν προχωρήσει κάποιος ερευνητής στην τμηματοποίηση της αγοράς στόχου πρέπει να έχει την δυνατότητα η συμβατική αγορά του εν λόγω προϊόντος να επεκταθεί ώστε να μπορεί να πραγματοποιηθεί και η τμηματοποίησης της [6]. Η διαδικασία της τμηματοποίησης σε πρώτο επίπεδο περιλαμβάνει τον εντοπισμό των κριτηρίων της τμηματοποίησης και δεύτερον την ανάπτυξη της περιγραφής κάθε τμήματος ξεχωριστά. Η στόχευση περιλαμβάνει την αξιολόγηση της ελκυστικότητας του κάθε τμήματος και την επιλογή τμήματος-στόχου [7].

Η τμηματοποίηση της αγοράς δεν έχει νόημα όταν η αγορά είναι πολύ μικρή σε μέγεθος, όταν οι προτιμήσεις των καταναλωτών είναι αρκετά ομοιογενείς και όταν οι βαριοί χρήστες του προϊόντος αντιπροσωπεύουν το συντριπτικά μεγαλύτερο ποσοστό του συνολικού όγκου πωλήσεων στην αγορά οπότε αποτελούν και το μοναδικό τμήμα-στόχο [6].

Σε περίπτωση που κρίνεται μη απαραίτητη η τμηματοποίηση της αγοράς, ακολουθείται μια στρατηγική η οποία σσκοπεύει στην ενίσχυση της αρχικής ζήτησης για το προϊόν. Αυτό σημαίνει ότι πρώτον εντοπίζονται καινούργιοι χρήστες για το προϊόν με σκοπό την αύξηση του μεριδίου της αγοράς, δεύτερον, αύξηση της χρήσης του προϊόντος βελτιώνοντας την τωρινή κατανάλωση και τρίτον επέκταση των περιστάσεων χρήσης του προϊόντος, εντοπίζοντας εναλλακτικές χρήσεις [8].

2.3. Επίπεδα τμηματοποίησης της αγοράς

Η τμηματοποίηση της αγοράς μπορεί έχει τέσσερις μορφές. Μπορεί να γίνει δηλαδή σε επίπεδο τμημάτων (segments), σε επίπεδο φωλεών (niches), σε τοπικό και τέλος σε ατομικό επίπεδο [9].

Τα επίπεδα αυτά αναλύονται παρακάτω:

- **Μάρκετινγκ τμήματος:** Η έννοια του τμήματος της αγοράς σχετίζεται μια μεγάλη ομάδα μέσα στην αγορά. Χρησιμοποιώντας το μάρκετινγκ τμήματος μία επιχείρηση, συμφωνεί ότι οι αγοραστές διαφέρουν ως προς την αγοραστική τους δύναμη ανάλογα με το εισόδημα τους παρουσιάζοντας διαφορετικές επιθυμίες και αγοραστικές συνήθειες. Οι καταναλωτές που ανήκουν σε κάθε τμήμα θεωρούνται ότι είναι αρκετά όμοιοι ως προς τις επιθυμίες τους και τις ανάγκες τους, χωρίς αυτό να σημαίνει έχουν κάποιες διαφορές και δεν είναι πανομοιότυποι.
- **Μάρκετινγκ φωλέας:** Η φωλέα αγοράς είναι περισσότερο συγκεκριμένη ομάδα μέσα σε ένα τμήμα και αποτελείται από καταναλωτές που μπορεί να έχουν περισσότερο εξειδικευμένες ανάγκες από το προϊόν σε σχέση με άλλους καταναλωτές που βρίσκονται στο τμήμα. Η φωλέα είναι μία μικρή αγορά που δεν εξυπηρετείται από τα ανταγωνιστικά προϊόντα. Οι φωλέες δημιουργούνται με τον διαχωρισμό ενός τμήματος σε μικρότερα τμήματα ή επισημαίνοντας μια ομάδα μέσα στο τμήμα με συγκεκριμένα χαρακτηριστικά, η οποία χρειάζεται κάποιο αγαθό ή υπηρεσία με ιδιαίτερα πλεονεκτήματα [8]. Ένα παράδειγμα φωλέας αποτελεί μια ομάδα τουριστών, η οποία έχει ανάγκη κατά τη διάρκεια των διακοπών της να προσέξει τη διατροφή της λόγω προβλημάτων υγείας. Τα επίπεδα φωλεών μπορούν να αναπτυχθούν και να εξυπηρετηθούν με επιτυχία από μικρές επιχειρήσεις, οι οποίες έχουν αντιληφθεί τις ανάγκες της φωλέας. Για τούτο το λόγο, οι πελάτες μπορεί να πληρώσουν κάτι παραπάνω για την καλύτερη εξυπηρέτηση. Παράλληλα και οι επιχειρηματίες, έχοντας μεγαλύτερη εμπειρία από τους πελάτες και με τη βελτίωση της διαδικασίας μάρκετινγκ που ακολουθούν, επιτυγχάνουν μεγαλύτερη απόδοση και τελικά ανάπτυξη της φωλέας.

- **Τοπικό μάρκετινγκ:** Το μάρκετινγκ μπορεί να παρέχει μια ισχυρή δύναμη στις επιχειρήσεις σε επίπεδο τοπικών περιοχών διότι είναι σχεδιασμένο και προσαρμοσμένο για συγκεκριμένη περιοχή με αποτέλεσμα οι καταναλωτές να ανταποκρίνονται άμεσα [10]. Συνεπώς, χρειάζεται αρχικά να καθοριστούν τα όρια της τοπικής περιοχής. Πρόκειται για τοπική κοινότητα, προάστιο, κράτος ή χώρα; Γνωρίζοντας η επιχείρηση όλα αυτά, μπορεί να βοηθηθεί στο να επιλέξει ποια θα είναι τα μέσα εκείνα, όπως τοπικές εφημερίδες, τοπικό ραδιόφωνο, που θα χρησιμοποιήσει με στόχο την προώθηση του προϊόντος/υπηρεσίας σε τοπικό επίπεδο.
- **Ατομικό μάρκετινγκ:** Το ατομικό μάρκετινγκ αναφέρεται στο μεμονωμένο μάρκετινγκ και παραπέμπει σε μια εποχή όπου η απουσία της τεχνολογίας, άρα και η έλλειψη μηχανημάτων, ανάγκαζε τους καταναλωτές να εξυπηρετούνται ο καθένας ξεχωριστά. Παραδείγματα χάρη, για τη κατασκευή ρούχων χρησιμοποιούταν ο ράφτης, για την κατασκευή υποδημάτων ο υποδηματοποιός κ.α. Έτσι ο καθένας είχε μια εξυπηρέτηση προσαρμοσμένη στις απαιτήσεις του. Η συγκεκριμένη τεχνική εμφανίζεται και σήμερα για διαφορετικούς λόγους και για να καλύψει ξεχωριστές ανάγκες. Η πρόοδος της τεχνολογίας μας βοηθάει στη γρήγορη, άμεση και ευκολότερη εξυπηρέτηση των αναγκών του πελάτη [8].

2.4. Διαδικασία τμηματοποίησης της αγοράς

Τα τμήματα της αγοράς μπορούν να εντοπισθούν από τους υπεύθυνους κάθε επιχείρησης μέσω διάφορων μεταβλητών και μεθόδων. Ωστόσο υπάρχει μία επίσημη διαδικασία για τον εντοπισμό των βασικότερων τμημάτων της αγοράς. Η διαδικασία αυτή αποτελείται από επόμενα τρία κύρια στάδια:

I. Έρευνα

Ο κάθε ερευνητής μάρκετινγκ πληροφορείται σχετικά με τα κίνητρα και τη συμπεριφορά των καταναλωτών, μέσα από συνεντεύξεις και συζητήσεις με τους καταναλωτές ή μέσω ερωτηματολογίων και με βάση αυτά είναι σε θέση να συλλέξει πολύτιμα στοιχεία, όπως για παράδειγμα τον τρόπο χρησιμοποίησης του προϊόντος.

II. Ανάλυση

Εδώ αυτή ο ερευνητής προσπαθεί να αναλύσει τα στοιχεία που έχει συλλέξει από την έρευνα. Μέσα από αυτή την ανάλυση στόχος είναι να εντοπίσει τους παράγοντες που παρουσιάζουν μεγαλύτερη συσχέτιση ανάμεσά τους και να δημιουργήσει έπειτα έναν αριθμό τμημάτων που διαφέρουν αρκετά μεταξύ τους.

III. Δημιουργία προφίλ

Τέλος, η κάθε ομάδα καταναλωτών που χρησιμοποιείται χαρακτηρίζεται και από ένα προφίλ το οποίο δημιουργείται με βάση τα χαρακτηριστικά ή τη συμπεριφορά τους [11]. Για να είναι η διαφορά τους σημαντική και χρήσιμη ως προς τη τμηματοποίηση της αγοράς, θα πρέπει να σχετίζεται με διαφορετικά πρότυπα συμπεριφοράς ή με διαφορετικού συνδυασμούς του μείγματος μάρκετινγκ για την προσέλκυση των καταναλωτών.[12] Επιπροσθέτως χρειάζεται τα επιλεγμένα τμήματα να ξεχωρίζουν και να είναι απομονωμένα από την υπόλοιπη αγορά προκειμένου να διευκολύνουν τις επιχειρήσεις να τα στοχεύουν περισσότερο αποτελεσματικά [13].

2.5. Κριτήρια τμηματοποίησης της αγοράς καταναλωτών

Το εισόδημα, η μόρφωση, η ηλικία, τα ενδιαφέροντα είναι μερικά μόνο από τα κριτήρια με τα οποία κατηγοριοποιείται η αγορά καταναλωτών. Ο αριθμός αυτών των κριτηρίων είναι μεγάλος και αυξάνεται συνεχώς. Οι μεταβλητές λοιπόν που χρησιμοποιούνται για την τμηματοποίηση των καταναλωτών σε διάφορα τμήματα είναι σχετικές με τα χαρακτηριστικά τους (γεωγραφικά, δημογραφικά και ψυχογραφικά) και με τις αντιδράσεις τους απέναντι στο συγκεκριμένο προϊόν [12].

Για να πραγματοποιηθεί η τμηματοποίηση των καταναλωτικών αγορών χρησιμοποιούνται δυο ομάδες μεταβλητών, η πρώτη ομάδα αφορά τα χαρακτηριστικά των καταναλωτών που αναφέραμε πιο πάνω και η δεύτερη ομάδα αφορά στις αντιδράσεις και στη συμπεριφορά των καταναλωτών ως προς τα πλεονεκτήματα που ψάχνουν από την κατανάλωση συγκεκριμένων αγαθών. Οι ομάδες είναι οι εξής:

Γεωγραφική τμηματοποίηση

Η αγοραστική συμπεριφορά ενός καταναλωτή επηρεάζεται από τον τόπο που μένει. Στην ομάδα των γεωγραφικών κριτηρίων περιλαμβάνονται τα εξής κριτήρια:

1. Κλίμα (π.χ. θερμό, ήπιο, ψυχρό).
2. Περιοχή (π.χ. αστική, ημιαστική, αγροτική),
3. Μέγεθος πόλης (π.χ. μεγαλούπολη, πόλη, κωμόπολη)
4. Πυκνότητα πληθυσμού (π.χ. υψηλή, μέτρια, αραιή).
5. Διαμόρφωση εδάφους (π.χ. πεδινό, ημιορεινό, ορεινό,)

Μερικά παραδείγματα θα δείξουν τη χρησιμότητα των κριτηρίων αυτών στην τμηματοποίηση. Οι πωλήσεις συσκευών ψύξης κλιματισμού εξαρτώνται από το κλίμα. Το ίδιο συμβαίνει και με τα είδη ρούχων (π.χ. παλτά, αδιάβροχα). Ανάλογα με την περιοχή που μένει κάποιος επηρεάζονται οι πωλήσεις διαφόρων προϊόντων γιατί σε κάθε περιοχή διαμορφώνονται ειδικές προτιμήσεις.

Για παράδειγμα οι κάτοικοι αστικών περιοχών καταναλώνουν περισσότερα καλλυντικά από τους κατοίκους ημιαστικών ή αγροτικών περιοχών. Επιπλέον, το κεραμίδι ή η πέτρα συνηθίζεται περισσότερο σαν οικοδομικό υλικό σε αγροτικές περιοχές και λιγότερο σε αστικές.

Τα προϊόντα πολυτελείας ή υπερπολυτελείας αγοράζονται από πολύ λίγους, γι' αυτό και βρίσκονται εκεί όπου το μέγεθος της πόλης είναι αρκετό, ώστε να δικαιολογούνται τα έξοδα διανομής και προβολής. Το ίδιο ισχύει και για τις αντιπροσωπείες ΙΧ αυτοκινήτων, αλλά και για τις μεγάλες αλυσίδες σουπερ μάρκετ που τουλάχιστον προς το παρόν δεν τους συμφέρει να ανοίξουν μεγάλης επιφάνειας καταστήματα σε μικρές πόλεις.

Όταν αυξάνεται η πυκνότητα του πληθυσμού (άτομα ανά τετραγωνικό χιλιόμετρο) μιας πόλης, η αξία της γης και της κατοικίας συνήθως μεγαλώνει, με συνέπεια οι κάτοικοι να περιορίζονται σε μικρού εμβαδού κατοικίες. Οι μικρές κατοικίες με τη σειρά τους, δεν χρειάζονται πολλά ή μεγάλα έπιπλα. Άρα, οι κατά κεφαλή πωλήσεις επίπλων επηρεάζονται από την πυκνότητα του πληθυσμού.

Τέλος, και η διαμόρφωση του εδάφους παίζει ρόλο στην κατανάλωση ορισμένων προϊόντων, για παράδειγμα οι κάτοικοι ορεινών περιοχών προτιμούν τρόφιμα με πολλές θερμίδες σε αντίθεση με εκείνους των νησιών.

Τα γεωγραφικά κριτήρια διακρίνονται για την ευκολία εφαρμογής τους στην τμηματοποίηση. Μειονεκτούν ωστόσο στο ότι δίνουν αριστερές πληροφορίες για τους καταναλωτές.

Δημογραφική τμηματοποίηση

Τα δημογραφικά κριτήρια είναι τα πιο συχνά χρησιμοποιούμενα κριτήρια τμηματοποίησης, γιατί είναι εύκολα και αποτελεσματικά. Ένα κριτήριο είναι εύκολο, όταν τα στοιχεία που χρειάζονται είναι δυνατό να βρεθούν γρήγορα και χωρίς κόπο και έξοδα και οι υπολογισμοί που θα γίνουν με βάση τα στοιχεία αυτά δεν είναι δύσκολοι. Στην ομάδα των δημογραφικών κριτηρίων περιλαμβάνονται ειδικότερα τα εξής:

1. Ηλικία
2. Φύλο
3. Εισόδημα
4. Επάγγελμα

5. Επίπεδο μόρφωσης
6. Καταγωγή (ή εθνικότητα, ή φυλή)
7. Θρησκεία
8. Οικογενειακή κατάσταση (έγγαμος, άγαμος, αριθμός παιδιών)
9. Κοινωνική τάξη
10. Τόπος κύριας κατοικίας

Οι πωλήσεις πολλών προϊόντων εξαρτώνται από ένα ή περισσότερα από τα δημογραφικά χαρακτηριστικά των καταναλωτών, για παράδειγμα η ηλικία επηρεάζει τις πωλήσεις ρούχων, παιχνιδιών, αυτοκινήτων, φαρμάκων και τροφίμων.

Το εισόδημα παίζει καθοριστικό ρόλο στη διαμόρφωση των αναγκών των αγοραστών, στον προσδιορισμό της ικανότητας ικανοποίησης των αναγκών και στις φιλοδοξίες των ατόμων. Επιχειρήσεις που παράγουν προϊόντα, όπως κατοικίες, ρούχα, αυτοκίνητα, είδη αθλητισμού και έπιπλα, χρησιμοποιούν το εισόδημα για κριτήριο τμηματοποίησης των αντίστοιχων αγορών. Το επίπεδο της μόρφωσης επηρεάζει τις πωλήσεις των υπηρεσιών αναψυχής και διασκέδασης. Το επάγγελμα επηρεάζει τις πωλήσεις προϊόντων που είναι σχετικά με τα διάφορα επαγγέλματα (π.χ. εργαλεία, είδη γραφείου).

Τέλος, η καταγωγή σαν δημογραφικό κριτήριο τμηματοποίησης θα αναβαθμίζεται συνεχώς εξαιτίας της αθρόας εισόδου και παραμονής στη χώρα μας νόμιμων ή παράνομων μεταναστών από πολλές χώρες. Οι μετανάστες αυτοί, εκτός από την εθνικότητα που έχουν, ανήκουν και σε κάποια θρησκεία που και αυτή, με τον πολλαπλασιασμό των μελών της θα αποκτά σημασία σαν κριτήριο τμηματοποίησης (π.χ. μουσουλμάνοι, ινδουιστές, καθολικοί).

Ψυχογραφική τμηματοποίηση

Η ψυχογραφική τμηματοποίηση πραγματοποιείται στους καταναλωτές κυρίως με βάση τον τρόπο ζωής του ή και τα χαρακτηριστικά τους. Η τμηματοποίηση αυτή παρουσιάζει πολλές φορές ότι καταναλωτές που μπορεί να έχουν ίδια δημογραφικά προφίλ, να διαφέρουν αρκετά ως προς τα ψυχογραφικά χαρακτηριστικά τους [8]. Μια σημαντική ψυχογραφική μεταβλητή είναι ο τρόπος ζωής των καταναλωτών καθότι επηρεάζει σημαντικά τα ενδιαφέροντά τους και τις καταναλωτικές τους συνήθειες. Γι' αυτό τον λόγο οι επιχειρήσεις κατηγοριοποιούν τους καταναλωτές με βάση τον τρόπο ζωής τους.

Σύμφωνα με την τμηματοποίηση με βάση τη συμπεριφορά των καταναλωτών, οι καταναλωτές κατηγοριοποιούνται σύμφωνα με τη θέση που έχουν απέναντι στο προϊόν και την ανταπόκρισή τους προς αυτό. Αρκετοί επιχειρηματίες θεωρούν ότι οι μεταβλητές που σχετίζονται με τη συμπεριφορά των καταναλωτών δημιουργούν τις καλύτερες προϋποθέσεις για τη δημιουργία τμημάτων της αγοράς.

Οι μεταβλητές αυτές είναι:

- I. **Πλεονεκτήματα:** Τα πλεονεκτήματα είναι η ικανοποίηση των αναγκών του ατόμου. Ένα βασικό πλεονέκτημα αυτού του είδους τμηματοποίησης, είναι ότι τα τμήματα που δημιουργούνται με βάση τα οφέλη που απολαμβάνουν οι καταναλωτές στηρίζονται σε περιγραφικούς παράγοντες και αυτό δίνει μεγαλύτερη ευελιξία στον καταμερισμό της αγοράς [12]. Η τμηματοποίηση με βάση τα οφέλη χωρίζει μια ετερογενή αγορά σε ομοιογενή ομάδες σύμφωνα με τα πλεονεκτήματα ενός προϊόντος που πιστεύουν οι καταναλωτές ότι είναι σημαντικά για αυτούς με την κατανάλωση αυτού του προϊόντος. Η συγκεκριμένη διαδικασία τμηματοποίησης, προσφέρει μια περισσότερο άμεση μέτρηση των διαφορών των προτιμήσεων των καταναλωτών και αποτελεί μια πιο προσανατολισμένη δράση ανάλυσης στους υπεύθυνους των επιχειρήσεων [14].
- II. **Περιστάσεις:** Ένας σημαντικός διαχωρισμός των καταναλωτών σε τμήματα είναι αυτός που έχει σχέση με τις περιστάσεις κατά τις οποίες οι καταναλωτές αγοράζουν ένα προϊόν και στις περιστάσεις που δημιουργούν μια ανάγκη τους για να προβούν στην αγορά ενός προϊόντος [8].
- III. **Κοινωνική θέση χρήστη:** Η τμηματοποίηση της αγοράς σύμφωνα με αυτή τη μέθοδο απευθύνεται στις ομάδες καταναλωτών που αποτελούν τωρινούς χρήστες, πρώην χρήστες, χρήστες που δεν το χρησιμοποιούν πλέον, χρήστες που το χρησιμοποιούν για πρώτη φορά, πιθανούς χρήστες μελλοντικά ή και τακτικούς χρήστες. Κάθε συγκεκριμένη ομάδα αντιστοιχεί σε μια διαφορετική στρατηγική μάρκετινγκ που ακολουθούν οι επιχειρήσεις προκειμένου να προσελκύει αυτό το τμήμα. Το είδος κάθε ομάδας, που θα εστιάσει μια εταιρεία, βασίζεται και στην θέση που κατέχει στην αγορά. Παραδείγματος χάρη, μια εταιρεία που κατέχει μεγάλο μερίδιο αγοράς στον κλάδο που δραστηριοποιείται, θα προσπαθήσει να εστιάσει σε νέους χρήστες, ενώ μια λιγότερο ανερχόμενη επιχείρηση θα προσπαθήσει να προσελκύσει χρήστες που ήδη υπάρχουν στον κλάδο και καταναλώνουν παρόμοια προϊόντα με απόσπαση τους από τον ηγέτη της αγοράς.
- IV. **Ρυθμός χρησιμοποίησης:** Σε αυτήν την περίπτωση τμηματοποίησης οι αγορές μπορούν να κατηγοριοποιηθούν σε ομάδες καταναλωτών ελαφριάς, μέτριας χρήσης, έντονης χρήσης του προϊόντος, ανάλογα με το ρυθμό χρήσης του προϊόντος. Οι καταναλωτές που χρησιμοποιούν με έντονο και συχνό ρυθμό ένα προϊόν, αποτελούν ένα όχι ιδιαίτερα μεγάλο ποσοστό της αγοράς, αλλά ένα μεγάλο ποσοστό σύμφωνα με την συνολική κατανάλωση αυτού [8]. Συνεπώς, οι επιχειρήσεις στοχεύουν στο να προσελκύσουν έναν έντονο χρήστη παρά πολλούς ελαφρούς χρήστες.
- V. **Αφοσίωση:** Σύμφωνα με αυτή τη μεταβλητή τμηματοποίησης, η αγορά χωρίζεται σε κατηγορίες ανάλογα με την αφοσίωση του καταναλωτή στη μάρκα του κάθε προϊόντος. Εάν, για παράδειγμα, υπάρχουν πέντε μάρκες ενός τύπου προϊόντος (Α, Β, Γ, Δ και Ε), οι υποψήφιοι καταναλωτές χωρίζονται σε τέσσερις ομάδες:

- **Πάρα πολύ αφοσιωμένοι:** Οι καταναλωτές που ανήκουν σε αυτή την ομάδα, αγοράζουν μια συγκεκριμένη μάρκα πάντοτε.
 - **Αφοσιωμένοι:** Οι καταναλωτές αυτής της κατηγορίας είναι αφοσιωμένοι σε δύο μάρκες που κατά τη γνώμη τους είναι παρόμοιες.
 - **Μεταβαλλόμενη αφοσίωση:** Στην κατηγορία αυτή ανήκουν οι καταναλωτές που στρέφονται από τη μία μάρκα στην άλλη.
 - **Άστατοι:** Εδώ ανήκουν οι καταναλωτές που δεν παρουσιάζουν κάποια αφοσίωση, αντιθέτως είναι αρκετά άστατοι στις καταναλωτικές τους συνήθειες και αγοράζουν σε περιόδους προσφορών ή αναζητούν την ποικιλία μέσα από τις αγορές τους.
- VI. **Φάση ετοιμότητας του αγοραστή:** Η μεταβλητή αυτή αναφέρεται στο ότι οι καταναλωτές διαφέρουν ως προς φάσεις ετοιμότητας προκειμένου να προβούν ή όχι στην αγορά ενός αγαθού. Υπάρχουν οι καταναλωτές οι οποίοι δεν γνωρίζουν καθόλου για την ύπαρξη ενός προϊόντος στην αγορά, υπάρχουν άλλοι οι οποίοι το γνωρίζουν καλά, εκείνοι που είναι μόνο ενημερωμένοι ως προς την ύπαρξη του και άλλοι που ενδιαφέρονται να το αγοράσουν. Οι διαφορές αυτές παίζουν σημαντικό ρόλο στον σχεδιασμό ενός προγράμματος μάρκετινγκ ανάλογα σε ποια ομάδα καταναλωτών απευθύνεται.
- VII. **Στάση:** Σε αυτήν την κατηγορία, οι καταναλωτές μπορούν να χωριστούν σε πέντε ομάδες στάσης απέναντι σε κάποιο προϊόν. Σε τέτοιες ομάδες ανήκουν εκείνοι που αντιδρούν με ενθουσιασμό σε ένα προϊόν, εκείνοι οι οποίοι είναι απλώς θετικοί, όσοι είναι αδιάφοροι, οι αρνητικοί και τέλος οι εχθρικοί.

2.6. Οφέλη τμηματοποίησης

Οι επιχειρήσεις που υιοθετούν την μέθοδο της τμηματοποίησης μπορούν να αποκομίσουν πολλά οφέλη. Ένα από τα θετικά αποτελέσματα αυτής της μεθόδου, είναι η κατανόηση των αναγκών των καταναλωτών, δίνοντας έτσι την ευκαιρία στις επιχειρήσεις να ανταποκριθούν καλύτερα στην προσφορά προϊόντων [15].

Επιπλέον, η τμηματοποίηση της αγοράς, επιτρέπει στις επιχειρήσεις να κατανοήσουν βαθύτερα τα εικάστοτε τμήματα στα οποία μπορεί να θέλει να στοχεύσει, καθώς και τη φύση του ανταγωνισμού που έχει να αντιμετωπίσει σε κάθε περίπτωση. Επιπροσθέτως, η μέθοδος της τμηματοποίησης της αγοράς μπορεί να κάνει τη διαδικασία του σχεδιασμού μάρκετινγκ περισσότερο σαφή [15]. Με την τμηματοποίηση, οι επιχειρήσεις είναι μπορούν να εντοπίσουν και να συγκρίνουν ευκαιρίες στην αγορά καθώς και να δημιουργήσουν καλύτερες ρυθμίσεις μεταξύ των προϊόντων τους και της ζήτησης από τη μεριά των καταναλωτών. Επίσης, μπορούν να αναπτύξουν προϋπολογισμούς έχοντας μια πιο σαφή κατανόηση των χαρακτηριστικών απόκρισης των καταναλωτών στα τμήματα που επιθυμούν να εισέλθουν. Ένα άλλο πλεονέκτημα της τμηματοποίησης της αγοράς, είναι ότι συμβάλλει στο να εντοπισθούν τυχόν κενά στην αγορά, δίνοντας έμφαση στην ανάπτυξη νέων προϊόντων ή επεκτάσεων των ήδη υπαρχόντων.

Επίσης, η τμηματοποίηση της αγοράς, διευκολύνει τις επιχειρήσεις να ταιριάζουν ένα προϊόν ή υπηρεσία πιο αποτελεσματικά με τις ανάγκες των καταναλωτών, και έτσι να δημιουργείται μια ισχυρή ανταγωνιστική θέση στην αγορά. . Λόγο του ότι οι καταναλωτές διαφέρουν ως προς την ευαισθησία τους στις τιμές των προϊόντων που υπάρχουν στην αγορά, με δεδομένη την τμηματοποίηση της αγοράς οι επιχειρήσεις μπορούν να αυξήσουν τις μέσες τιμές των προϊόντων τους άρα και τα κέρδη τους. Η τμηματοποίηση, τέλος, βοηθάει στην ενίσχυση των ευκαιριών για ανάπτυξη των επιχειρήσεων και διατήρηση των πελατών κάτι το οποίο επιτυγχάνεται μέσω της προσφοράς κατάλληλων προϊόντων στα επιλεγμένα τμήματα της αγοράς.

2.7. Η βιομηχανική αγορά

Η βιομηχανική αγορά έχει ένα βασικό χαρακτηριστικό: οι αγορές δεν γίνονται για την άμεση ικανοποίηση των αναγκών των αγοραστών, αλλά για την έμμεση ικανοποίηση των αναγκών τρίτων. Οι αγορές, με άλλα λόγια, είναι εισροές μιας παραγωγικής διαδικασίας που καταλήγει σε τελικά αγαθά, που μπορούν να ικανοποιήσουν ανάγκες. Η παραγωγή θα πρέπει να νοείται έτσι, ώστε να συμπεριλαμβάνει τόσο την τεχνική (μετασχηματισμός υλικών ή και άυλων εισροών), όσο και την οικονομική (πρόσθεση χρησιμότητας) πλευρά της [16].

Στη μεταποιητική αγορά οι αγοραστές είναι βιομηχανίες, βιοτεχνίες, αγροτικές μονάδες, μονάδες παροχής υπηρεσιών (ηλεκτρισμός, τηλεπικοινωνίες, ύδρευση, τράπεζες κ.ά.) κλπ. Αυτοί που αγοράζουν περιλαμβάνουν, μεταξύ άλλων, πρώτες ύλες, ημι-κατεργασμένα υλικά, κεφαλαιουχικά αγαθά, τμήματα προϊόντων, διάφορα εφόδια, υλικά συσκευασίας και διάφορες υπηρεσίες.

Στην κυβερνητική αγορά αγοραστής είναι το κράτος, η τοπική αυτοδιοίκηση και άλλα νομικά πρόσωπα δημοσίου δικαίου. Ο μειοδοτικός διαγωνισμός (με κλειστές ή ανοικτές προσφορές) είναι η διαδικασία αγοράς που κυριαρχεί. Αυτή η διαδικασία περιλαμβάνει τα εξής στάδια:

- Εντοπισμός των αναγκών
- Προγραμματισμός ικανοποίησής τους
- Αναλυτικός καθορισμός των προδιαγραφών (δηλ. των τεχνικών χαρακτηριστικών) των προϊόντων που πρόκειται να αγορασθούν
- Ευρεία δημοσίευση αυτών των προδιαγραφών, διαγωνισμός
- Αποδοχή προσφορών από διάφορους προμηθευτές
- Αξιολόγηση - επιλογή μιας προσφοράς
- Υπογραφή συμβολαίου προμήθειας
- Παραλαβή και έλεγχος των προϊόντων

Στη θεσμική αγορά αγοραστές είναι διάφοροι θεσμοί κοινωνικού αποστολής, όπως νοσοκομεία, γηροκομεία, εκκλησίες, μουσεία, αθλητικοί σύλλογοι, συνδικαλιστικά σωματεία, ιδρύματα, επιτροπές κ.ά. Όλοι αυτοί οι θεσμοί έχουν σαφώς προκαθορισμένους στόχους που προσπαθούν να επιτύχουν απαλλαγμένοι από το κίνητρο του κέρδους.

Τέλος, στη μεταπωλητική αγορά αγοραστές είναι οι ενδιαμέσιοι των διαύλων διανομής και κυρίως οι χονδρέμποροι και λιανέμποροι. Οι ενδιαμέσιοι αυτοί, αγοράζουν τα προϊόντα για να τα μεταπωλήσουν ως έχουν έναντι κέρδους. Οι πολύ λίγες εξαιρέσεις (π.χ. συσκευασία σε μικρότερες μονάδες όγκου ή βάρους) δεν ανατρέπουν τον κανόνα, δηλαδή ότι οι μεσάζοντες δεν επιφέρουν οποιαδήποτε αλλαγή στα τεχνικά χαρακτηριστικά των προϊόντων που μεταπωλούν.

2.8. Κριτήρια τμηματοποίησης της βιομηχανικής αγοράς

Τα πιο συχνά χρησιμοποιούμενα κριτήρια τμηματοποίησης της βιομηχανικής αγοράς είναι τα εξής:

- Η γεωγραφική τοποθεσία του βιομηχανικού αγοραστή,
- Ο κλάδος της δραστηριότητας του βιομηχανικού αγοραστή,
- Το μέγεθος του βιομηχανικού αγοραστή,
- Το μέγεθος παραγγελίας και η συχνότητα αγοράς,
- Η χρήση του προϊόντος.

Η γεωγραφική τοποθεσία του βιομηχανικού αγοραστή

Το πού είναι εγκατεστημένος ένας βιομηχανικός αγοραστής είναι το ευκολότερο κριτήριο τμηματοποίησης. Η γεωγραφική τοποθεσία δεν επηρεάζει μόνο το τι αγοράζει, αλλά και το από πού το αγοράζει. Στην Ελλάδα για παράδειγμα το ξύλο χρησιμοποιείται περισσότερο (σαν οικοδομικό στοιχείο) στην επαρχία παρά στις πόλεις. Επομένως, οι έμποροι οικοδομική υλικών της επαρχίας αναλογικά αγοράζουν περισσότερη ξυλεία από τους συναδέλφους τους στην Αθήνα ή άλλες πόλεις.

Το ίδιο ισχύει για τα κεραμίδια, τα πλαστικά δάπεδα, τους πλαστικούς σωλήνες και πολλά άλλα. Το κόστος μεταφοράς, το κόστος διατήρησης αποθέματος και η ανάγκη για άμεση παράδοση σε έκτακτες περιπτώσεις είναι παράγοντες που υποχρεώνουν τους βιομηχανικούς αγοραστές να προτιμούν να αγοράζουν τα αναγκαία γι' αυτούς προϊόντα από πηγές που είναι κοντά τους. Έτσι, και το κόστος είναι μικρότερο και η λειτουργία των μονάδων τους πιο άνετη. Το γεωγραφικό κριτήριο είναι μεγάλης σημασίας, όταν οι διάφορες οικονομικές δραστηριότητες συγκεντρώνονται σε ορισμένες περιοχές (π.χ. ξενοδοχεία Ρόδου, Κέρκυρας, ναυπηγεία Αττικής, θερμοκήπια Κρήτης, παραγωγή εσπεριδοειδών Κορίνθου, Άρτας).

Ο βιομηχανικός πωλητής τμηματοποιώντας γεωγραφικά μια βιομηχανική αγορά επιλέγει τμήματα σαν αγορές στόχους. Το βασικό επιχείρημα που υποστηρίζεται από το αντίστοιχο μίγμα Μάρκετινγκ είναι η ανταγωνιστική προσφορά του σαν πακέτο. Αυτό ακριβώς θέλει και ο βιομηχανικός αγοραστής, για να ενισχύσει και αυτός με τη σειρά του την ανταγωνιστικότητα της μονάδας του στην εφοδιαστική αλυσίδα. Αυτό είναι ο κανόνας αν και υπάρχουν και εξαιρέσεις. Οι μονάδες στη Χίο που παράγουν γλυκά προμηθεύονται σχεδόν όλες τις υλικές εισροές από πολύ μακριά. Τόσο μακριά (π.χ. κεράσια Τρίπολης, φιαλίδια Αττικής) όσο μακριά στέλνουν στη συνέχεια τα τελικά προϊόντα τους. Προφανώς, όσο οι γραμμές εφοδιασμού γίνονται μακρύτερες, τόσο αυξάνονται και τα προβλήματα της εφοδιαστικής (Logistics).

Εκτός από αυτές τις εξαιρέσεις άλλοι λόγοι που εξασθενούν τη σημαντικότητα αυτού του κριτηρίου τμηματοποίησης είναι η βελτίωση των υποδομών, όπως για παράδειγμα αυτοκινητόδρομοι, ταχεία σιάνη, η ανάπτυξη του διαδικτύου και η χρήση του σαν εναλλακτικού διαύλου πώλησης βιομηχανικών προϊόντων από βιομηχανικούς πωλητές σε βιομηχανικούς αγοραστές και η δημιουργία ενός αυξανόμενου αριθμού μονάδων που παρέχουν εξειδικευμένες υπηρεσίες εφοδιαστικής με ανταγωνιστικούς όρους, παραδείγματος χάρη μεταφορικές, αποθηκευτικές, διανεμητικές. Όλες αυτές οι εξελίξεις υποβαθμίζουν τη σημαντικότητα του παράγοντα απόσταση, δεν την εξαφανίζουν. Γι' αυτό και ο ρόλος της Ελλάδος σαν εφοδιαστικού συνεταιίρου της Βαλκανικής ενδοχώρας.

Ο κλάδος της δραστηριότητας του βιομηχανικού αγοραστή

Ο κλάδος της οικονομικής δραστηριότητας είναι το σημαντικότερο κριτήριο. Κάθε κλάδος έχει και ειδικές ανάγκες για εισροές. Έτσι, μια επιχείρηση, τμηματοποιώντας την αγορά με το κριτήριο αυτό, μπορεί και ικανοποιεί καλύτερα αυτές τις ειδικές ανάγκες. Παραδείγματος χάρη ένας κατασκευαστής μοκετών, μπορεί να τμηματοποιήσει την αγορά στα εξής τμήματα: αυτοκινητοβιομηχανίες, εργολάβοι κατοικιών, οικοδομικές επιχειρήσεις κατασκευής μεγάλων κτιρίων (γραφεία, νοσοκομεία), υπερβολάβοι, τοποθετήσεις μοκετών, χονδρέμποροι, λιανέμποροι, δημόσιο. Μια επιχείρηση παραγωγής κρέατος μπορεί να χωρίσει την αγορά στα εξής τμήματα: νοσοκομεία, ξενοδοχεία, εστιατόρια, ένοπλες δυνάμεις, φοιτητικές εστίες, χονδρέμποροι, λιανέμποροι, βιομηχανίες κατασκευής έτοιμων φαγητών, πλοία.

Το μέγεθος του βιομηχανικού αγοραστή: Η βιομηχανική αγορά περιλαμβάνει επιχειρήσεις που το μέγεθος τους διαφέρει πολύ. Το μέγεθος του βιομηχανικού αγοραστή δεν επηρεάζει μόνο τις ποσότητες των προϊόντων που αγοράζει, αλλά και τον τρόπο με τον οποίο τις αγοράζει. Έτσι, για κάθε μέγεθος χρειάζεται και το ανάλογο μίγμα Μάρκετινγκ. Για παράδειγμα, για μια βιομηχανία μύρας, η αγορά μπορεί να διαιρεθεί στα εξής τμήματα: Αλυσίδες Super Market, αλυσίδες παραδοσιακών παντοπωλείων,

παντοπωλεία, καταστήματα ευκολίας, εστιατόρια, αλυσίδες fast food, καφετέριες, καντίνες, κάβες, χονδρέμποροι, πλοία, τροφοδοσίες.

Το μέγεθος παραγγελίας και η συχνότητα αγοράς: Το μέγεθος του αγοραστή έχει σχέση και με το μέγεθος της παραγγελίας. Οι μεν μικρού μεγέθους συνήθως αγοράζουν μικρές ποσότητες πολύ συχνά, οι δε μεγάλοι μεγάλες ποσότητες αραιά. Η όξυνση του ανταγωνισμού πίεσε τους μεγάλους αγοραστές (π.χ. αλυσίδες σούπερ μάρκετ) να πυκνώσουν τις παραγγελίες μικραίνοντας το μέγεθος τους, ώστε να ρίξουν το κόστος τους. Η τμηματοποίηση της αγοράς με κριτήριο το μέγεθος παραγγελίας αλλά και τη συχνότητα αγοράς, δημιουργεί διαφορετικά τμήματα με διαφορετικές ειδικές ανάγκες εφοδιασμού το καθένα και άρα τις αντίστοιχες ευκαιρίες για εξειδικευμένα μίγματα Μάρκετινγκ βιομηχανικών πωλητών.

Η χρήση του προϊόντος: Οι βιομηχανικοί αγοραστές αγοράζουν τα προϊόντα για διάφορους λόγους. Ανάλογα με τη χρήση του προϊόντος, μπορεί να τμηματοποιηθεί η αγορά, γιατί πίσω από κάθε χρήση υπάρχουν και οι ιδιαίτερες ανάγκες που απαιτούν ένα ειδικό μίγμα Μάρκετινγκ. Για παράδειγμα, το αλεύρι αγοράζεται από τους καταναλωτές, αλλά και από τους άλλους ενδιαμέσους που κάθε ένας από αυτούς το χρησιμοποιεί για διαφορετικό σκοπό. Για μια αλευροβιομηχανία, η αγορά μπορεί να διαιρεθεί στα εξής τμήματα:

- Χονδρέμποροι - λιανέμποροι (θα το μεταπωλήσουν ως έχει),
- αρτοποιεία,
- εργαστήρια ζαχαροπλαστικής,
- ζαχαροπλαστεία,
- βιομηχανίες παραγωγής μπισκότου
- αρτοβιομηχανίες,
- βιομηχανίες κατασκευής φρυγανιών,
- άλλες βιομηχανίες που χρησιμοποιούν αλεύρι (π.χ. έτοιμα φαγητά),
- πιτσαρίες,
- δυνάμεις, δημόσιο, πλοία, τροφοδοσίες.

Τα κριτήρια τμηματοποίησης που αναφέρθηκαν είναι τα πιο σημαντικά. Αυτό δε σημαίνει ότι δεν υπάρχουν άλλα. Με το μέγεθος, αλλά και το δυναμισμό που διακρίνει τη βιομηχανική αγορά, πώς θα ήταν δυνατόν η στρατηγική τμηματοποίησης να στερηθεί των αναλογών όπλων. Αν και σε αριθμό δεν είναι τόσα όσο στα καταναλωτικά, οι ειδικές συνθήκες κάτω από τις οποίες δραστηριοποιούνται τόσοι βιομηχανικοί αγοραστές είναι τόσο διαφοροποιημένες που νομοτελειακά αποκτούν και ειδικά μίγματα ΜΚΤ που να απευθύνονται σε αγορές στόχους που προέκυψαν από κριτήρια εξειδικευμένα.

3. Κριτήρια τμηματοποίησης

Προκειμένου να προχωρήσουμε στην τμηματοποίηση των δεδομένων σε ομάδες, είναι απαραίτητο να ορίσουμε πότε θεωρούμε ότι δύο αντικείμενα είναι όμοια. Στην βιβλιογραφία υπάρχουν δύο κύριες κατηγορίες κριτηρίων που εξυπηρετούν αυτόν το σκοπό: Τα μέτρα απόστασης και τα μέτρα ομοιότητας.

3.1. Μέτρα Απόστασης

Πολλές τεχνικές τμηματοποίησης χρησιμοποιούν την απόσταση δύο αντικειμένων στο χώρο, προκειμένου να αποφανθούν εάν αυτά τα δύο αντικείμενα ανήκουν στην ίδια ομάδα ή όχι. Ένα ορθό μέτρο απόστασης πρέπει να είναι συμμετρικό και να λαμβάνει την τιμή μηδέν, όταν δύο αντικείμενα είναι πανομοιότυπα [17]. Εάν με $d(x_i, x_j)$ συμβολίσουμε την απόσταση των αντικειμένων x_i και x_j ένα μέτρο απόστασης λέγεται και μετρικό μέτρο απόστασης όταν ισχύουν και οι παρακάτω σχέσεις:

$$d(x_i, x_k) \leq d(x_i, x_j) + d(x_j, x_k) \quad \forall x_i, x_j, x_k \in S \quad (1)$$

$$d(x_i, x_j) = 0 \Rightarrow x_i = x_j \quad \forall x_i, x_j \in S \quad (2)$$

Απόσταση Minkowski

Εάν ορίσουμε δύο αντικείμενα σε διανυσματική μορφή p -διαστάσεων, δηλαδή $x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})$ και $x_j = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{jp})$ η απόσταση μεταξύ τους δίνεται από την παρακάτω σχέση:

$$d(x_i, x_j) = \left(|x_{i1} - x_{j1}|^g + |x_{i2} - x_{j2}|^g + \dots + |x_{ip} - x_{jp}|^g \right)^{\frac{1}{g}} \quad (3)$$

Για $g = 2$ η παραπάνω απόσταση ονομάζεται και Ευκλείδεια απόσταση μεταξύ των δύο αντικειμένων. Ένα από τα πλεονεκτήματα αυτού του μέτρου είναι ότι μας δίνει τη δυνατότητα να χρησιμοποιήσουμε την σταθμισμένη απόσταση δύο αντικειμένων. Με άλλα λόγια, σε περίπτωση που οι διάφορες διαστάσεις του διανύσματος έχουν διαφορετικά επίπεδα σημαντικότητας, το συγκεκριμένο μέτρο μας δίνει τη δυνατότητα να τα σταθμίσουμε αναλόγως.

Ειδικές περιπτώσεις

Σε περίπτωση όπου τα διανύσματα παίρνουν μη συνεχείς τιμές, όπως για παράδειγμα όταν παίρνουν δυαδικές, κατηγορικές ή ονομαστικές η παραπάνω μέθοδος πρέπει να τροποποιηθεί κατάλληλα προκειμένου να έχει νόημα η τμηματοποίηση. Συγκεκριμένα στην περίπτωση των δυαδικών τιμών στο διάνυσμα, το μέτρο απόστασης μεταξύ δύο αντικειμένων μπορεί να δοθεί από τον πίνακα συνάφειας (contingency table). Έτσι χρησιμοποιώντας το λεγόμενο simple matching coefficient μπορούμε να εκτιμήσουμε το μέτρο διαφοροποίησης (απόσταση) των δύο αντικειμένων.

$$d(x_i, x_j) = \frac{r + s}{q + r + s + t} \quad (4)$$

όπου q είναι ο αριθμός των χαρακτηριστικών που έχουν την τιμή 1 και στα δύο αντικείμενα, t είναι ο αριθμός των χαρακτηριστικών που είναι 0 και για τα δύο αντικείμενα και s, r είναι ο αριθμός των χαρακτηριστικών που δεν είναι ίδιος και στα δύο αντικείμενα. Ο παραπάνω συντελεστής ισχύει για συμμετρικά χαρακτηριστικά, χαρακτηριστικά δηλαδή για τα οποία η τιμή 1 είναι το ίδιο σημαντική με την τιμή 0. Για μη συμμετρικά χαρακτηριστικά, μπορούμε να αγνοήσουμε το t , δηλαδή τον αριθμό των μη σημαντικών όμοιων αρνητικών τιμών. Σε αυτή τη περίπτωση η σχέση ονομάζεται συντελεστής Jaccard και δίνεται από τον τύπο:

$$d(x_i, x_j) = \frac{r + s}{q + r + s} \quad (5)$$

Ονομαστικά χαρακτηριστικά

Στην περίπτωση όπου τα χαρακτηριστικά παίρνουν ονομαστικές τιμές, δύο κυρίως προσεγγίσεις έχουν προταθεί στη βιβλιογραφία. Η πρώτη είναι το απλό ταίριασμα, δηλαδή εξετάζεται το σε πόσα από τα χαρακτηριστικά τα δύο αντικείμενα έχουν ίδια τιμή. Η απόσταση δίνεται από τον τύπο:

$$d(x_i, x_j) = \frac{p - m}{p} \quad (6)$$

Όπου p είναι ο συνολικός αριθμός χαρακτηριστικών και m ο αριθμός των όμοιων χαρακτηριστικών. Η δεύτερη προσέγγιση είναι να δημιουργηθεί μία δυαδική αναπαράσταση της λύσης και να εφαρμοστεί ο τύπος της προηγούμενης παραγράφου.

Διατάξιμα Χαρακτηριστικά

Στην περίπτωση διατάξιμων χαρακτηριστικών, η αλληλουχία των τιμών είναι σημαντική. Για αυτόν το λόγο, έχει προταθεί η κανονικοποίησή τους στο διάστημα $[0,1]$ προκειμένου να τις αντιστοιχίσουμε με αριθμητικές τιμές τις οποίες μπορούμε να επεξεργαστούμε πιο εύκολα. Η μετατροπή αυτή γίνεται με βάση την παρακάτω σχέση

$$z_{i,n} = \frac{r_{i,n} - 1}{M_n - 1} \quad (7)$$

Όπου $z_{i,n}$ είναι η κανονικοποιημένη τιμή του χαρακτηριστικού a_n του αντικειμένου i , $r_{i,n}$ είναι η τιμή πριν την κανονικοποίηση και M_n είναι το άνω όριο του χαρακτηριστικού a_n με δεδομένο ότι το κάτω όριο είναι η τιμή 1.

Συναρτήσεις Ομοιότητας

Ένα εναλλακτικό κριτήριο τμηματοποίησης, είναι η χρήση της συνάρτησης ομοιότητας $s(x_i, x_j)$, η οποία συγκρίνει τα διανύσματα x_i και x_j . Η συγκεκριμένη συνάρτηση πρέπει να έχει μία μεγάλη τιμή όταν τα δύο διανύσματα είναι όμοια και τη μέγιστη τιμή όταν τα δύο αυτά διανύσματα είναι πανομοιότυπα.

Μέτρο ημιτόνου

Εάν θεωρήσουμε ότι δύο διανύσματα «μοιάζουν» όταν έχουν σχεδόν ίδια γωνία, τότε το κανονικοποιημένο εσωτερικό τους γινόμενο μπορεί να χρησιμοποιηθεί ως μέτρο ομοιότητας:

$$s(x_i, x_j) = \frac{x_i^T \cdot x_j}{\|x_i\| \cdot \|x_j\|} \quad (8)$$

Επίσης παρακάτω παρατίθενται και διάφορα άλλα μέτρα που αναφέρονται στην βιβλιογραφία, χωρίς να γίνει όμως περαιτέρω ανάλυση.

Pearson Correlation Measure

$$s(x_i, x_j) = \frac{(x_i - \bar{x}_i)^T \cdot (x_j - \bar{x}_j)}{\|x_i - \bar{x}_i\| \cdot \|x_j - \bar{x}_j\|} \quad (9)$$

Extended Jaccard Measure

$$s(x_i, x_j) = \frac{x_i^T \cdot x_j}{\|x_i\|^2 + \|x_j\|^2 - x_i^T \cdot x_j} \quad (10)$$

Dice Coefficient Measure

$$s(x_i, x_j) = \frac{2x_i^T \cdot x_j}{\|x_i\|^2 + \|x_j\|^2} \quad (11)$$

3.2. Αξιολόγηση της ποιότητας της τμηματοποίησης

Η ερώτηση εάν και κατά πόσο μία δεδομένη τμηματοποίηση των δεδομένων είναι καλή ή όχι είναι πολύ δύσκολο να απαντηθεί μιας και δεν υπάρχει ένα σαφές αντικειμενικό κριτήριο που να την απαντάει. Έτσι πολλές φορές η αξιολόγηση σχετικά με το εάν μία τμηματοποίηση είναι καλή γίνεται καθαρά εμπειρικά, χρησιμοποιώντας μάλιστα διαφορετικά κριτήρια σε κάθε πρόβλημα [17]. Παρόλα αυτά έχουν προταθεί αρκετά κριτήρια στη βιβλιογραφία, τα οποία χωρίζονται σε δύο κατηγορίες: Τα εσωτερικά και τα εξωτερικά.

Εσωτερικά κριτήρια ποιότητας

Τα εσωτερικά κριτήρια ποιότητας μετράνε συνήθως την συνοχή των τμημάτων χρησιμοποιώντας κάποιο μέτρο ομοιότητας. Συνήθως υπολογίζονται η ομοιογένεια εντός του τμήματος, η διαχωρισιμότητα ανάμεσα στα τμήματα ή ένας συνδυασμός των δύο. Δεν απαιτείται κάποια επιπλέον πληροφορία εκτός των δεδομένων αυτών καθ' αυτών.

Sum of Squared Error (SSE)

Το άθροισμα του τετραγώνου του σφάλματος είναι το πιο απλό και το πιο διαδεδομένο μέτρο ποιότητας όσον αφορά την τμηματοποίηση. Υπολογίζεται ως εξής:

$$SSE = \sum_{k=1}^K \sum_{\forall x_i \in C_k} \|x_i - \mu_k\|^2 \quad (12)$$

όπου C_k είναι το σύνολο των αντικειμένων στο τμήμα k και μ_k είναι ο μέσος των διανυσμάτων που ανήκουν στο τμήμα k . Το διάνυσμα μ_k υπολογίζεται ως εξής:

$$\mu_{k,j} = \frac{1}{N_k} \sum_{\forall x_i \in C_k} x_{i,j} \quad (13)$$

όπου με $N_k = |C_k|$ είναι ο αριθμός των αντικειμένων που ανήκουν στο τμήμα k . Οι μέθοδοι που ελαχιστοποιούν το SSE καλούνται συχνά και μέθοδοι ελάχιστης διακύμανσης, μιας και μπορεί να αποδειχθεί ότι η ελαχιστοποίηση του SSE είναι ισοδύναμη με την ελαχιστοποίηση της διακύμανσης. Το κριτήριο αυτό είναι κατάλληλο για περιπτώσεις στις οποίες τα δεδομένα δημιουργούν συμπαγή «νέφη» τα οποία είναι επαρκώς διαχωρισμένα το ένα με το άλλο [18].

Μέτρα Διασποράς

Τα βαθμωτά κριτήρια διασποράς προκύπτουν από τους πίνακες διασποράς οι οποίοι αφορούν στη διασπορά εντός του τμήματος, μεταξύ των τμημάτων αλλά και το άθροισμα των δύο. Για ένα τμήμα k ο πίνακας διασποράς υπολογίζεται ως εξής:

$$S_k = \sum_{x \in C_k} (x - \mu_k) \cdot (x - \mu_k)^T \quad (14)$$

Ο πίνακας διασποράς που αφορά στην διασπορά εντός του τμήματος υπολογίζεται με βάση τον προηγούμενο τύπο ως εξής:

$$S_W = \sum_{k=1}^K S_k \quad (15)$$

Ενώ ο πίνακας διασποράς μεταξύ των διαφόρων τμημάτων δίνεται:

$$S_B = \sum_{k=1}^K N_k (\mu_k - \mu) \cdot (\mu_k - \mu)^T \quad (16)$$

Όπου μ είναι ο μέσος όλων των δεδομένων και ορίζεται ως εξής:

$$\mu = \frac{1}{m} \sum_{k=1}^K N_k \mu_k \quad (17)$$

Τέλος, ο πίνακας συνολικής διασποράς μπορεί να υπολογιστεί από τον τύπο:

$$S_T = \sum_{x \in C_1, C_2, \dots, C_K} (x - \mu) \cdot (x - \mu)^T \quad (18)$$

Με βάση λοιπόν τους τρεις πίνακες S_W, S_B, S_T μπορούμε να ορίσουμε τρία διαφορετικά βαθμωτά κριτήρια.

- Το κριτήριο του ίχνους. Η ελαχιστοποίηση του αθροίσματος των διαγώνιων στοιχείων του πίνακα S_W είναι ισοδύναμη με την ελαχιστοποίηση του SSE και κατά συνέπεια το κριτήριο είναι κατάλληλο. Υπολογίζεται ως εξής:

$$J_e = \text{tr}[S_W] = \sum_{k=1}^K \sum_{x \in C_k} \|x - \mu_k\|^2 \quad (19)$$

Με όμοιο τρόπο, ένα άλλο κριτήριο με βάση το ίχνος, είναι η μεγιστοποίησή του στον πίνακα S_B , η μεγιστοποίηση δηλαδή της ποσότητας:

$$tr[S_B] = \sum_{k=1}^K N_k \|\mu_k - \mu\|^2 \quad (20)$$

- Το κριτήριο της οριζουσας. Η οριζουσα ενός πίνακα διασποράς μετράει κατά προσέγγιση το τετράγωνο του όγκου διασποράς.
- Το κριτήριο της ανεξαρτησίας. Αν $\lambda_1, \lambda_2, \dots, \lambda_d$ είναι οι ιδιοτιμές του πίνακα $S_W^{-1}S_B$ λέμε ότι έχουμε «καλό» διαχωρισμό, όταν οι μη μηδενικές ιδιοτιμές είναι επαρκώς μεγάλες. Κατά συνέπεια μπορούμε να ορίσουμε αρκετά κριτήρια με βάση τις ιδιοτιμές, όπως για παράδειγμα τα παρακάτω:

$$1. \quad tr[S_W^{-1}S_B] = \sum_{i=1}^d \lambda_i \quad (21)$$

$$2. \quad J_f = tr[S_T^{-1}S_W] = \sum_{i=1}^d \frac{1}{1 + \lambda_i} \quad (22)$$

$$3. \quad \frac{|S_W|}{|S_T|} = \prod_{i=1}^d \frac{1}{1 + \lambda_i} \quad (23)$$

Εξωτερικά Κριτήρια Ποιότητας

Τα εξωτερικά κριτήρια ποιότητας χρησιμοποιούνται συνήθως όταν έχουμε ήδη μία τμηματοποίηση των δεδομένων και θέλουμε να δούμε το κατά πόσο η δομή των τμημάτων ταιριάζει με την δεδομένη. Δεν θα γίνει περαιτέρω ανάλυση όμως στην παρούσα εργασία.

4. Μέθοδοι τμηματοποίησης

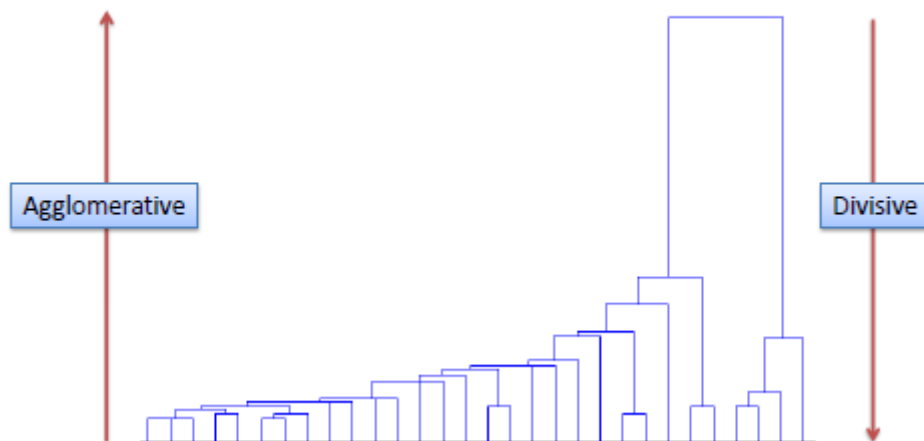
Σε αυτό το κομμάτι της εργασίας, θα αναφερθούμε στους πιο διαδεδομένους αλγορίθμους τμηματοποίησης. Όπως έχει ήδη αναφερθεί, δεν υπάρχει σαφής ορισμός της έννοιας του «τμήματος» και κατά συνέπεια αναλόγως της εκάστοτε ερμηνείας του όρου, προκύπτει και ένας αντίστοιχος αλγόριθμος. Για αυτόν ακριβώς το λόγο, υπάρχει ένα πλήθος αλγορίθμων στη βιβλιογραφία τους οποίους θα προσπαθήσουμε να κατατάξουμε σε κατηγορίες. Οι Fraley και Raftery [19] πρότειναν την κατάταξη των αλγορίθμων σε δύο κύριες κατηγορίες: τους ιεραρχικούς και τους διαχωριστικούς (partitioning). Μερικά χρόνια αργότερα, οι Han and Kamber [20] πρότειναν τον περαιτέρω διαχωρισμό των αλγορίθμων σε τρεις πρόσθετες κατηγορίες: μέθοδοι με βάση την πυκνότητα, μέθοδοι με βάση το μοντέλο και μέθοδοι με βάση το δίκτυο.

4.1. Ιεραρχικές Μέθοδοι

Το κύριο γνώρισμα των συγκεκριμένων μεθόδων είναι η κατασκευή των τμημάτων με αναδρομικό διαχωρισμό των δεδομένων είτε από επάνω προς τα κάτω είτε το ανάποδο. Οι μέθοδοι αυτοί μπορούν να διαχωριστούν περεταίρω στις ακόλουθες κατηγορίες:

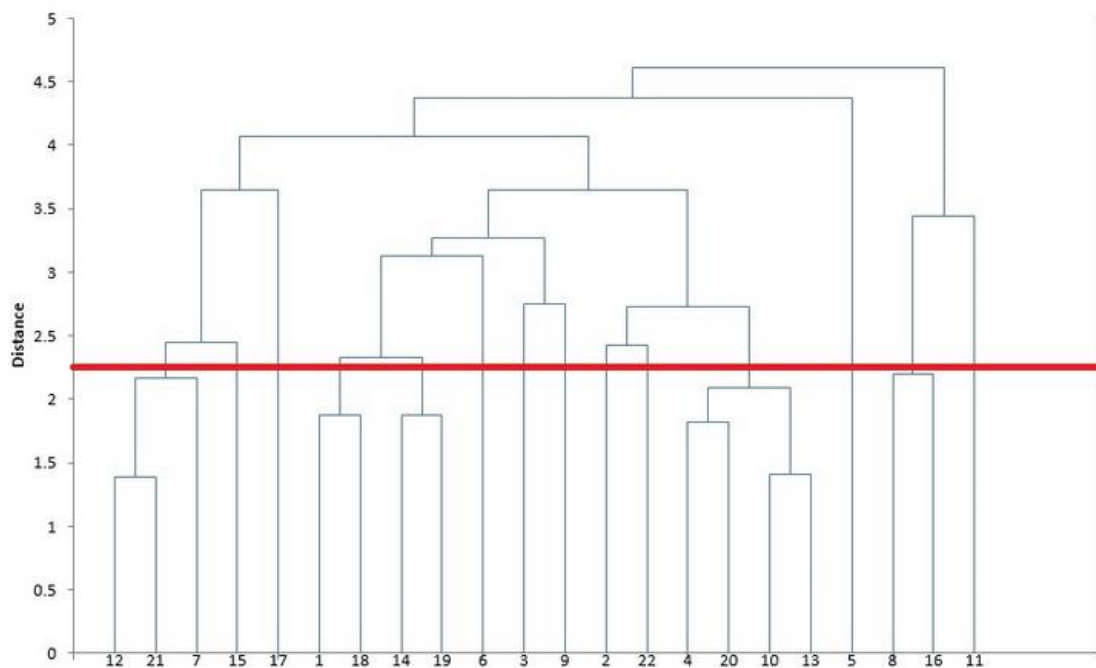
- **Agglomerative ιεραρχική τμηματοποίηση:** Κάθε αντικείμενο αρχικά αντιπροσωπεύει ένα τμήμα. Στη συνέχεια καθώς ο αλγόριθμος εκτελείται, τα τμήματα αυτά συγχωνεύονται με βάση κάποιους κανόνες μέχρις ότου επιτευχθεί η επιθυμητή δομή των τμημάτων.
- **Divisive ιεραρχική τμηματοποίηση:** Όλα τα αντικείμενα ανήκουν αρχικά σε ένα μόνο τμήμα. Κατά την εκτέλεση του αλγορίθμου, τα τμήματα διαιρούνται σε υποτμήματα έως ότου επιτευχθεί η επιθυμητή δομή των τμημάτων.

Hierarchical Clustering



Εικόνα 1 – Agglomerative και divisive τμηματοποίηση

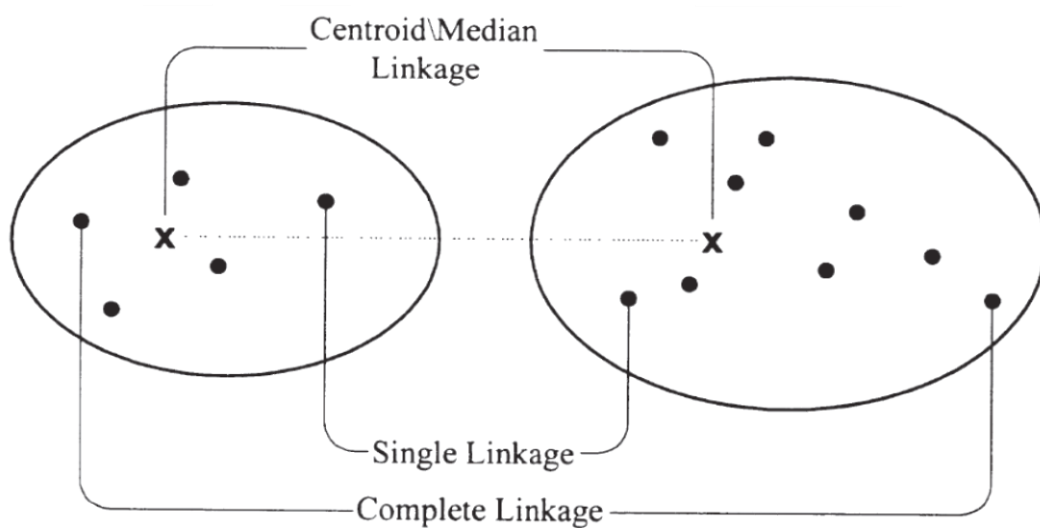
Dendrogram



Εικόνα 2 – Δενδρόγραμμα για μία τμηματοποίηση. Η κόκκινη γραμμή αντιστοιχεί στο ανεκτό επίπεδο αποστάσεων.

Το αποτέλεσμα και των δύο τρόπων είναι ένα δενδρόγραμμα, το οποίο αναλόγως του σημείου που θα κοπεί, προκύπτουν διαφορετικά επίπεδα ομοιότητας. Όσον αφορά τώρα τον τρόπο μέτρησης του επιπέδου ομοιότητας υπάρχουν οι εξής διαφορετικοί τρόποι:

- Τμηματοποίηση μονής ακμής: Η μέθοδος αυτή συναντάται συχνά και ως «connectedness», η μέθοδος του ελαχίστου ή η μέθοδος του πλησιέστερου γείτονα. Ως απόσταση μεταξύ των τμημάτων λαμβάνεται η ελάχιστη απόσταση μεταξύ οποιουδήποτε σημείου του πρώτου τμήματος με οποιοδήποτε σημείο του δεύτερου τμήματος. Εάν όμως δεν χρησιμοποιείται μέτρο απόστασης, αλλά μέτρο ομοιότητας τότε η ομοιότητα μεταξύ δύο τμημάτων είναι ίση με τη μέγιστη ομοιότητα οποιουδήποτε σημείου του πρώτου τμήματος με οποιοδήποτε σημείο του δεύτερου τμήματος [21].
- Τμηματοποίηση πλήρους ακμής: Η μέθοδος αυτή συναντάται συχνά και ως μέθοδος διαμέτρου, ως η μέθοδος του μεγίστου ή μέθοδος του πιο μακρινού γείτονα. Όμοια με πριν, σε αυτή τη μέθοδο ως απόσταση μεταξύ των τμημάτων λαμβάνεται η μέγιστη απόσταση μεταξύ οποιουδήποτε σημείου του πρώτου τμήματος με οποιοδήποτε σημείο του δεύτερου τμήματος [22].
- Τμηματοποίηση μέσου όρου ακμής: Η μέθοδος αυτή καλείται και μέθοδος ελάχιστης διακύμανσης. Σε αυτή τη μέθοδο ως απόσταση μεταξύ των τμημάτων λαμβάνεται ο μέσος όρος της απόστασης μεταξύ οποιουδήποτε σημείου του πρώτου τμήματος με οποιοδήποτε σημείο του δεύτερου τμήματος [23].



Εικόνα 3 – Γραφική αναπαράσταση των μεθόδων μονής, πλήρους και μέσου όρου ακμής.

Η τμηματοποίηση μονής ακμής έχει ένα μειονέκτημα γνωστό και ως «φαινόμενο αλυσίδας». Εάν υπάρχουν έστω και ελάχιστα σημεία που δημιουργούν μια «γέφυρα» μεταξύ των τμημάτων, τότε η απόστασή τους είναι πολύ μικρή και κατά συνέπεια τα δύο τμήματα συγχωνεύονται σε ένα [24].

Η τμηματοποίηση μέσου όρου έχει και αυτή ένα μειονέκτημα. Μπορεί να οδηγήσει στην κατάρτιση ενός τμήματος εάν αυτό έχει σχήμα μακρόστενο και εξαιτίας αυτού να οδηγήσει στην συγχώνευση των νέων τμημάτων με άλλα γειτονικά.

Η τμηματοποίηση πλήρους ακμής συνήθως οδηγεί σε πιο συμπαγή τμήματα αλλά οι άλλες μέθοδοι είναι πιο εύκολα προσαρμόσιμες.

Συνοπτικά θα παραθέσουμε τα πλεονεκτήματα και τα μειονεκτήματα των ιεραρχικών μεθόδων. Τα πλεονεκτήματα είναι τα εξής:

- Προσαρμοστικότητα: Οι μέθοδοι μονής ακμής συνεχίζουν να δίνουν καλά αποτελέσματα όταν τα δεδομένα απαρτίζονται από τμήματα που είναι σαφώς διαχωρισμένα, συμπεριλαμβανομένου τμημάτων σε σχήμα αλυσίδας ή ομόκεντρα.
- Πολλαπλά επίπεδα διαχωρισμού: Οι μέθοδοι αυτές δημιουργούν πολλά partitions (κατατμήσεις) και με αυτόν τον τρόπο δίνεται στο χρήστη η ευχέρεια να επιλέξει πιο επίπεδο διαχωρισμού είναι το καταλληλότερο για την εκάστοτε εφαρμογή.

Τα κυριότερα μειονεκτήματα των μεθόδων είναι τα εξής:

- Δυσκολία στην εφαρμογή τους σε μεγάλα δεδομένα. Η πολυπλοκότητα όσον αφορά το χρόνο είναι τουλάχιστον $O(m^2)$ όπου m είναι ο συνολικός αριθμός δεδομένων, η οποία είναι μη γραμμική (τετραγωνική) ως προς το χρόνο. Κατά συνέπεια για μεγάλο m ο χρόνος εκτέλεσης αυξάνεται δυσανάλογα. Εκτός αυτού απαιτείται και μεγάλη ποσότητα μνήμης προκειμένου να εκτελεστεί σωστά ο αλγόριθμος.
- Οι ιεραρχικές μέθοδοι δεν μπορούν να αναιρέσουν κάποια ενέργεια που έγινε σε προηγούμενο στάδιο του αλγορίθμου. Δεν υπάρχει δηλαδή τρόπος να επιστρέψει σε μία προηγούμενη κατάσταση, εκτός και εάν αποθηκεύουμε το αποτέλεσμα μετά από κάθε επανάληψη του αλγορίθμου.

4.2. Διαχωριστικές (Partitioning) Μέθοδοι

Οι συγκεκριμένες μέθοδοι λειτουργούν με το να επανατοποθετούν τα δεδομένα από το ένα τμήμα στο άλλο, ξεκινώντας από έναν αρχικό διαχωρισμό. Είναι λοιπόν απαραίτητο να ορίσει αρχικά ο χρήστης τον αριθμό των τμημάτων. Η εύρεση του ολικού βέλτιστου, ανεξάρτητα από το ποιο είναι το μέτρο ποιότητας, απαιτεί την εξαντλητική αναζήτηση όλων των πιθανών partitions. Αν και αυτό μπορεί να είναι εφικτό για

μικρά δεδομένα, είναι πρακτικά αδύνατο εάν ο αριθμός των δεδομένων είναι μεγάλος. Για αυτόν τον λόγο χρησιμοποιούνται ευριστικές τεχνικές βελτιστοποίησης μερικές από τις οποίες παρατίθενται παρακάτω.

4.3. Αλγόριθμοι που ελαχιστοποιούν κάποιο σφάλμα

Το συγκεκριμένο είδος αλγορίθμων λειτουργούν αρκετά αποτελεσματικά όταν τα τμήματα είναι συμπαγή και απομονωμένα. Η βασική τους ιδέα είναι να ελαχιστοποιήσουν κάποιο σφάλμα, το οποίο συνήθως έχει να κάνει με την απόσταση των δεδομένων από το κέντρο του τμήματος. Το πιο γνωστό κριτήριο τέτοιου είδους είναι το άθροισμα του τετραγωνικού σφάλματος (Sum of Squared Error – SSE), το οποίο μετράει την ευκλείδεια απόσταση.

Ο πιο απλός και ευρέως χρησιμοποιούμενος αλγόριθμος αυτής της κατηγορίας είναι ο αλγόριθμος K-means. Για να εκτελεστεί, πρέπει ο χρήστης να ορίσει τον αριθμό των τμημάτων K . Το κέντρο του κάθε τμήματος ορίζεται ως ο μέσος όρος των αντικειμένων που περιέχει. Ο αλγόριθμος ξεκινάει επιλέγοντας τυχαία ως κέντρα των K τμημάτων κάποια από τα αντικείμενα που υπάρχουν στο σύνολο δεδομένων. Στη συνέχεια υπολογίζει τις αποστάσεις κάθε άλλου αντικειμένου από τα κέντρα αυτά, και το τοποθετεί στο πλησιέστερό του. Κατόπιν, τα κέντρα των τμημάτων επανυπολογίζονται ως οι μέσοι όροι των αντικειμένων που περιέχουν με βάση την παρακάτω σχέση:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (24)$$

Η διαδικασία αυτή επαναλαμβάνεται έως ότου να μην υπάρχει αλλαγή στο τμήμα που ανήκει το κάθε αντικείμενο ή έως ότου να μην υπάρχει αλλαγή στο συνολικό σφάλμα ή όταν φθάσει στον μέγιστο αριθμό επαναλήψεων. Ένας απλός ψευδοκώδικας του αλγορίθμου παρατίθεται παρακάτω.

Αρχή
Αρχικοποίηση των K τμημάτων
Όσο (συνθήκη τερματισμού = ΨΕΥΔΗΣ)
Τοποθέτηση των αντικειμένων στο πλησιέστερο τμήμα
Επανυπολογισμός των κέντρων των τμημάτων
Τέλος
Τέλος

Πολλοί ερευνητές θεωρούν τον K-means σαν έναν αλγόριθμο gradient descent, ο οποίος ξεκινάει από ένα αρχικό σύνολο τμημάτων (τα οποία αντιπροσωπεύονται από το κέντρο τους) και με επαναληπτικό τρόπο το επανυπολογίζει με σκοπό να ελαττώσει το σφάλμα. Η πολυπλοκότητα του αλγορίθμου είναι γραμμική με την ποσότητα των αντικειμένων και αυτό είναι ένας από τους κυριότερους λόγους για την ευρεία χρήση του. Εκτός αυτού όμως, ο αλγόριθμος είναι εξαιρετικά εύκολο να προγραμματιστεί, συγκλίνει σε σχετικά μικρό αριθμό επαναλήψεων και μπορεί εύκολα να προσαρμοστεί σε αραιά δεδομένα [25].

Το κυριότερο του μειονέκτημα, είναι ότι η ποιότητα της τελικής τμηματοποίησης εξαρτάται πολύ από την αρχική επιλογή των κέντρων. Αυτό είναι και ο λόγος που μπορεί να οδηγήσει τον αλγόριθμο να συγκλίνει σε ένα τοπικό βέλτιστο και όχι στο ολικό. Άλλα μειονεκτήματα είναι η ευαισθησία του αλγορίθμου όσον αφορά τον θόρυβο στα δεδομένα και στις ακραίες τιμές. Για παράδειγμα εάν όλες οι τιμές είναι συγκεντρωμένες γύρω από το κέντρο του τμήματος, αλλά μία ή δύο τιμές έχουν τιμές εξαιρετικά μακριά από αυτό θα οδηγήσουν τον αλγόριθμο στο να μετακινήσει το κέντρο του τμήματος σε λάθος περιοχή. Για αυτό το λόγο λειτουργεί καλά όταν τα δεδομένα χωρίζονται σε όσο το δυνατόν πιο ισοτροπικά τμήματα. Επίσης, ο αλγόριθμος λόγω των υπολογισμών που απαιτούνται, είναι αναγκαίο τα δεδομένα να είναι καθαρά αριθμητικά. Διαφορετικά δεν είναι δυνατός ο υπολογισμός των αποστάσεων και των νέων κέντρων που προκύπτουν μετά από κάθε επανάληψη του αλγορίθμου. Τέλος ο αλγόριθμος απαιτεί τον εκ των προτέρων ορισμό του αριθμού των τμημάτων, πράγμα που μπορεί να αποδειχθεί δύσκολο σε περίπτωση που δεν υπάρχει αρκετή γνώση, εμπειρία ή δεδομένα.

4.4. Μέθοδοι που βασίζονται σε θεωρία γράφων

Οι συγκεκριμένες μέθοδοι επιδιώκουν το σχηματισμό τμημάτων με βάση θεωρία γράφων. Οι ακμές των γράφων ενώνουν δύο αντικείμενα, τα οποία αποτελούν τους κόμβους του γράφου. Ένας από τους πιο γνωστούς αλγορίθμους αυτής της κατηγορίας είναι ο αλγόριθμος Minimal Spanning Tree [26]. Οι inconsistent ακμές είναι αυτές οι οποίες έχουν πολύ μεγαλύτερο μήκος από ότι οι γειτονικές τους.

4.5. Μέθοδοι που βασίζονται στην πυκνότητα

Η βασική ιδέα των συγκεκριμένων μεθόδων είναι ότι τα σημεία που ανήκουν σε κάθε τμήμα, προκύπτουν από μία συγκεκριμένη συνάρτηση πυκνότητας πιθανότητας [27]. Η συνολική κατανομή των δεδομένων

θεωρείται ότι ακολουθεί ένα μίγμα κατανομών. Ο στόχος της συγκεκριμένης μεθόδου είναι να αναγνωριστούν τα τμήματα καθώς και η κατανομή που ακολουθούν, ακόμη και εάν τα τμήματα έχουν ακαθόριστο σχήμα στο χώρο (χωρίς μάλιστα να είναι απαραίτητα κυρτά). Στην βιβλιογραφία συναντάει κανείς εκτενείς ερευνητικές προσεγγίσεις που βασίζονται στην υπόθεση ότι οι συναρτήσεις πιθανότητας ακολουθούν πολυμεταβλητές Γκαουσιανές κατανομές.

Σε αυτές τις περιπτώσεις ένας αποδοκός τρόπος λύσης του προβλήματος είναι να χρησιμοποιηθεί η αρχή της μέγιστης πιθανότητας. Σύμφωνα με αυτήν, επιλέγεται μία συγκεκριμένη δομή των τμημάτων και αντίστοιχες παράμετροι τέτοιες ώστε να μεγιστοποιείται η πιθανότητα τα δεδομένα να προέρχονται από μία τέτοια κατανομή. Ο αλγόριθμος μεγιστοποίησης των προσδοκιών (Expectation Maximization –EM) είναι ένας αλγόριθμος μέγιστης πιθανότητας γενικού σκοπού και έχει εφαρμοστεί στο πρόβλημα της εκτίμησης των παραμέτρων. Ο αλγόριθμος DBSCAN (density-based spatial clustering of applications with noise) είναι ικανός να «ανακαλύπτει» τμήματα με ακανόνιστο σχήμα στο χώρο και είναι αρκετά αποδοτικός για μεγάλα σετ δεδομένων. Η βασική ιδέα είναι να ελέγχει τη γειτονική περιοχή του κάθε αντικειμένου και να εξασφαλίζει ότι περιέχει τουλάχιστον τον ελάχιστον αριθμό αντικειμένων. Ο αλγόριθμος AUTOCLASS είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος ο οποίος καλύπτει μεγάλο εύρος κατανομών όπως Gauss, Bernoulli, Poisson και άλλες. Επίσης γνωστές μέθοδοι είναι οι SNOB και MCLUST.

5. Προσδιορισμός του αριθμού των τμημάτων

Όπως έχει ήδη αναφερθεί σε προηγούμενα κεφάλαια, πολλοί από τους αλγόριθμους απαιτούν τον ορισμό των τμημάτων εκ των προτέρων. Είναι επίσης ευρέως γνωστό ότι η σωστή επιλογή του αριθμού αυτού μπορεί να παίζει πολύ σημαντικό ρόλο και να επηρεάσει σε μεγάλο βαθμό την ποιότητα της τμηματοποίησης. Επίσης, πέρα από το μαθηματικό και αλγοριθμικό κομμάτι η γνώση του αριθμού των τμημάτων μπορεί να προσφέρει στην εκάστοτε εταιρία μία πολύτιμη πληροφορία η οποία θα την βοηθήσει στην λήψη αποφάσεων. Για αυτούς τους λόγους λοιπόν, δημιουργείται εύκολα η απορία πώς μπορούμε να εκτιμήσουμε τον αριθμό των τμημάτων όταν η γνώση και η εμπειρία ως προς τα δεδομένα μας είναι περιορισμένη.

Σε αυτό το σημείο αξίζει να σημειωθεί πώς τα περισσότερα από τα κριτήρια που χρησιμοποιούνται για την εκτίμηση της ποιότητας της τμηματοποίησης, όπως για παράδειγμα το μέσο τετραγωνικό σφάλμα είναι φθίνουσες συναρτήσεις όσον αφορά τον αριθμό τμημάτων K . Όσο δηλαδή αυξάνονται τα τμήματα, τόσο μειώνεται το σφάλμα. Έτσι η «βέλτιστη» λύση θα ήταν να έχουμε όσα τμήματα όσα και τα δεδομένα

μας. Είναι προφανές όμως ότι σε αυτήν την περίπτωση παρόλο που το σφάλμα είναι μηδέν, η εταιρία ή ο αποφασίζων δεν μπορεί να εξάγει κανένα απολύτως χρήσιμο συμπέρασμα. Κατά συνέπεια είναι επιτακτική η ανάγκη χρήσης διαφορετικών κριτηρίων. Υπάρχουν αρκετές προτάσεις στη βιβλιογραφία οι οποίες μπορούν να εκτιμήσουν το K . Είναι κατά το πλείστον ευριστικές μέθοδοι οι οποίες υπολογίζουν τα μέτρα ποιότητας για τις διάφορες τιμές του K , έτσι ώστε να βρεθεί το βέλτιστο.

5.1. Μέθοδοι που βασίζονται στην διασπορά εντός των τμημάτων

Πολλές από τις μεθόδους που έχουν προταθεί στη βιβλιογραφία, βασίζονται στη διασπορά των αντικειμένων εντός του ίδιου τμήματος. Ο αλγόριθμος within cluster depression decay [28] ανήκει σε αυτήν την κατηγορία και βασίζεται στον υπολογισμό του σφάλματος W_K :

$$W_K = \sum_{k=1}^K \frac{1}{2N_k} D_k \quad (25)$$

Όπου το D_k είναι το άθροισμα των αποστάσεων ανά δύο όλων των αντικειμένων εντός του τμήματος k .

$$D_k = \sum_{x_i, x_j \in C_k} \|x_i - x_j\| \quad (26)$$

Όσο ο αριθμός των τμημάτων αυξάνεται, το within cluster depression decay μειώνεται απότομα μέχρι ένα σημείο όπου η καμπύλη αρχίζει και γίνεται επίπεδη. Δηλαδή περαιτέρω αύξηση του αριθμού των τμημάτων δίνει πολύ μικρή μείωση της εν λόγω ποσότητας. Σύμφωνα με τους συγγραφείς αυτό το σημείο είναι ο προτιμώμενος αριθμός τμημάτων.

Μία άλλη μέθοδος που ανήκει σε αυτήν την κατηγορία είναι η μέθοδος αναλογικής μείωσης σφάλματος (Proportional Reduction in Error – PRE). Η συγκεκριμένη μέθοδος έχει αποδειχθεί ότι λειτουργεί αποδοτικά στα περισσότερα σετ δεδομένων. Η μέθοδος συγκρίνει το κατά πόσο μειώνεται το συνολικό τετραγωνικό σφάλμα εάν αυξήσουμε τα τμήματα από K σε $K + 1$. Ο αλγόριθμος θεωρεί ότι αύξηση του αριθμού των τμημάτων είναι αποδεκτή μόνο εάν η αντίστοιχη μείωση του τετραγωνικού σφάλματος είναι μεγαλύτερη από ένα κατώφλι (συνήθως 0.4 ή μεγαλύτερο).

5.2. Μέθοδοι που βασίζονται στην διασπορά εντός και μεταξύ των τμημάτων

Οι προαναφερθείσες μέθοδοι εάν και βασίζονται σε λογικές υποθέσεις, πάσχουν από ένα πολύ σημαντικό παράγοντα: το ότι δεν λαμβάνουν καθόλου τι γίνεται μεταξύ των τμημάτων παρὰ μόνο εντός των τμημάτων. Με βάση λοιπόν αυτήν την σκέψη ως κίνητρο, πολλοί ερευνητές κατέληξαν ότι εκτός από την ομοιογένεια εντός του ίδιου τμήματος απαιτείται ταυτόχρονα και όσο το δυνατόν μεγαλύτερη ετερογένεια μεταξύ των διαφόρων τμημάτων. Οι Ray και Turi [29] για παράδειγμα, χρησιμοποίησαν ένα μέτρο που ορίζεται ως ο λόγος μεταξύ της διασποράς εντός του τμήματος, με τη διασπορά μεταξύ των τμημάτων. Η ελαχιστοποίηση αυτού του μέτρου είναι ισοδύναμη με την ελαχιστοποίηση της διασποράς εντός του τμήματος και με την μεγιστοποίηση της διασποράς μεταξύ των τμημάτων.

Μία δεύτερη μέθοδος αυτής της κατηγορίας είναι η μέθοδος του δείκτη εγκυρότητας (validity index) η οποία προτάθηκε από τους Kim et al. [30]. Δύο μέτρα είναι κατάλληλα για αυτή τη χρήση:

- Η μέση απόσταση εντός του τμήματος (mean intra-cluster distance) – MICD η οποία ορίζεται για το k τμήμα ως εξής:

$$MD_k = \sum_{x_i \in C_k} \frac{\|x_i - \mu_k\|}{N_k} \quad (27)$$

- Η ελάχιστη απόσταση μεταξύ των τμημάτων που ορίζεται:

$$d_{min} = \min_{i \neq j} \|\mu_i - \mu_j\| \quad (28)$$

Προκειμένου να οριστεί ο δείκτης εγκυρότητας, πρέπει να μελετηθεί η συμπεριφορά των δύο αυτών μεγεθών γύρω από τον πραγματικό αριθμό τμημάτων, έστω K^* . Όταν τα δεδομένα έχουν χωριστεί σε λιγότερα τμήματα από το K^* , δηλαδή $K < K^*$ τότε τουλάχιστον ένα τμήμα θα έχει μεγάλη τιμή MD . Καθώς όμως ο αριθμός των τμημάτων αυξάνεται παραπάνω από το K^* , τότε η τιμή αυτή μειώνεται απότομα. Όσον αφορά τώρα το δεύτερο μέγεθος, είναι μεγάλο όταν τα δεδομένα είναι χωρισμένα σε $K \leq K^*$ τμήματα. Γίνεται απότομα μικρό όταν το K γίνει μεγαλύτερο από το K^* καθώς τότε τουλάχιστον ένα από τα συμπαγή τμήματα διασπάται.

Προκειμένου να βρούμε τις καταστάσεις με λιγότερα τμήματα από το βέλτιστο αλλά και τις καταστάσεις με περισσότερα τμήματα από το βέλτιστο, έχουν οριστεί δύο επιπρόσθετες συναρτήσεις. Αυτές οι

συναρτήσεις εξαρτώνται και από τα κέντρα των τμημάτων τα οποία να θυμίσουμε ότι ορίζονται ως εξής $\mu = [\mu_1, \mu_2, \dots, \mu_K]^T$.

Η πρώτη λοιπόν συνάρτηση ορίζεται από τον εξής τύπο:

$$v_u(K, \mu; X) = \frac{\sum_{k=1}^K MD_k}{K} \quad 2 \leq K \leq K_{max} \quad (29)$$

Η συνάρτηση αυτή λαμβάνει πολύ μικρές τιμές όταν $K \geq K^*$ και σχετικά μεγάλες τιμές όταν $K < K^*$. Κατά αυτόν τον τρόπο μας εξυπηρετεί στο να αποφανθούμε εάν τα δεδομένα είναι χωρισμένα σε λιγότερα τμήματα από το βέλτιστο.

Η δεύτερη συνάρτηση ορίζεται από τον τύπο:

$$v_o(K, \mu) = \frac{K}{d_{min}} \quad 2 \leq K \leq K_{max} \quad (30)$$

Η συνάρτηση αυτή παίρνει μεγάλες τιμές για $K \geq K^*$ ενώ σχετικά μικρές τιμές για $K < K^*$. Έτσι καταλαβαίνουμε εάν τα δεδομένα είναι χωρισμένα σε μεγαλύτερα τμήματα από το βέλτιστο.

Ο δείκτης εγκυρότητας βασίζεται στο γεγονός ότι και οι δύο συναρτήσεις λαμβάνουν μικρές τιμές για $K = K^*$. Ορίζουμε δύο διανύσματα, που περιέχουν τις τιμές των παραπάνω συναρτήσεων για όλες τις τιμές του $K \in (0, K_{max})$. Δηλαδή:

$$V_u = [v_u(2, \mu; X), \dots, v_u(K_{max}, \mu; X)]$$

$$V_o = [v_o(2, \mu), \dots, v_o(K_{max}, \mu)]$$

Τέλος, προτού υπολογιστεί ο δείκτης εγκυρότητας, κάθε στοιχείο του διανύσματος κανονικοποιείται στο διάστημα $[0,1]$ με βάση τον παρακάτω τύπο:

$$v_u^*(K, \mu; X) = \frac{v_u(K, \mu; X)}{\max_{K=2, \dots, K_{max}} \{v_u(K, \mu; X)\} - \min_{K=2, \dots, K_{max}} v_u(K, \mu; X)} \quad (31)$$

Η διαδικασία κανονικοποίησης επαναλαμβάνεται με τον ίδιο τρόπο για το διάνυσμα V_o . Ο δείκτης εγκυρότητας δίνεται λοιπόν από το άθροισμα των δύο κανονικοποιημένων διανυσμάτων, δηλαδή:

$$v_{sv}(K, \mu; X) = v_u^*(K, \mu; X) + v_o^*(K, \mu) \quad (32)$$

Αφού όπως προαναφέρθηκε και οι δύο συναρτήσεις έχουν μικρή τιμή όταν ο αριθμός των τμημάτων γίνει ίσος με τον βέλτιστο, είναι προφανές ότι η ελάχιστη τιμή του δείκτη εγκυρότητας προκύπτει όταν το K γίνει ίσο με το K^* .

6. Αλγόριθμοι εμπνευσμένοι από τη φύση

Καθώς ο αριθμός των διαθέσιμων δεδομένων αυξάνεται, η ανάγκη τμηματοποίησης των δεδομένων αυτών όπως προαναφέρθηκε, είναι πιο επιτακτική από ποτέ. Είναι αδύνατον ο οποιοσδήποτε χρήστης να μπορεί να λειτουργήσει αποδοτικά και να λάβει σωστές αποφάσεις, εάν πρώτα δεν διαχωρίσει τα δεδομένα σε αντίστοιχες ομάδες που μοιράζονται παρόμοια χαρακτηριστικά. Προκειμένου να αντιληφθούμε το μέγεθος του προβλήματος, αρκεί να σκεφτούμε ότι εάν έχουμε n αντικείμενα που θέλουμε να τα κατατάξουμε σε k ομάδες οι διαφορετικοί τρόποι ταξινόμησης των αντικειμένων είναι σύμφωνα με τον Liu [31]:

$$N(n, k) = \frac{1}{k!} \sum_{i=0}^k (-1)^i \binom{k}{i} (k-i)^n \quad (33)$$

Για παράδειγμα, εάν έχουμε 25 αντικείμενα που θέλουμε να χωρίσουμε σε 5 ομάδες, υπάρχουν 2.436.684.974.110.751 διαφορετικοί τρόποι για να γίνει αυτό. Ειδικά δε στην περίπτωση που δεν γνωρίζουμε τον αριθμό των ομάδων και πρέπει να εξετάσουμε ένα πλήθος από k οι συνδυασμοί είναι ακόμη περισσότεροι και δίνονται από τον τύπο:

$$\sum_{i=1}^n N(n, i) \quad (34)$$

Οι παραδοσιακοί αλγόριθμοι τμηματοποίησης, εξερευνούν ένα μικρό υποσύνολο του πεδίου λύσεων, το οποίο ορίζεται από διάφορες παραμέτρους όπως για παράδειγμα ο αριθμός των τμημάτων, το κριτήριο τμηματοποίησης και τη μέθοδο που ακολουθείται. Αλγόριθμοι όπως ο μονής ακμής λειτουργούν ντετερμινιστικά και θα δώσουν πάντα το ίδιο αποτέλεσμα, ενώ αλγόριθμοι όπως ο γνωστός k-means λειτουργούν ξεινώντας από ένα τυχαίο (ή μη) αρχικό σημείο, με αποτέλεσμα η τελική λύση να εξαρτάται από την επιλογή αρχικού σημείου. Σε κάθε περίπτωση η λύση είναι τοπικά βέλτιστη, πράγμα που ασφαλώς δεν εξασφαλίζει ότι είναι και ολικά βέλτιστη.

Εξαιτίας λοιπόν αυτής της φύσης του προβλήματος, οι ερευνητές διερευνούν αλγόριθμους οι οποίοι μπορούν να εξερευνήσουν σχετικά αποτελεσματικά μεγάλους χώρους λύσεων. Πρόσφατα, οι αλγόριθμοι βελτιστοποίησης που βασίζονται στην συμπεριφορά των μελισσών έχουν αποδειχθεί αρκετά αποτελεσματικοί και σκοπός της συγκεκριμένης εργασίας είναι να διερευνηθεί η απόδοσή τους με εφαρμογή στο πρόβλημα της τμηματοποίησης της αγοράς.

6.1. Αλγόριθμος Βελτιστοποίησης Ζευγαρώματος Μελισσών (Honey Bees Mating Optimization –HBMO)

Ο αλγόριθμος HBMO βασίζεται στην συμπεριφορά ζευγαρώματος των μελισσών και συγκεκριμένα στη διαδικασία με την οποία η βασίλισσα ζευγαρώνει με τους κηφήνες της κυψέλης. Ο εν λόγω αλγόριθμος, παρότι το γεγονός ότι είναι αρκετά πρόσφατος, έχει τραβήξει το ενδιαφέρον πολλών ερευνητικών ομάδων και έχει εφαρμοστεί σε διάφορα προβλήματα βελτιστοποίησης. Ενδεικτικά ο αναγνώστης παραπέμπεται στα [32]–[43].

Αρχικά, θα γίνει μία σύντομη περιγραφή του τρόπου με τον οποίο λειτουργεί μία κοινότητα μελισσών. Ως γνωστόν οι μέλισσες ζουν, εργάζονται, αναπτύσσονται και αναπαράγονται σε ομάδες. Κάθε ομάδα μελισσών, δουλεύει και κατασκευάζει φωλιές, οι οποίες ονομάζονται κυψέλες. Μία τυπική αποικία μελισσών αποτελείται συνήθως από μία βασίλισσα, η οποία είναι επιφορτισμένη με το καθήκον να παράγει αυγά, από μηδέν έως μερικές χιλιάδες κηφήνες (αναλόγως της περιόδου) και συνήθως από 10 έως 60 χιλιάδες εργάτριες. Η βασίλισσα αποτελεί το σημαντικότερο κομμάτι της κοινότητας μιας και είναι υπεύθυνη για την αναπαραγωγή. Συγκεκριμένα, μπορεί να γεννήσει μέχρι και 1500 αυγά ανά ημέρα και να ζήσει περίπου 5 με 6 έτη σε αντίθεση με τους κηφήνες και τις εργάτριες που δεν ζουν παραπάνω από 6 μήνες.

Η βασίλισσα τρέφεται με το λεγόμενο «βασιλικό πολτό» ο οποίος την κάνει μεγαλύτερη και πιο ανθεκτική από τις υπόλοιπες μέλισσες. Ο ρόλος των κηφήνων στην κυψέλη είναι αποκλειστικά η γονιμοποίηση της βασίλισσας και για αυτόν το λόγο πεθαίνουν μόλις ολοκληρωθεί η διαδικασία γονιμοποίησης στο τέλος της σεζόν. Ο ρόλος των εργατριών είναι η εκτέλεση όλων των απαραίτητων εργασιών που διασφαλίζουν την επιβίωση εντός της κυψέλης. Όταν είναι νέες μένουν στην κυψέλη και πραγματοποιούν κατασκευαστικές, κυρίως, εργασίες ενώ όταν ενηλικιωθούν συμμετέχουν στην αναζήτηση τροφής και νερού εκτός της κυψέλης.

Όσον αφορά τη διαδικασία ζευγαρώματος τώρα, που είναι και η ιδέα πάνω στην οποία βασίζεται ο συγκεκριμένος αλγόριθμος βελτιστοποίησης, χωρίζεται σε δύο φάσεις. Κατά την πρώτη φάση, η βασίλισσα εκτελεί έναν «χορό» προκειμένου να αντιληφθούν οι κηφήνες ότι ξεκινάει η διαδικασία ζευγαρώματος. Στη συνέχεια η βασίλισσα εκτελεί την «πτήση ζευγαρώματος» εκτός της κυψέλης. Κατά τη διάρκεια αυτής της πτήσης, η βασίλισσα καταδιώκεται από ένα μεγάλο σμήνος κηφήνων. Το ζευγάρι γίνεται στον αέρα από όσους κηφήνες καταφέρουν να «πιάσουν» τη βασίλισσα, οι οποίοι είναι συνήθως περίπου στους 18. Το σπέρμα όλων των κηφήνων που θα καταφέρουν να ζευγαρώσουν με τη βασίλισσα αποθηκεύεται στην σπερματοθήκη της, ενώ αμέσως μετά το ζευγάρι οι αντίστοιχοι κηφήνες πεθαίνουν. Κάθε φορά που η βασίλισσα γεννάει γονιμοποιημένα αυγά, χρησιμοποιεί μία μείξη από το σπέρμα που έχει αποθηκεύσει από όλους τους κηφήνες στην σπερματοθήκη της. Με βάση την παραπάνω περιγραφή λοιπόν, μπορούμε να προχωρήσουμε στην περιγραφή του αλγορίθμου.

Ο χρήστης αρχικά ορίζει το μέγεθος του αρχικού πληθυσμού της κυψέλης και τον αριθμό των βασιλισσών. Κάθε μέλος του πληθυσμού αντιστοιχεί σε μία λύση του προβλήματος βελτιστοποίησης που εξετάζουμε. Είναι άξιο αναφοράς σε αυτό το κομμάτι, ότι ενώ στην πραγματικότητα η βασίλισσα είναι μόνο μία στην υλοποίηση του αλγορίθμου επιτρέπεται να έχουμε και παραπάνω βασιλίσσες. Ο αρχικός πληθυσμός μπορεί να παραχθεί είτε τυχαία είτε όχι, συνήθως όμως επιλέγεται τυχαίος αρχικός πληθυσμός. Τα καλύτερα μέλη του πληθυσμού αυτού ορίζονται ως βασιλίσσες ενώ όλα τα υπόλοιπα αποτελούν τους κηφήνες. Οι εργάτριες σε αυτόν τον αλγόριθμο δεν αποτελούν μέλη του πληθυσμού, δεν είναι δηλαδή λύσεις του προβλήματος. Αντίθετα, κάθε εργάτρια αντιστοιχεί σε μία διαφορετική μέθοδο τοπικής αναζήτησης όπως θα περιγραφεί και στη συνέχεια. Ο γονότυπος της κάθε μέλισσας ορίζεται ως ένα διάνυσμα p διαστάσεων ενώ η απόδοσή της εκτιμάται από μία συνάρτηση ποιότητας (fitness function).

Εκτός από τις παραμέτρους που αναφέρθηκαν παραπάνω, ο χρήστης πρέπει να ορίσει και τον μέγιστο αριθμό ζευγαρώματων για κάθε βασίλισσα σε μία πτήση ζευγαρώματος. Αυτή η παράμετρος αντιστοιχεί ουσιαστικά στην χωρητικότητα της σπερματοθήκης της. Έτσι σε κάθε πτήση, η βασίλισσα ζευγαρώνει με έναν ή περισσότερους κηφήνες, αποθηκεύοντας τον γονότυπό τους στην σπερματοθήκη της. Επιπλέον ορίζεται και ο αριθμός των νεογνών που θα γεννήσουν οι βασιλίσσες, ο οποίος μπορεί να είναι ίσος, μικρότερος ή και μεγαλύτερος από το μέγεθος της σπερματοθήκης. Το τελευταίο είναι εφικτό για δύο κυρίως λόγους. Αφενός το κάθε νεογνό λαμβάνει γονίδια από τη βασίλισσα και έναν ή πολλούς κηφήνες. Αφετέρου, το γεγονός ότι ένα νεογνό έχει προέλθει από διασταύρωση των γονιδίων των γονέων του, δεν σημαίνει ότι θα είναι πανομοιότυπο με οποιοδήποτε άλλο νεογνό προκύψει από τους ίδιους γονείς.

Η αλγοριθμική υλοποίηση της πτήσης ζευγαρώματος της βασίλισσας μπορεί να θεωρηθεί ως ένα σύνολο από μεταβάσεις στο χώρο λύσεων που ορίζεται από τους κηφήνες. Όταν μία βασίλισσα βρεθεί στη συγκεκριμένη θέση που βρίσκεται ο κηφήνας, η πιθανότητα να ζευγαρώσει μαζί του δίνεται από τον τύπο:

$$P(D) = e^{\left| -\frac{\Delta(f)}{Speed(t)} \right|} \quad (35)$$

Το $P(D)$ είναι η πιθανότητα να προστεθεί το σπέρμα του κηφήνα D στην σπερματοθήκη της βασίλισσας. $\Delta(f)$ είναι η διαφορά στην τιμή της συνάρτησης ποιότητας μεταξύ της βασίλισσας και του κηφήνα. Ο όρος αυτός κάνει την πιθανότητα ζευγαρώματος μεγάλη, όταν η τιμή της συνάρτησης ποιότητας του κηφήνα είναι κοντά ή μεγαλύτερη από αυτήν της βασίλισσας. Ο όρος $Speed(t)$ αντιστοιχεί στην τιμή της ταχύτητας της βασίλισσας και πρακτικά επηρεάζει την πιθανότητα ζευγαρώματος ως εξής. Εάν η ταχύτητα είναι μεγάλη, σημαίνει ότι η βασίλισσα βρίσκεται στην αρχή της πτήσης της και κατά συνέπεια η πιθανότητα ζευγαρώματος είναι μεγάλη. Η ταχύτητα της βασίλισσας μειώνεται μετά από κάθε μετάβαση στον χώρο των λύσεων και αυτό υλοποιείται αλγοριθμικά με βάση τον παρακάτω τύπο.

$$Speed(t + 1) = \alpha \cdot Speed(t) \quad (36)$$

Εκτός από την ταχύτητα, ορίζουμε και την ενέργεια της βασίλισσας η οποία καθορίζει το πότε τελειώνει η πτήση ζευγαρώματος. Μετά από κάθε μετάβαση στο χώρο λύσεων η ενέργεια μειώνεται και αυτή όπως και η ταχύτητα, με βάση τον παρακάτω τύπο:

$$energy(t + 1) = energy(t) - step \quad (37)$$

Οι παράμετροι α και $step$ μπορούν είτε να είναι προκαθορισμένοι από τον χρήστη είτε να επιλέγονται τυχαία κατά την εκτέλεση του αλγορίθμου ή ακόμη και κατά την έναρξη της κάθε μεμονωμένης πτήσης. Μία πτήση ζευγαρώματος ολοκληρώνεται είτε εάν γεμίσει η σπερματοθήκη της βασίλισσας είτε εάν τελειώσει η ενέργειά της.

Η επόμενη φάση του αλγορίθμου είναι η παραγωγή των νεογνών. Όπως προαναφέρθηκε, η σπερματοθήκη της βασίλισσας περιέχει σπέρμα από όλους τους κηφήνες με τους οποίους ζευγάρωσε. Με τη χρήση ενός οποιουδήποτε τελεστή διασταύρωσης ικανού να χρησιμοποιεί περισσότερους από δύο γονείς, είναι πλέον εφικτή η δημιουργία των νεογνών (οι τελεστές που θα χρησιμοποιηθούν παρατίθενται σε επόμενο κεφάλαιο).

Οι εργάτριες στο συγκεκριμένο αλγόριθμο, σε αντίθεση με άλλους αλγορίθμους μελισσών, δεν αποτελούν μέλη του πληθυσμού δηλαδή πιθανές λύσεις του προβλήματος αλλά λειτουργούν ως διαδικασίες τοπικής αναζήτησης. Κατ' αναλογία με τις πραγματικές εργάτριες στην κυψέλη οι οποίες ταΐζουν τα νεογνά μέχρι να μεγαλώσουν, έτσι και οι εργάτριες στον αλγόριθμο αναλαμβάνουν να βελτιώσουν τις νέες λύσεις (νεογνά). Οι διαφορετικές ικανότητες των εργατριών αντιστοιχούν σε διαφορετικές διαδικασίες τοπικής

αναζήτησης, οι οποίες καθορίζονται από τον χρήστη. Η ανάθεση του κάθε νεογνού σε μία ή περισσότερες εργάτριες μπορεί να γίνεται είτε με τυχαίο τρόπο είτε με κάποιο συγκεκριμένο κανόνα, όπως για παράδειγμα να επιλέγεται η εργάτρια αναλόγως της αποτελεσματικότητάς της.

Μετά το πέρας της διαδικασίας ανατροφής των νεογνών, αξιολογείται η τιμή της συνάρτησης ποιότητάς τους και εφόσον κάποιο από αυτά ξεπεράσει την χειρότερη βασίλισσα, γίνεται βασίλισσα. Η τρέχουσα βασίλισσα φεύγει από την κυψέλη, διαγράφεται δηλαδή από τον πληθυσμό. Εάν η τιμή της συνάρτησης ποιότητας δεν ξεπεράσει της βασίλισσας τότε το συγκεκριμένο νεογνό θα έχει το ρόλο του κηφήνα στην επόμενη επανάληψη του αλγορίθμου. Ο αριθμός των κηφήνων πρέπει να διατηρείται σταθερός σε κάθε επανάληψη του αλγορίθμου και για αυτόν το λόγο στην περίπτωση που τα νεογνά που δημιουργήθηκαν είναι περισσότερα από τους κηφήνες που διαγράφηκαν, επιλέγονται μόνο όσοι κηφήνες χρειάζονται για να συμπληρωθεί ο πληθυσμός. Για την καλύτερη κατανόηση του αλγορίθμου παρατίθεται ο αντίστοιχος ψευδοκώδικας:

Αλγόριθμος Βελτιστοποίησης Ζευγαρώματος Μελισσών

Αρχικοποίηση

Δημιουργία αρχικού πληθυσμού

Επιλογή των βέλτιστων μελών του πληθυσμού ως βασίλισσες

Επιλογή μέγιστου αριθμού πτήσεων ζευγαρώματος (M)

Όσο $i \leq M$

 Για κάθε βασίλισσα

 Αρχικοποίηση της σπερματοθήκης, της ενέργειας και της ταχύτητας της βασίλισσας

 Επιλογή των παραμέτρων a και $step$

 Όσο $energy > 0$ και η σπερματοθήκη δεν είναι γεμάτη

 Επιλογή ενός κηφήνα από τον πληθυσμό

 Εάν το ζευγάρι είναι επιτυχές τότε

 Προσθήκη του σπέρματος του κηφήνα στην σπερματοθήκη της βασίλισσας

 Διαγραφή του κηφήνα από τον πληθυσμό

 Τέλος αν

$speed = a \times speed$

$energy = energy - step$

 Τέλος Όσο

 Για $j = 1$ έως το μέγεθος της σπερματοθήκης

 Επιλογή σπέρματος από τη σπερματοθήκη

 Δημιουργία νεογνού με χρήση ενός τελεστή διασταύρωσης

 Επιλογή μίας εργάτριας

 Βελτίωση της λύσης του νεογνού με χρήση της εργάτριας

 Εάν η νέα λύση είναι καλύτερη από τη βασίλισσα

 Αντικατάσταση της βασίλισσας από το νεογνό

 Διαγραφή της παλιάς βασίλισσας

 Αλλιώς

 Προσθήκη του νεογνού στον πληθυσμό των κηφήνων

 Τέλος αν

 Τέλος

Τέλος

Τέλος

Επιστροφή Βέλτιστης Λύσης (καλύτερη βασίλισσα)

6.2. Αλγόριθμος Βελτιστοποίησης Ζευγαρώματος Μπάμπουρων (Bumble Bees Mating Optimization – BBMO)

Η συμπεριφορά των μπάμπουρων έχει και αυτή με τη σειρά της εμπνεύσει ερευνητές στο να την μιμηθούν και να την εφαρμόσουν σε προβλήματα βελτιστοποίησης. Ο αντίστοιχος αλγόριθμος ονομάστηκε Αλγόριθμος Βελτιστοποίησης Ζευγαρώματος Μπάμπουρων (Bumble Bees Mating Optimization) [40], [44], [45] και βασίζεται στην συμπεριφορά ζευγαρώματος της βασίλισσας των μπάμπουρων.

Οι μπάμπουρες, όμοια με τις μέλισσες, είναι κοινωνικά έντομα των οποίων οι κοινότητες αποτελούνται από τη βασίλισσα, πολλές εργάτριες και τους κηφήνες. Οι βασίλισσες είναι τα μόνα μέλη του πληθυσμού που επιβιώνουν από τη μία χρονιά στην άλλη, περνώντας τον χειμώνα σε χειμερία νάρκη. Με τον ερχομό της θερινής περιόδου οι βασίλισσες ξυπνούν από τη χειμερία νάρκη, συγκεντρώνουν γύρη και νέκταρ και στη συνέχεια προετοιμάζουν την κυψέλη όπου θα γεννήσουν τα αυγά τους. Η βασίλισσα γεννάει δύο ειδών αυγά: τα γονιμοποιημένα και τα μη γονιμοποιημένα. Τα πρώτα περιέχουν γονίδια από τη βασίλισσα και από έναν ή παραπάνω κηφήνες με τους οποίους ζευγάρωσε η βασίλισσα τον προηγούμενο χρόνο. Από τα αυγά αυτά προκύπτουν οι εργάτριες του επόμενου χρόνου. Το δεύτερο είδος αυγών περιλαμβάνει γονίδια μόνο από τη βασίλισσα και οι μπάμπουρες που προκύπτουν θα είναι οι κηφήνες του επόμενου έτους.

Με τη δημιουργία της κοινότητας, τον ρόλο της εύρεσης τροφής τον αναλαμβάνουν πλέον κάποιες από τις εργάτριες. Μία δεύτερη ομάδα εργατριών αναλαμβάνει την ανατροφή των νεογνών. Τέλος, οι εργάτριες έχουν τη δυνατότητα παραγωγής αυγών, χωρίς όμως να είναι γονιμοποιημένα, όταν η ικανότητα της βασίλισσας να παράγει αυγά μειωθεί. Από τα αυγά της βασίλισσας αναδεικνύονται οι νέες βασίλισσες. Στη συνέχεια καθώς το τέλος του έτους πλησιάζει οι κηφήνες αποχωρούν από τη φωλιά όπως και οι νέες βασίλισσες και η αποικία αρχίζει να παρακμάζει. Οι νέες βασίλισσες ζευγαρώνουν με τους κηφήνες, αποθηκεύουν το σπέρμα τους στην σπερματοθήκη τους και πέφτουν σε χειμερία νάρκη.

Όσον αφορά την αλγοριθμική υλοποίηση των παραπάνω τώρα, αρχικά χωρίζουμε τον πληθυσμό σε τρεις διαφορετικές ομάδες: στις βασίλισσες, στις εργάτριες και στους κηφήνες. Ο αρχικός πληθυσμός λύσεων επιλέγεται τυχαία και όλα τα μέλη του έχουν το ρόλο του κηφήνα. Στη συνέχεια υπολογίζουμε την τιμή της συνάρτησης ποιότητας και ο μπάμπουρας με την καλύτερη τιμή γίνεται κηφήνας. Η βασίλισσα, αναζητάει στο πλήθος λύσεων τους κηφήνες με τους οποίους θα ζευγαρώσει με βάση το πόσο καλές λύσεις αντιπροσωπεύουν. Έτσι, κηφήνες οι οποίοι έχουν καλές τιμές της συνάρτησης ποιότητας είναι πιο πιθανό να ζευγαρώσουν με τη βασίλισσα. Κάθε φορά που πραγματοποιείται ένα ζευγάρωμα, ο γονότυπος του

κηφήνα αποθηκεύεται στη σπερματοθήκη της βασίλισσας μέχρι να επιτευχθεί ο μέγιστος αριθμός από ζευγαρώματα, τον οποίο τον ορίζει ως παράμετρο ο χρήστης.

Κατόπιν, η βασίλισσα πέφτει σε χειμερία νάρκη έως την επόμενη θερινή περίοδο όποτε και θα γεννήσει τα αυγά της. Από τα αυγά αυτά, θα γεννηθούν νέες βασίλισσες και εργάτριες μέσω της διασταύρωσης του γονότυπου της βασίλισσας και των κηφήνων με τους οποίους έχει ζευγαρώσει. Οι ισχυρότεροι από τους απογόνους προορίζονται για νέες βασίλισσες ενώ οι υπόλοιποι για νέες εργάτριες. Προτού όμως οι νέες βασίλισσες πάρουν το ρόλο τους, τρέφονται αρχικά από τις παλιές βασίλισσες και στη συνέχεια από τις εργάτριες. Αυτή η διαδικασία υλοποιείται αλγοριθμικά με βάση τον παρακάτω τύπο και αποτελεί ουσιαστικά μία διαδικασία τοπικής αναζήτησης.

$$nq_{ij} = nq_{ij} + \left(\frac{b_{max} - (b_{max} - b_{min}) * lsi}{lsi_{max}} \right) * (nq_{ij} - q) + \frac{1}{M} \sum_{k=1}^M \left(\frac{b_{min} - (b_{min} - b_{max}) * lsi}{lsi_{max}} \right) * (nq_{ij} - w_{kj}) \quad (38)$$

Στην παραπάνω εξίσωση με nq_{ij} συμβολίζεται η νέα λύση της βασίλισσας i , με q η λύση της παλιάς βασίλισσας, με w_{kj} η λύση της εργάτριας k , με M τον μέγιστο αριθμό εργατριών από τις οποίες θα τραφεί η νέα βασίλισσα και οι παράμετροι b_{min} και b_{max} που καθορίζουν εάν και κατά πόσο η νέα βασίλισσα θα τρέφεται από τις εργάτριες και την παλιά βασίλισσα. Ο μέγιστος αριθμός επαναλήψεων της τοπικής αναζήτησης ορίζεται με lsi_{max} ενώ με lsi η τρέχουσα επανάληψη.

Οι κηφήνες, όπως προαναφέρθηκε παράγονται με μετάλλαξη του γονότυπου των παλαιών βασιλισσών ή των εργατριών. Η υλοποίηση της μετάλλαξης μπορεί να είναι μία τυχαία αλλαγή σε κάποιο ή κάποια από τα στοιχεία του διανύσματος. Δεν αποκλείεται όμως να είναι και μία διαδικασία τοπικής αναζήτησης ανάλογα με το πρόβλημα. Η κίνηση των κηφήνων καθώς απομακρύνονται από την κυψέλη προκειμένου να ζευγαρώσουν γίνεται σε σχηματισμό σμήνους. Αυτό μοντελοποιείται αλγοριθμικά ως εξής:

$$d_{ij} = d_{ij} + a * (d_{kj} - d_{lj}) \quad (39)$$

Όπου τα d_{ij} , d_{kj} , d_{lj} αποτελούν τις λύσεις των κηφήνων i , k και l αντίστοιχα ενώ με a συμβολίζουμε μία παράμετρο που ελέγχει το κατά πόσο επηρεάζεται ο κηφήνας i από τους υπόλοιπους. Η κάθε νέα βασίλισσα επιλέγει τους κηφήνες με τους οποίους θα ζευγαρώσει ανάλογα με το πόσο ισχυροί είναι και

αποθηκεύει το σπέρμα τους στην σπερματοθήκη της. Στη συνέχεια όλα τα μέλη του πληθυσμού πεθαίνουν και επιβιώνουν μόνο οι βασίλισσες.

Αλγόριθμος Βελτιστοποίηση Ζευγαρώματος Μπάμπουρων

Καθορισμός μέγιστου αριθμού επαναλήψεων

Καθορισμός μέγιστου αριθμού ζευγαρώματων

Καθορισμός μέγιστου αριθμού βασιλισσών

Αρχικοποίηση

Δημιουργία αρχικού πληθυσμού

Υπολογισμός της αντικειμενικής συνάρτησης για κάθε μπάμπουρα

Επιλογή των καλύτερων μπάμπουρων ως βασίλισσες

Επιλογή των υπόλοιπων μπάμπουρων ως κηφήνες

Ταξινόμηση των κηφήνων με βάση την τιμή της αντικειμενικής συνάρτησης

Επιλογή των κηφήνων που θα ζευγαρώσουν με την κάθε βασίλισσα

Αποθήκευση του γονότυπου των κηφήνων στις σπερματοθήκες των βασιλισσών

Κύρια Φάση

Μέχρι τον μέγιστο αριθμό επαναλήψεων

Δημιουργία απογόνων με χρήση ενός τελεστή διασταύρωσης

Υπολογισμός της αντικειμενικής συνάρτησης για κάθε απόγονο

Ταξινόμηση των απογόνων με βάση την τιμή της αντικειμενικής συνάρτησης

Επιλογή των καλύτερων απογόνων ως νέες βασίλισσες

Επιλογή των υπόλοιπων απογόνων ως εργάτριες

Ανατροφή κάθε νέας βασίλισσας από τις εργάτριες

Δημιουργία ενός ποσοστού κηφήνων από μετάλλαξη των παλαιών βασιλισσών

Δημιουργία των υπόλοιπων κηφήνων από μετάλλαξη των εργατριών

Υπολογισμός της τιμής της αντικειμενικής συνάρτησης για κάθε κηφήνα

Ταξινόμηση των κηφήνων με βάση την τιμή της αντικειμενικής συνάρτησης

Μέχρι τον μέγιστο αριθμό ζευγαρώματων ανά βασίλισσα

Επιλογή των κηφήνων που θα ζευγαρώσουν με τη βασίλισσα

Αποθήκευση του γονότυπου των κηφήνων στην σπερματοθήκη της βασίλισσας

Τέλος

Επιβίωση των βασιλισσών

Διαγραφή όλων των υπολοίπων μελών του πληθυσμού

Τέλος

Επιστροφή της καλύτερης βασίλισσας

6.3. Αλγόριθμος Τεχνητής Αποικίας Μελισσών (Artificial Bee Colony Optimization – ABC)

Ο αλγόριθμος τεχνητής αποικίας μελισσών αντίθετα με τους δύο προηγούμενους ανήκει στην κατηγορία αλγορίθμων που βασίζονται στην προσομοίωση της διαδικασίας εύρεσης τροφής. Είναι ο πιο διαδεδομένος αλγόριθμος μελισσών σε αυτήν την κατηγορία και έχει χρησιμοποιηθεί με επιτυχία σε αρκετές μελέτες [46], [47]. Εφαρμόζεται κυρίως σε συνεχή προβλήματα και στηρίζεται στη λογική ότι περιοχές με μεγάλες ποσότητες τροφής προσελκύουν περισσότερες μέλισσες. Ο πληθυσμός των μελισσών απαρτίζεται από τρεις κατηγορίες. Η πρώτη κατηγορία είναι οι εξερευνητρίες (employed bees), ο ρόλος των οποίων είναι να εξερευνούν τις διάφορες πηγές τροφής που υπάρχουν στο χώρο και να ενημερώνουν τις υπόλοιπες μέλισσες σχετικά με τα ευρήματά τους. Η δεύτερη κατηγορία είναι οι μέλισσες «θεατές» (onlooker bees), οι οποίες περιμένουν στην κυψέλη τις εξερευνητρίες προκειμένου να ενημερωθούν σχετικά με τις «καλές» θέσεις τροφής και να κινηθούν προς αυτές. Τέλος, η τρίτη κατηγορία μελισσών ονομάζεται «ανιχνευτές» (scout bees), οι οποίες είναι στην ουσία εξερευνητρίες των οποίων η πηγή τροφής τελείωσε και αναζητούν τυχαία στο χώρο για μία νέα πηγή τροφής.

Κατά την αρχικοποίηση του αλγορίθμου, επιλέγονται τυχαία κάποιες πηγές τροφής από τις εξερευνητρίες και αξιολογούνται ως προς την ποσότητα τροφής που περιέχουν, πράγμα που αντιστοιχεί στην τιμή της αντικειμενικής συνάρτησης. Στη συνέχεια, οι εξερευνητρίες επιστρέφουν στην κυψέλη και ενημερώνουν τις μέλισσες θεατές σχετικά με τις θέσεις που βρίσκονται οι πηγές τροφής. Οι μέλισσες θεατές επιλέγουν να πάνε σε μία συγκεκριμένη πηγή τροφής βασιζόμενοι σε έναν πιθανοτικό κανόνα, ο οποίος πρακτικά σημαίνει ότι όσο καλύτερη είναι η πηγή τροφής τόσο μεγαλύτερη είναι η πιθανότητα να την επισκεφτούν οι μέλισσες θεατές. Η πιθανότητα να επιλεγεί μία πηγή τροφής δίνεται από τον παρακάτω τύπο:

$$p_i = \frac{f_i}{\sum_{n=1}^N f_n} \quad (40)$$

Όπου με p_i συμβολίζουμε την πιθανότητα επιλογής της πηγής και με f_i την τιμή της αντικειμενικής συνάρτησης της πηγής i . Οι εξερευνητρίες αλλά και οι θεατές επισκέπτονται και αξιολογούν και γειτονικές πηγές τροφής εκτός από την τρέχουσα. Αυτό αλγοριθμικά μοντελοποιείται με βάση την παρακάτω εξίσωση:

$$x_{ij} = x_{ij} + rand * (x_{ij} - x_{kj}) \quad (41)$$

Με x_{ij} συμβολίζουμε την τρέχουσα πηγή τροφής, με x_{kj} μία διαφορετική πηγή τροφής από την i και με $rand$ έναν τυχαίο αριθμό εντός του διαστήματος $(0,1)$.

Κατόπιν αξιολογείται η νέα αυτή πηγή τροφής ως προς την τιμή της συνάρτησης ποιότητας και σε περίπτωση που είναι καλύτερη από την αρχική, την αντικαθιστά. Μόλις ολοκληρωθεί η διαδικασία για όλες τις μέλισσες, επιστρέφουν στην κυψέλη για να ενημερώσουν τις θεατές σχετικά με τις νέες πηγές τροφής πλέον. Εάν μετά από κάποιο αριθμό επαναλήψεων η λύση δεν έχει βελτιωθεί, θεωρείται ότι έχει εξαντληθεί και η εξερευνητρια παίρνει το ρόλο ανιχνεύτριας, ψάχνοντας για τροφή σε τυχαίο σημείο του χώρου λύσεων. Μία ιδιομορφία του συγκεκριμένου αλγορίθμου είναι ότι κατ' ουσίαν οι μέλισσες δεν αποτελούν λύσεις του προβλήματος, αλλά διαδικασίες που ενεργούν πάνω σε μία λύση (πηγή τροφής) προκειμένου να την βελτιώσουν. Για παράδειγμα, εάν μία πηγή τροφής επιλεγεί από δέκα θεατές μέλισσες, αυτό σημαίνει ότι θα γίνουν δέκα προσπάθειες βελτίωσης της συγκεκριμένης λύσης.

Στη συνέχεια παρουσιάζουμε τον ψευδοκώδικα του αλγορίθμου προκειμένου να γίνει ευκολότερη η κατανόησή του.

Αλγόριθμος Τεχνητής Αποικίας Μελισσών

Καθορισμός Παραμέτρων

- Καθορισμός μέγιστου αριθμού επαναλήψεων
- Καθορισμός του αριθμού πηγών τροφής
- Καθορισμός του αριθμού των εξερευνητριών
- Καθορισμός του αριθμού των θεατών

Αρχικοποίηση

- Δημιουργία του αρχικού πληθυσμού (πηγές τροφής)
- Επιλογή μίας πηγής από κάθε εξερευνητρια
- Υπολογισμός της τιμής της αντικειμενικής συνάρτησης για κάθε πηγή τροφής

Κύρια Φάση

- Μέχρι τον μέγιστο αριθμό επαναλήψεων
 - Επιστροφή της εξερευνητριας στην κυψέλη
 - Επιλογή πηγής τροφής από κάθε μέλισσα θεατή
 - Αποστολή των θεατών και της εξερευνητριας στην πηγή τροφής
 - Δημιουργία μίας καινούριας πηγής τροφής με βάση την προηγούμενη
 - Υπολογισμός της τιμής της αντικειμενικής συνάρτησης για κάθε νέα πηγή τροφής
 - Εάν βρεθεί καλύτερη πηγή τροφής
 - Αντικατάσταση της παλαιάς πηγής από τη νέα
 - Τέλος αν
 - Εάν οι προσπάθειες βελτίωσης σε μία συγκεκριμένη πηγή υπερβούν το όριο
 - Τοποθέτηση μίας ανιχνεύτριας σε τυχαία θέση
 - Τέλος αν
 - Διατήρηση της βέλτιστης λύσης

Τέλος

Επιστροφή της καλύτερης πηγής τροφής

7. Προσαρμογή των αλγορίθμων στο πρόβλημα

7.1. Συνάρτηση ποιότητας

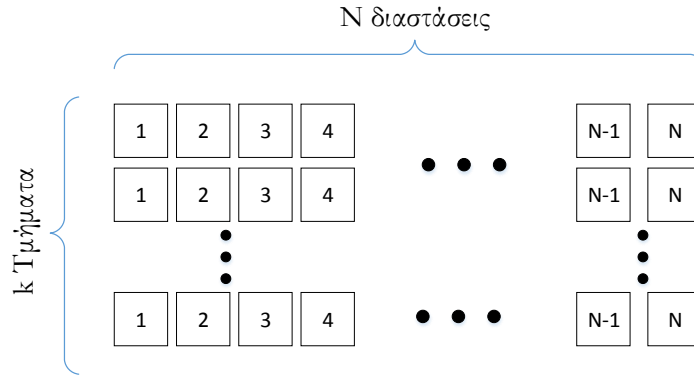
Ως αντικειμενική συνάρτηση ποιότητας θα χρησιμοποιήσουμε το άθροισμα των τετραγώνων των αποστάσεων του κάθε αντικειμένου από το κέντρο του τμήματος στο οποίο ανήκει. Προφανώς θεωρούμε ότι ένα αντικείμενο ανήκει σε ένα τμήμα, όταν βρίσκεται πλησιέστερα στο εν λόγω τμήμα από ότι σε οποιοδήποτε άλλο.

Πρέπει να αναφερθεί σε αυτό το σημείο, ότι σε κάθε αξιολόγηση της συνάρτησης ποιότητας ελέγχεται εάν υπάρχουν άδεια τμήματα. Σε περίπτωση που υπάρχουν άδεια τμήματα, παρακάμπτεται ο τρόπος υπολογισμού και ως τιμή της αντικειμενικής συνάρτησης θέτουμε μία πολύ μεγάλη τιμή, έτσι ώστε να την απορρίψει ο αλγόριθμος.

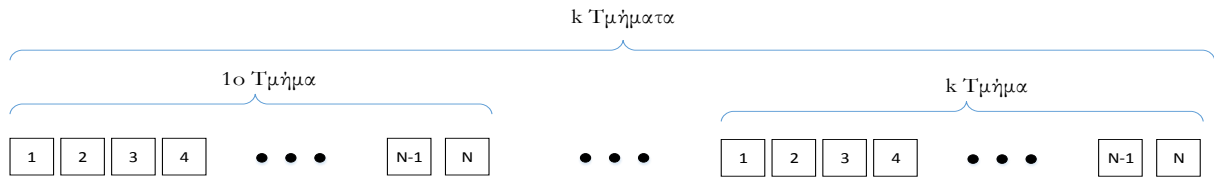
7.2. Αναπαράσταση λύσεων

Κατά καιρούς έχουν προταθεί στη βιβλιογραφία διάφοροι μέθοδοι αναπαράστασης της λύσης του προβλήματος. Προτού όμως προχωρήσουμε στην αναφορά και τον σχολιασμό τους, κρίνεται σκόπιμο να αναφέρουμε κάποιες βασικές προδιαγραφές που πρέπει να πληρούν. Αρχικά ο τρόπος αναπαράστασης της λύσης πρέπει να είναι πλήρης, να προσφέρει δηλαδή τη δυνατότητα αναπαράστασης όλων των πιθανών λύσεων του χώρου του προβλήματος. Μία δεύτερη απαίτηση είναι η εγκυρότητα, με την έννοια ότι η εκάστοτε λύση με τον τρόπο που αναπαρίσταται είναι έγκυρη και εντός του πεδίου λύσεων του προβλήματος [48].

Στην παρούσα εργασία ελέγχουμε την επίδοση διαφόρων μορφών αναπαράστασης. Αρχικά θεωρούμε ότι η λύση αναπαρίσταται από τα κέντρα των εκάστοτε τμημάτων, που είναι πραγματικοί αριθμοί. Έτσι για παράδειγμα εάν τα δεδομένα μας έχουν N διαστάσεις και επιθυμούμε την ταξινόμησή τους σε k τμήματα, τότε η λύση μας μπορεί να αναπαρασταθεί με τους εξής τρόπους:



Εικόνα 4 - Αναπαράσταση Λύσης μέσω των συντεταγμένων του κέντρου του τμήματος (α' τρόπος)



Εικόνα 5 - Αναπαράσταση Λύσης μέσω των συντεταγμένων του κέντρου του τμήματος (β' τρόπος)

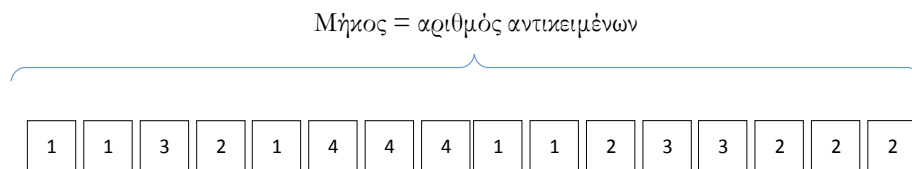
Μία δεύτερη προσέγγιση στο πρόβλημα της αναπαράστασης είναι η αποκαλούμενη δυαδική [48] η οποία βασίζεται σε ένα διάνυσμα μήκους M , όπου M ο αριθμός των δεδομένων. Ουσιαστικά αυτή η μέθοδος επιλέγει ποια από τα δεδομένα θεωρούνται «πρότυπα» ή με άλλα λόγια ποια από τα δεδομένα αποτελούν ουσιαστικά τα κέντρα του αντίστοιχου τμήματος. Έτσι μαρκάρονται με «1» τα δεδομένα που είναι πρότυπα και με «0» όλα τα υπόλοιπα. Προφανώς ο αριθμός των «1» θα πρέπει να είναι ίσος με τον αριθμό των επιθυμητών τμημάτων.

Συνεχίζοντας στην κατηγορία της δυαδικής αναπαράστασης, έχει προταθεί η εναλλακτική της δημιουργίας ενός πίνακα $k \times M$ του οποίου οι γραμμές αντιστοιχούν στα τμήματα και οι στήλες είναι τα δεδομένα. Έτσι, εάν το j -οστό στοιχείο ανήκει στο i -οστό τμήμα τότε το στοιχείο (i, j) του πίνακα λαμβάνει την τιμή «1». Όλες οι υπόλοιπες τιμές του πίνακα λαμβάνουν την τιμή «0».

Όπως είναι προφανές, η χρήση του πίνακα, αυξάνει τις απαιτήσεις σε μνήμη μιας και απαιτεί $O(k \cdot M)$ σε χώρο σε αντίθεση με το $O(M)$ που απαιτεί η χρήση του διανύσματος. Από την άλλη, ο χρόνος που

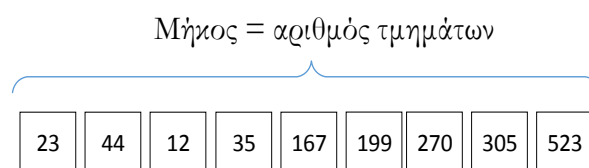
απαιτείται για να «αποκρυπτογραφηθεί» η λύση και να κατατάξουμε κάθε αντικείμενο σε ένα τμήμα είναι $O(k \cdot M)$ για την περίπτωση του πίνακα, ενώ $O(k \cdot n \cdot M)$ για την περίπτωση του διανύσματος, μιας και πρέπει για κάθε αντικείμενο, να εντοπίσουμε το πλησιέστερο πρότυπο και να το κατατάξουμε στο αντίστοιχο τμήμα.

Μία τρίτη προσέγγιση στο πρόβλημα της τμηματοποίησης είναι αντί να αποθηκεύουμε τα κέντρα των τμημάτων, να αποθηκεύουμε το σε ποιο τμήμα ανήκει το κάθε αντικείμενο από τα δεδομένα μας. Το μειονέκτημα αυτής της μεθόδου είναι ότι η ίδια λύση μπορεί να έχει διαφορετικές αναπαράστασεις. Για παράδειγμα η λύση [111122233] είναι απόλυτα ισοδύναμη με τις λύσεις [222233311], [111133322], [333311122], [333322211], [2222111133]. Κατά συνέπεια, ο χώρος λύσεων που πρέπει να εξερευνηθεί μεγαλώνει, πράγμα που μπορεί να οδηγήσει στη μείωση της αποτελεσματικότητας του αλγορίθμου.



Εικόνα 6 - Αναπαράσταση λύσης μέσω των τμημάτων που ανήκει το κάθε αντικείμενο

Στην ίδια κατηγορία αναπαράστασης, έχει προταθεί ο εξής τρόπος αναπαράστασης. Χρησιμοποιείται ένα διάνυσμα, του οποίου τα στοιχεία αντιστοιχούν το κάθε ένα σε ένα αντικείμενο πρότυπο. Για παράδειγμα, ένα τέτοιο διάνυσμα μπορεί να είναι για παράδειγμα το [5 9 12 22], το οποίο σημαίνει ότι τα αντικείμενα 5, 9, 12 και 22 είναι τα πρότυπα και αντιπροσωπεύουν τα τμήματα. Στη συνέχεια όλα τα υπόλοιπα αντικείμενα κατατάσσονται στο πλησιέστερο σε αυτά τμήμα. Πρέπει να τονίσουμε εδώ, ότι σε περίπτωση που ο αλγόριθμος επιτρέπει την ύπαρξη μη ταξινομημένων διανυσμάτων εμφανίζεται πάλι το πρόβλημα της πολλαπλής αναπαράστασης της ίδιας λύσης. Για παράδειγμα οι λύσεις [9 5 12 22], [12 5 9 22], [22 5 9 12] κ.ο.κ. αναπαριστούν την ίδια ακριβώς λύση.



Εικόνα 7 - Αναπαράσταση λύσης μέσω της αναπαράστασης των αντικειμένων-προτύπων

Τώρα, όσον αφορά το υπολογιστικό κόστος και συγκεκριμένα το κόστος σε χώρο μνήμης απαιτείται $O(M)$ στην πρώτη περίπτωση ενώ $O(k)$ στη δεύτερη. Παρόλα αυτά, δεν είναι ασφαλές να καταλήξει κάποιος στο συμπέρασμα ότι η χρήση της δεύτερης αναπαράστασης θα οδηγήσει σε καλύτερη αποτελεσματικότητα του αλγορίθμου. Αυτό οφείλεται στο γεγονός ότι ανάλογα με τον αλγόριθμο και τις ενέργειες που εκτελεί πάνω στα δεδομένα, απαιτείται διαφορετικό κάθε φορά υπολογιστικό κόστος. Για παράδειγμα ο υπολογισμός της fitness function είναι πολύ ευκολότερος στην πρώτη περίπτωση από ότι στη δεύτερη, μιας και γνωρίζουμε εκ των προτέρων την κατάταξη των αντικειμένων.

7.3. Ειδικά θέματα στη διακριτή αναπαράσταση της λύσης

Σε αυτό το υποκεφάλαιο της εργασίας θα εξετάσουμε θέματα που πρέπει να αντιμετωπιστούν όταν χρησιμοποιείται κάποια ακέραια αναπαράσταση της λύσης. Όπως προαναφέρθηκε, αντί μία λύση του προβλήματος να είναι οι συντεταγμένες των κέντρων των τμημάτων, η λύση είναι ένα διάνυσμα ίδιου μήκους με τον αριθμό των αντικειμένων στο σετ δεδομένων, του οποίου οι τιμές σε κάθε διάσταση αντιστοιχούν στο τμήμα στο οποίο ανήκει το κάθε αντικείμενο. Για παράδειγμα εάν το πλήθος των αντικειμένων είναι N τότε ένα διάνυσμα λύσης θα έχει διαστάσεις $1 \times N$ και εάν το i -οστό στοιχείο του έχει τιμή k , σημαίνει ότι το αντικείμενο i ανήκει στο k τμήμα. Ένας δεύτερος τρόπος αναπαράστασης που θα εξεταστεί είναι η αναπαράσταση βασισμένη σε «πρότυπα» ή αλλιώς medoids. Κατά αυτόν τον τρόπο μία λύση είναι ένα διάνυσμα μήκους k , η οποία περιέχει τους αύξοντες αριθμούς των αντικειμένων τα οποία θεωρούμε κέντρα των τμημάτων. Όπως είναι φυσιολογικό, οποιαδήποτε διακριτή αναπαράσταση και αν επιλέξουμε, είναι απαραίτητο να γίνουν κάποιες μετατροπές στους αλγορίθμους, προκειμένου να είναι σε θέση να λειτουργήσουν με τη συγκεκριμένη αναπαράσταση λύσης. Αρχικά θα αναφέρουμε κάποιες αλλαγές στον τρόπο υπολογισμού του κόστους, και στη συνέχεια θα αναφερθούμε στις αλλαγές που θα γίνουν στους αλγορίθμους.

Υπολογισμός Κόστους

Αυτό που αλλάζει στον υπολογισμό του κόστους είναι ότι πρέπει να προστεθεί ένα βήμα το οποίο θα υπολογίζει τις συντεταγμένες των κέντρων των τμημάτων. Έτσι για κάθε αντικείμενο, ελέγχουμε σε ποιο τμήμα ανήκει και στη συνέχεια υπολογίζουμε τον μέσο όρο των τιμών των αντικειμένων του εν λόγω τμήματος. Οι συντεταγμένες

που θα βρούμε αντιστοιχούν στο κέντρο του τμήματος. Κατόπιν υπολογίζουμε την ευκλείδεια απόσταση του κάθε αντικειμένου από το κέντρο του τμήματος που ανήκει, όπως και στην προηγούμενη περίπτωση. Το τελικό κόστος είναι το άθροισμα των τετραγώνων των αποστάσεων. Ένας εναλλακτικός τρόπος υπολογισμού του εν λόγω κόστους είναι να εκμεταλλευτούμε την εξής ιδιότητα:

$$D_r = \sum_i \sum_j \|x_i - x_j\|^2 = 2 \cdot n_r \sum_i \|x_i - \bar{x}\|^2 \quad (42)$$

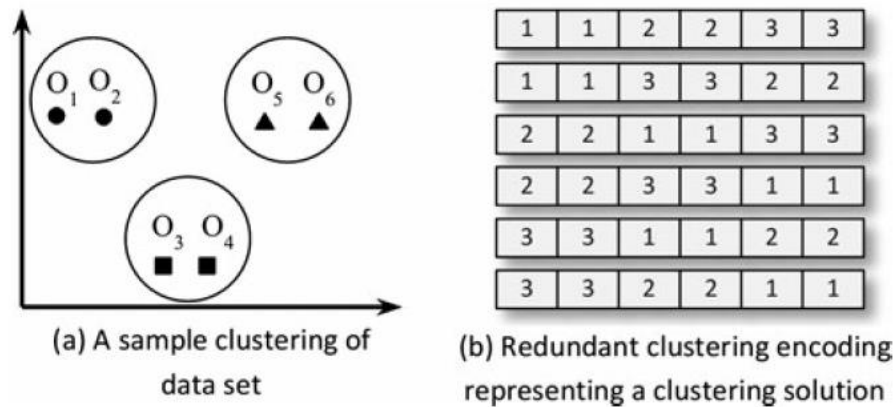
Όπου τα x_i, x_j είναι τα αντικείμενα που ανήκουν στο τμήμα r και n_r είναι το πλήθος τους. Κατά αυτόν τον τρόπο μπορούμε κατά την έναρξη του αλγορίθμου να υπολογίσουμε τον πίνακα αποστάσεων και να τον αποθηκεύσουμε, αποφεύγοντας έτσι δαπανηρούς υπολογισμούς κατά τη διάρκεια εκτέλεσης του εκάστοτε αλγορίθμου.

Αναπαράσταση Λύσης

Ένα άλλο πρόβλημα, όπως έχουμε ήδη αναφέρει είναι το εξής. Χρησιμοποιώντας τον συγκεκριμένο τρόπο αναπαράστασης, οι λύσεις $[\alpha \alpha \beta \beta]$ και $[\beta \beta \beta \alpha]$ είναι ακριβώς οι ίδιες, αφού πρακτικά δείχνουν ότι τα τρία πρώτα αντικείμενα ανήκουν στο ίδιο τμήμα και τα δύο τελευταία ανήκουν σε δεύτερο τμήμα. Όταν λοιπόν εφαρμόσουμε κάποιον τελεστή διασταύρωσης, όπως για παράδειγμα το νεογνό να κληρονομεί την τιμή της αντίστοιχης διάστασης από τον έναν από τους δύο γονείς μπορεί να καταλήξουμε στην εξής λύση $[\alpha \beta \alpha \beta]$, όπου κληρονομήθηκε το 1^ο, 3^ο και 4^ο στοιχείο από τον πρώτο γονέα και το 2^ο και το 5^ο από τον δεύτερο γονέα. Έτσι ενώ οι δύο γονείς είναι πανομοιότυποι, το νεογνό αντιπροσωπεύει μία εντελώς διαφορετική λύση.

Προκειμένου να αντιμετωπίσουμε αυτό το πρόβλημα, προτείνουμε την ακόλουθη τεχνική επαναρίθμησης των λύσεων [49]. Αρχικά επιλέγουμε μία από τις διαθέσιμες λύσεις του προβλήματος ως λύση αναφοράς. Κατόπιν εξετάζουμε την ομοιότητα μεταξύ της λύσεως αναφοράς με την λύση που θέλουμε να μετασχηματίσουμε. Η ομοιότητα του i -οστού τμήματος της πρώτης λύσης (αναφοράς) σε σχέση με το j -οστό τμήμα της δεύτερης λύσης (υπό μετασχηματισμό) ορίζεται ως ο αριθμός των κοινών αντικειμένων στα δύο αυτά τμήματα. Για παράδειγμα, έστω ότι στο 3^ο τμήμα της λύσης αναφοράς ανήκουν τα αντικείμενα $\{5, 13, 22, 23, 30, 55\}$ και στο 1^ο τμήμα της υπό μετασχηματισμό λύσης ανήκουν τα αντικείμενα $\{5, 6, 12, 22, 23, 31, 56\}$. Η ομοιότητα του τμήματος #3 της πρώτης λύσης με την ομοιότητα του τμήματος #1 της δεύτερης λύσης είναι ίση με 3, καθώς 3 είναι τα κοινά τους στοιχεία. Κατ'

αυτών τον τρόπο, σχηματίζουμε τον πίνακα ομοιότητας, διαστάσεων $K \times K$ όπου K ο αριθμός των τμημάτων.



Εικόνα 8 - Το πρόβλημα της πολλαπλής αναπαράστασης λύσεων

Στη συνέχεια, προκειμένου να βρούμε τις αντιστοιχίες, εξετάζουμε με φθίνουσα σειρά τις τιμές του πίνακα ομοιότητας. Έτσι ξεκινώντας από την μεγαλύτερη τιμή σχηματίζουμε τους κατάλληλους μετασχηματισμούς. Διακρίνονται οι εξής περιπτώσεις:

1. Η μεγαλύτερη τιμή είναι μία και βρίσκεται στη διαγώνιο του πίνακα. Αυτό σημαίνει ότι δεν απαιτείται κάποιος μετασχηματισμός για το συγκεκριμένο τμήμα, καθώς οι δύο λύσεις αναπαριστούν με τον ίδιο αριθμό το εν λόγω τμήμα.
2. Οι μεγαλύτερες τιμές είναι δύο και βρίσκονται σε συμμετρικές θέσεις στον πίνακα. Έτσι αντικαθιστούμε στην υπό μετασχηματισμό λύση το i_{max} με το j_{max} και αντίστροφα, όπου τα (i_{max}, j_{max}) και (j_{max}, i_{max}) είναι οι συντεταγμένες των σημείων όπου βρέθηκε η μέγιστη ομοιότητα.
3. Οι μεγαλύτερες τιμές είναι παραπάνω από δύο και βρίσκονται ανά δύο σε συμμετρικές θέσεις στον πίνακα. Σε αυτήν την περίπτωση επιλέγουμε ένα από τα ζευγάρια συμμετρικών θέσεων και εφαρμόζουμε ό,τι και στο 2.
4. Οι μεγαλύτερες τιμές είναι παραπάνω από δύο, αλλά δεν βρίσκονται ανά σε συμμετρικές θέσεις στον πίνακα. Σε αυτήν την περίπτωση, θα έχουμε ένα αριθμό ζευγαριών και κάποιες τιμές που θα περισσεύουν. Επιλέγουμε μία από τις τιμές που περισσεύουν και οι συντεταγμένες της μας δείχνουν τον μετασχηματισμό που πρέπει να εφαρμόσουμε.

Σε όλες τις περιπτώσεις, μετά από κάθε επιτυχή αντιστοίχιση 2 τμημάτων, διαγράφουμε τις αντίστοιχες γραμμές και στήλες από τον πίνακα ομοιότητας για να προχωρήσουμε ορθά με την αντιστοίχιση των

υπόλοιπων τμημάτων. Επαναλαμβάνουμε τη διαδικασία μέχρι αν αντιστοιχίσουμε όλα τα τμήματα της λύσης υπό μετασχηματισμό με κάποιο από τα τμήματα της λύσης αναφοράς.

7.4. Τελεστές αλγορίθμων

7.4.1. Τελεστές που δρουν στην πραγματική αναπαράσταση λύσης

Αλγόριθμος HBMO

Αρχικοποίηση

Στον συγκεκριμένο αλγόριθμο επιλέξαμε η αρχικοποίηση του πληθυσμού να γίνεται επιλέγοντας ως κέντρα των τμημάτων τυχαία αντικείμενα του σετ δεδομένων. Οι εναλλακτικές θα ήταν να εντοπίζουμε την μέγιστη και την ελάχιστη τιμή της κάθε διάστασης του σετ δεδομένων και να δημιουργούμε τυχαίες τιμές εντός του εκάστοτε διαστήματος.

Τελεστής διασταύρωσης

Ως τελεστή διασταύρωσης, χρησιμοποιούμε τον ακόλουθο: Έστω ότι ο γονότυπος το κηφήνα είναι $D_1 = [1.34 \ 0.67 \ 0.89]$ και της βασίλισσας $Q_1 = [1.1 \ 0.71 \ 0.95]$. Ορίζουμε έναν συντελεστή διασταύρωσης (crossover ratio - CR) και συγκρίνουμε το αποτέλεσμα μίας γεννήτριας τυχαίων αριθμών με τον συντελεστή αυτό. Εάν είναι μικρότερο, τότε το νεογνό κληρονομεί για τη συγκεκριμένη διάσταση την τιμή του κηφήνα, διαφορετικά της βασίλισσας. Εάν για παράδειγμα ορίσουμε το CR στο 0.30, τότε όταν το αποτέλεσμα της γεννήτριας είναι κάτω από 0.30 θα κληρονομείται η τιμή του κηφήνα ενώ όταν είναι πάνω από 0.30 της βασίλισσας. Για ευκολότερη κατανόηση θα ονομάσουμε αυτήν την τεχνική, «**Τεχνική Α**»

Μία παραλλαγή της παραπάνω τεχνικής προσαρμοσμένη στο εν λόγω πρόβλημα είναι αντί να κληρονομεί μόνο μία διάσταση το νεογνό, να κληρονομεί ολόκληρο το διάνυσμα του κέντρου του τμήματος. Θα ονομάσουμε αυτήν την τεχνική, «**Τεχνική Β**».

Ακόμη, ελέγχουμε και την επίδοση ενός τρίτου τελεστή διασταύρωσης, ο οποίος ουσιαστικά αποτελεί τον σταθμισμένο μέσο όρο των γονέων και ορίζεται με τον ακόλουθο τρόπο:

$$brood = queen + rand \cdot R \cdot (drone - queen) \quad (43)$$

Όπου το R είναι μία παράμετρος εντός του διαστήματος $(0,1)$ η οποία δείχνει το κατά πόσο επηρεάζεται το νεογνό από τον κηφήνα και από την βασίλισσα. Το $rand$ είναι ένας τυχαίος αριθμός εντός του διαστήματος $(0,1)$. Θα ονομάσουμε αυτήν την τεχνική «**Τεχνική Γ**».

Στη συνέχεια, πρέπει να εξειδικεύσουμε ποιες θα είναι οι εργάτριες του αλγορίθμου, οι διαδικασίες δηλαδή τοπικής αναζήτησης. Μία από τις πιο διαδεδομένες τεχνικές τοπικής αναζήτησης είναι η ακόλουθη, η οποία μοιάζει και με τον δεύτερο συντελεστή διασταύρωσης, όπως αυτός περιγράφηκε παραπάνω.

$$brood = brood \pm \beta \cdot (brood - queen) \quad (44)$$

Θα ονομάσουμε αυτήν την εργάτρια, «**Εργάτρια Α**»

Μία δεύτερη μέθοδος είναι η ακόλουθη:

$$brood = brood \pm (\delta + \varepsilon) \cdot brood \quad (45)$$

Όπου οι τιμές του $\varepsilon \in (0,1)$ είναι προκαθορισμένες, ενώ οι τιμές του δ προκύπτουν τυχαία κάθε φορά που εφαρμόζεται η μέθοδος. Η εργάτρια αυτή ονομάζεται «**Εργάτρια Β**». Πρέπει να σημειωθεί σε αυτό το σημείο ότι οι εργάτριες Α και Β εφαρμόζονται μόνο σε μία τιμή του διανύσματος λύσης.

Αλγόριθμος BBMO

Στον αλγόριθμο BBMO χρησιμοποιούμε την ίδια διαδικασία αρχικοποίησης όπως και στον HBMO και χρησιμοποιούμε τους ίδιους τελεστές διασταύρωσης με τον HBMO (Α,Β και Γ), αλλά για την διαδικασία τοπικής αναζήτησης χρησιμοποιούμε αυτή που περιγράφηκε στην εξίσωση (38) σε δύο παραλλαγές. Στην πρώτη (εργάτρια Α) αλλάζουμε μόνο μία τιμή του διανύσματος, ενώ στη δεύτερη (εργάτρια Β) αλλάζουμε όλο το κέντρο της εν λόγω λύσης.

Αλγόριθμος ABC

Στον αλγόριθμο ABC χρησιμοποιούμε τις τεχνικές που προτείνονται από τους συγγραφείς χωρίς να κάνουμε κάποια μεταβολή.

Τελεστές που λειτουργούν στην ακέραια αναπαράσταση λύσης

Αλγόριθμος HBMO

Οι τελεστές διασταύρωσης καθώς και οι εργάτριες που χρησιμοποιήσαμε στην προηγούμενη εφαρμογή των αλγορίθμων, δεν είναι κατάλληλοι για την ακέραια αναπαράσταση της λύσης. Αυτό συμβαίνει γιατί με βάση τον τρόπο υπολογισμού τους, μπορεί να οδηγήσουν σε μη ακέραια αποτελέσματα, πράγμα που είναι ανούσιο στην εν λόγω αναπαράσταση. Για παράδειγμα δεν γίνεται ένα αντικείμενο να ανήκει στο τμήμα 2,5. Για αυτόν το λόγο εφαρμόζουμε τους παρακάτω τρόπους αρχικοποίησης, τελεστές διασταύρωσης και εργάτριες:

Αρχικοποίηση

Προκειμένου να αποφύγουμε την ύπαρξη άδειων τμημάτων δημιουργούμε τα αρχικά διανύσματα λύσεων με τον ακόλουθο τρόπο. Αρχικά ορίζουμε έναν ελάχιστο αριθμό p αντικειμένων ανά τμήμα ο οποίος είναι ίσος με τον μεγαλύτερο ακέραιο αριθμό που είναι μικρότερος όμως από το πλήθος των αντικειμένων διά τον αριθμό των τμημάτων. Στη συνέχεια αναθέτουμε σε κάθε τμήμα p αντικείμενα με τυχαίο τρόπο και εάν μετά το τέλος αυτής της διαδικασίας υπάρχουν ακόμη αντικείμενα που δεν ανήκουν σε κάποιο τμήμα, τα αναθέτουμε και αυτά με τυχαίο τρόπο σε κάποιο.

Τελεστής διασταύρωσης

Ως τελεστή διασταύρωσης χρησιμοποιούμε την «Τεχνική Α» όπως περιγράφηκε παραπάνω, μιας και είναι κατάλληλη και για ακέραια προβλήματα.

Εργάτριες

Ως εργάτριες, χρησιμοποιούμε την «Εργάτρια Δ» καθώς και τους τελεστές μετάλλαξης και ΚΜΟ, όπως περιγράφονται σε επόμενη παράγραφο. Η «εργάτρια Δ» λειτουργεί πάνω στην λύση επιλέγοντας τυχαία ένα αντικείμενο και κατατάσσοντάς το σε ένα τυχαίο τμήμα, διαφορετικό από το αρχικό του.

7.4.1.1. Επιλογή

Ο τελεστής επιλογής χρησιμοποιείται στα κομμάτια των αλγορίθμων όπου απαιτείται η επιλογή ενός μέλους του πληθυσμού λύσεων. Σκοπός του είναι να λειτουργεί με τέτοιο τόπο ώστε να κάνει πιο πιθανή την επιλογή μίας καλής λύσης, χωρίς όμως να αποκλείει τις υπόλοιπες λύσεις καθώς και αυτές θα μπορούσαν εν δυνάμει να οδηγήσουν σε μία καλύτερη λύση καθώς ο αλγόριθμος εξελίσσεται. Στην παρούσα εργασία θα μελετήσουμε τον τελεστή που προτείνεται από τους Krishna et al. [50]. Αρχικά ορίζουμε μία συνάρτηση f τέτοια ώστε $f(s_i) = -S(s_i)$ και μία συνάρτηση $g(s_i) = f(s_i) - (\bar{f} - c \cdot \sigma)$, όπου S είναι το συνολικό τετραγωνικό σφάλμα, \bar{f} είναι η μέση τιμή της συνάρτησης f για όλον τον πληθυσμό, σ είναι η αντίστοιχη τυπική απόκλιση και c είναι μία σταθερά που παίρνει τιμές εντός του διαστήματος (1,3). Τέλος ορίζουμε τη συνάρτηση F με τον ακόλουθο τύπο:

$$F(s_i) = \begin{cases} g(s_i), & \text{if } g(s_i) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (46)$$

Έτσι, λύσεις οι οποίες έχουν σχετικά μικρό τετραγωνικό σφάλμα θα έχουν σχετικά μεγάλη τιμή της συνάρτησης F . Με βάση αυτήν την παρατήρηση ορίζουμε την πιθανότητα επιλογής του συγκεκριμένου μέλους του πληθυσμού ως εξής:

$$P(s_i) = \frac{F(s_i)}{\sum_{j=1}^N F(s_j)} \quad (47)$$

Οι συγκεκριμένες πιθανότητες ορίζουν μία διακριτή κατανομή πιθανότητας, με βάση την οποία επιλέγεται κάποιο μέλος του πληθυσμού.

7.4.1.2. Μετάλλαξη

Ο τελεστής της μετάλλαξης ανακατατάσσει κάποιο από τα δεδομένα σε άλλο τμήμα, με έναν πιθανοτικό κανόνα ο οποίος βασίζεται στις αποστάσεις μεταξύ του συγκεκριμένου αντικειμένου από τα κέντρα των τμημάτων. Με άλλα λόγια, η πιθανότητα κατάταξης του κάθε αντικειμένου σε ένα συγκεκριμένο τμήμα είναι μεγάλη εάν αυτό βρίσκεται κοντά στο κέντρο του τμήματος, ενώ είναι μικρή όταν η απόσταση είναι μεγάλη. Ορίζουμε με $d_j = d(x_i, c_j)$ την ευκλείδεια απόσταση του σημείου i από το κέντρο του τμήματος j . Ορίζουμε επίσης την παρακάτω κατανομή πιθανότητας:

$$p_j = \Pr\{s_i \in j\} = \frac{c_m d_{\max} - d_j}{\sum_{i=1}^K c_m d_{\max} - d_i} \quad (48)$$

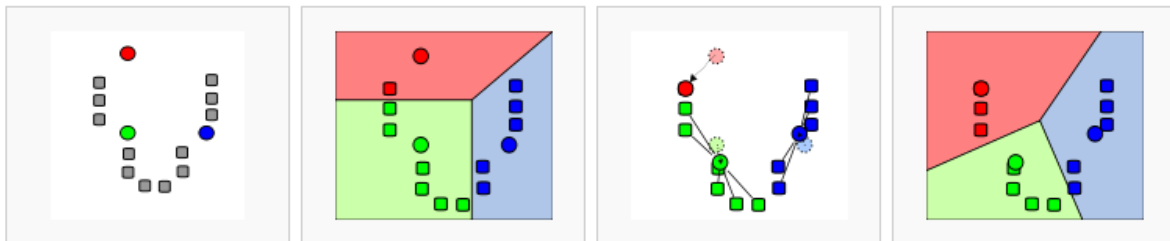
Όπου c_m είναι μία σταθερά ≥ 1 και $d_m = \max\{d_j\}$. Ο συγκεκριμένος τελεστής μπορεί να οδηγήσει σε λύσεις οι οποίες αποτελούνται από άδεια τμήματα. Αυτό συμβαίνει όταν πριν τη μετάλλαξη τουλάχιστον ένα από τα τμήματα περιέχει μονάχα ένα αντικείμενο. Αυτό προφανώς αποτελεί μία μη επιθυμητή κατάσταση και για να το αντιμετωπίσουμε εφαρμόζουμε το συγκεκριμένο τελεστή μετάλλαξης μόνο όταν $d_{s_i} > 0$, μόνο δηλαδή όταν το αντικείμενο δεν είναι ταυτόχρονα και κέντρο του τμήματος. Τέλος πρέπει να αναφερθεί ότι ο συγκεκριμένος τελεστής εφαρμόζεται με πιθανότητα P_m η οποία είναι προκαθορισμένη και λειτουργεί ως ποσοστό μετάλλαξης.

7.4.1.3. Τελεστής K-means

Με την εφαρμογή των παραπάνω τελεστών επιλογής και μετάλλαξης, υπάρχει η πιθανότητα να καθυστερήσει η σύγκλιση των αλγορίθμων μιας και βασίζονται σε πιθανοτικούς κανόνες και ο αρχικός πληθυσμός επιλέγεται τυχαία. Επιπρόσθετα, η πιθανότητα μετάλλαξης πρέπει να λαμβάνει σχετικά μικρές τιμές γιατί διαφορετικά οι αλγόριθμοι μπορεί να παρουσιάσουν ταλαντώσεις. Για να βελτιωθεί λοιπόν η επίδοση των αλγορίθμων έχει προταθεί η εφαρμογή μίας επανάληψης του αλγορίθμου K-means στην λύση. Ο αντίστοιχος τελεστής ονομάστηκε KMO (K-means operator). Δοσμένης μίας λύσης s_i ο KMO εφαρμόζεται ως εξής:

1. Υπολογισμός των κέντρων των τμημάτων με βάση τον τύπο (24).
2. Επανακατανομή των αντικειμένων στο τμήμα από το κέντρο του οποίου η απόσταση είναι ελάχιστη.

Παρά την απλότητα του τελεστή, υπάρχουν και κάποιες ανεπιθύμητες παρενέργειες. Για παράδειγμα, η εφαρμογή του τελεστή μπορεί να οδηγήσει σε άδεια τμήματα. Αυτό το πρόβλημα είναι γνωστό και μία από τις πιθανές λύσεις είναι να δημιουργήσουμε ένα τμήμα με ένα μόνο στοιχείο, το οποίο το αφαιρούμε από το τμήμα στο οποίο ανήκει. Η επιλογή του γίνεται με βάση την απόσταση από το κέντρο του τμήματος, οπότε προφανώς επιλέγεται το στοιχείο του οποίου η απόσταση από το αντίστοιχο κέντρο είναι η μεγαλύτερη.

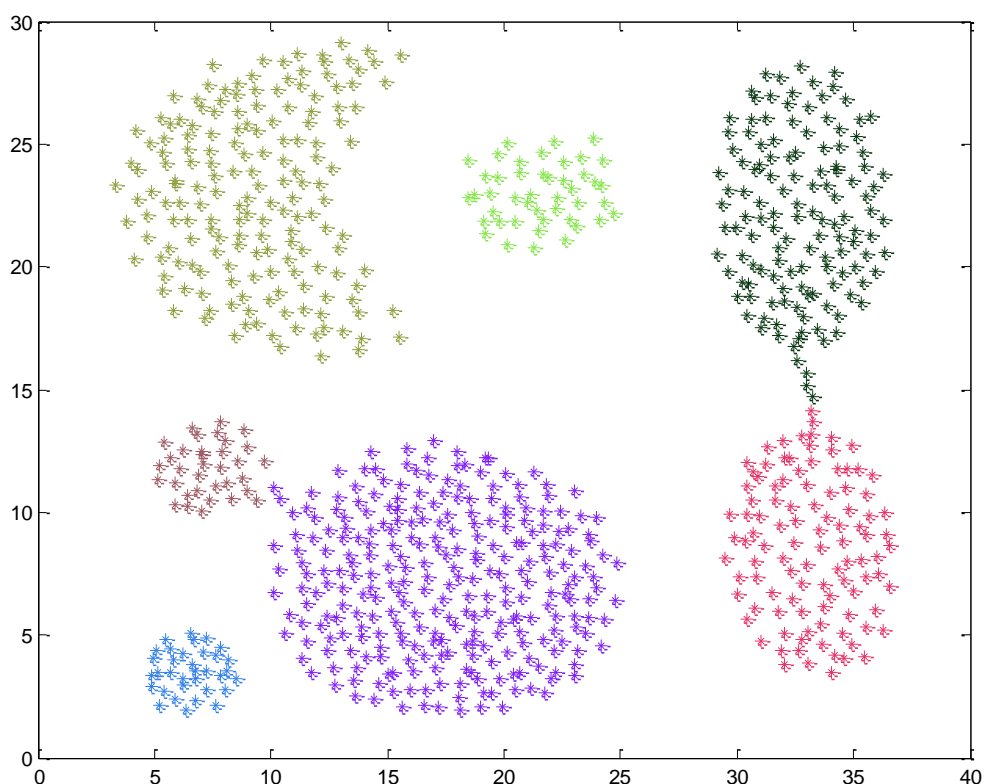


Εικόνα 9 – Γραφική αναπαράσταση μίας επανάληψης του αλγορίθμου K-means

8. Ρύθμιση των παραμέτρων των αλγορίθμων

8.1. Σετ δεδομένων

Προκειμένου να προχωρήσουμε στη ρύθμιση των παραμέτρων των αλγορίθμων για κάθε αναπαράσταση λύσης και κάθε διαφορετικό συνδυασμό τελεστών θα χρησιμοποιήσουμε το σετ δεδομένων “Aggregation” το οποίο παρατίθεται στο παράρτημα. Αποτελείται από 788 σημεία δύο διαστάσεων και ο λόγος που χρησιμοποιήθηκε είναι η δυνατότητα που μας δίνει να το απεικονίσουμε. Παρακάτω βλέπουμε το σετ αυτό, με χρωματισμένα τα τμήματα αναφοράς.



Εικόνα 10 - Το σετ δεδομένων Aggregation ταξινομημένο σωστά

Το κριτήριο με βάση το οποίο θα ομαδοποιήσουμε τα δεδομένα είναι το κριτήριο του τετραγώνου της απόστασης. Ο λόγος που επιλέχθηκε αυτό το κριτήριο είναι η απλότητά του, μέσω της οποίας είναι ευκολότερο να αναδειχθούν οι επιδόσεις των αλγορίθμων.

Για κάθε αλγόριθμο, θα δείξουμε την επίδραση της κάθε παραμέτρου στα τελικά αποτελέσματα καθώς και το υπολογιστικό φορτίο του, το οποίο εκφράζεται αφενός από τον χρόνο εκτέλεσης και αφετέρου από τον αριθμό κλήσεων της αντικειμενικής συνάρτησης.

8.2. Συνεχής Αναπαράσταση

Αλγόριθμος HBMO

Προκειμένου να γίνει σωστή ρύθμιση των παραμέτρων του προβλήματος, δοκιμάζουμε διάφορους συνδυασμούς. Για κάθε συνδυασμό, εκτελούμε 10 επαναλήψεις έτσι ώστε να βεβαιωθούμε ότι τυχόν διαφορές στην επίδοση δεν οφείλονται σε στοχαστικούς παράγοντες. Στον παρακάτω πίνακα παραθέτουμε τη μέση τιμή των αποτελεσμάτων για τα 10 τρεξίματα για κάθε συνδυασμό παραμέτρων.

Αποτελέσματα HBMO (Τεχνική Α – Εργάτρια Α)

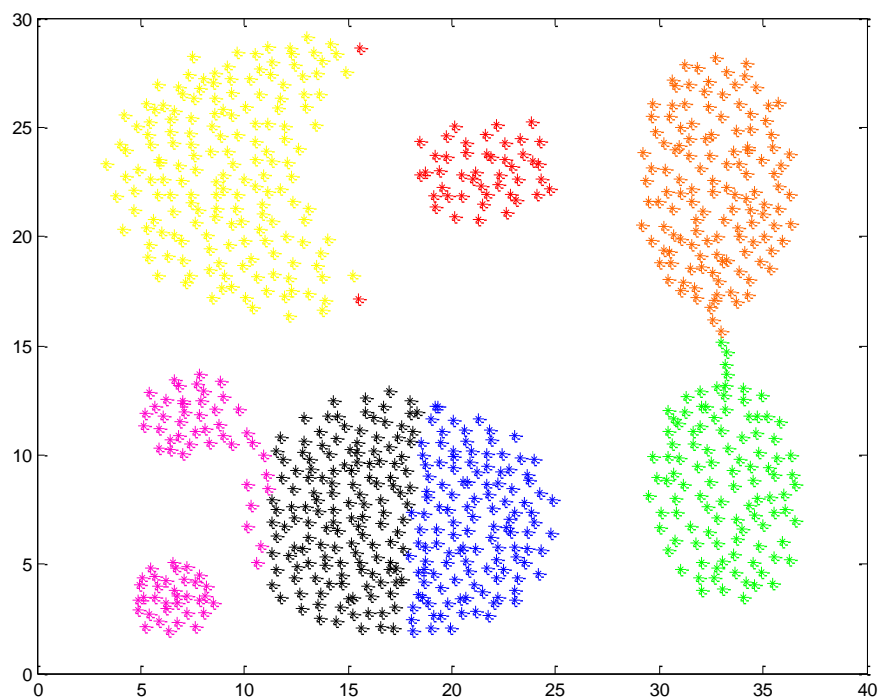
	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]				Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	25	1	50		12481,2	37,7	81176,1	676,7	0,1	20,2
2	25	1	150		11142,4	87,8	191188,2	666,9	0,2	72,0
3	25	1	300		11155,2	141,1	313765,8	638,8	0,3	135,5
4	25	3	50		11084,7	98,0	218631,3	539	0,2	85,5
5	25	3	150		11022,7	294,7	660133,8	520,3	0,6	237,5
6	25	3	300		11043,5	579,9	1305530,7	484	1,2	469,8
7	25	5	50		11019,1	165,6	369581,1	545,2	0,3	151,9
8	25	5	150		10997,9	503,1	1129305,9	496,2	1,0	484,6
9	25	5	300		10997,6	1023,6	2224423,5	269,2	2,0	979,0
10	50	1	50		11661,7	38,3	81583,5	750,3	0,1	29,4
11	50	1	150		11232,9	80,2	173543,4	684,9	0,16	73,40
12	50	1	300		11128,9	117,7	263712,9	689,8	0,24	110,99
13	50	3	50		11126,0	83,4	184460,7	634,8	0,17	81,31
14	50	3	150		10998,0	266,9	598230,9	633,2	0,53	236,90
15	50	3	300		10997,2	516,5	1161656,1	608,4	1,03	484,19
16	50	5	50		11015,9	137,8	308237,7	563,3	0,28	131,90
17	50	5	150		10997,3	486,6	1096259,1	468,7	0,97	453,02
18	50	5	300		10996,8	993,5	2156762,1	476,3	1,99	896,34
19	75	1	50		11660,1	39,4	86902,2	760,6	0,08	32,11
20	75	1	150		11150,7	73,4	163191,9	749,9	0,15	71,14
21	75	1	300		11172,5	175,8	262422,0	722,4	0,35	164,98
22	75	3	50		11063,0	110,6	1755657,0	754,3	0,22	106,76
23	75	3	150		11026,3	315,2	5492385,0	701	0,63	291,07
24	75	3	300		11047,0	631,1	1050471,0	740,6	1,26	618,24
25	75	5	50		11054,4	117,6	252234,0	658,2	0,2	114,9
26	75	5	150		10997,0	475,6	1035851,1	514,2	1,0	448,5
27	75	5	300		11021,1	929,4	2084134,2	510,8	1,9	880,7

Πίνακας 1 – Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Α-Α

Στις στήλες του Πίνακας 1 βλέπουμε την μέση βέλτιστη λύση, τον μέσο χρόνο ανά εκτέλεση του αλγορίθμου, τον μέσο αριθμό κλήσεων της αντικειμενικής συνάρτησης, την μέση επανάληψη όπου εντοπίστηκε η βέλτιστη λύση, ο μέσος χρόνος ανά επανάληψη και τέλος ο μέσος χρόνος μέχρι τον εντοπισμό της τελικής λύσης (λύση στην οποία συγκλίνει ο αλγόριθμος). Επιλέγουμε τον 14^ο συνδυασμό παραμέτρων, καθώς αποδείχτηκε ότι δίνει καλύτερη μέση τιμή αποτελεσμάτων στον μικρότερο δυνατό χρόνο. Στο Γράφημα 1 βλέπουμε το πώς μεταβάλλεται η ποιότητα της λύσης με τον συνολικό υπολογιστικό χρόνο (δηλαδή την πρώτη με την τελευταία στήλη του πίνακα). Παρατηρούμε ότι οι συνδυασμοί παραμέτρων που απαιτούν πάνω από 230 δευτερόλεπτα δεν βελτιώνουν περαιτέρω τη λύση. Στην Εικόνα 11 βλέπουμε την λύση που προκύπτει από την βέλτιστη επιλογή παραμέτρων.



Γράφημα 1 -Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Α-Α



Εικόνα 11 – Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος HBMO περ.Α-Α

Αποτελέσματα HBMO (Τεχνική Β – Εργάτρια Α)

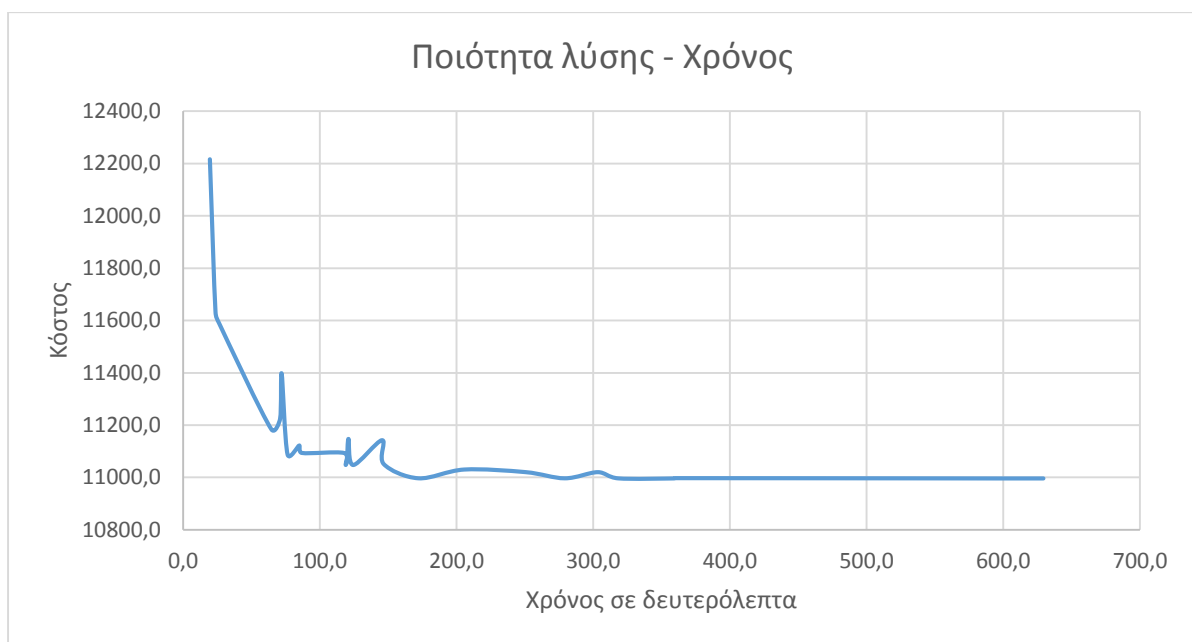
Αντικαθιστούμε τον τελεστή διασταύρωσης και εφαρμόζουμε την «Τεχνική Β» (βλ. κεφ. 7). Εκτελούμε εκ νέου τις απαραίτητες προσομοιώσεις και παραθέτουμε τα αποτελέσματα στον παρακάτω πίνακα.

	Συνδυασμός παραμέτρων [πληθυσμός, βασίλισσες, επαναλήψεις τοπικής αναζήτησης]				Βέλτιστη λύση	Χρόνος	Evaluation	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	25	1	50		12216,7	38,4	81879,9	254,0	0,1	19,5
2	25	1	150		11397,1	96,1	211482,6	374,8	0,2	72,0
3	25	1	300		11142,9	166,7	377427,3	436,4	0,3	145,5
4	25	3	50		11088,2	101,2	227918,4	377,3	0,2	76,4
5	25	3	150		11051,7	290,6	666339,9	252,5	0,6	146,8
6	25	3	300		11020,8	573,4	1310768,1	264,6	1,1	303,5
7	25	5	50		11061,8	169,1	384559,8	354,2	0,3	119,8
8	25	5	150		10997,3	497,1	1128581,1	280,9	1,0	279,3
9	25	5	300		10997,2	961,5	2228125,8	275,0	1,9	528,8
10	50	1	50		11590,1	36,5	84459,9	359,5	0,1	26,3
11	50	1	150		11224,0	86,3	189081,3	410,5	0,17	70,89
12	50	1	300		11093,9	133,5	302000,1	441,1	0,27	117,76
13	50	3	50		11122,8	95,3	216743,7	445,3	0,19	84,90
14	50	3	150		10997,4	270,5	616713,3	318,0	0,54	172,04
15	50	3	300		10996,8	535,9	1227665,4	335,4	1,07	359,48

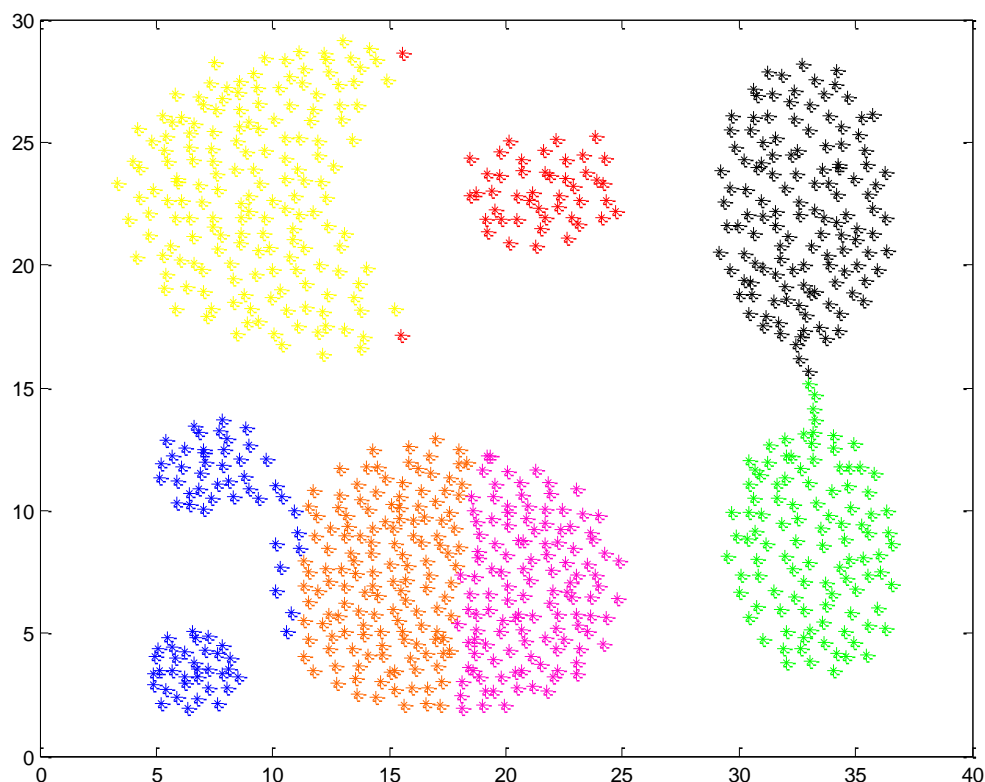
16	50	5	50		11048,2	152,7	346564,2	407,0	0,31	124,34
17	50	5	150		11020,9	483,7	1099203,6	258,9	0,97	250,45
18	50	5	300		10996,8	937,3	2159109,9	317,7	1,87	595,57
19	75	1	50		11633,6	39,7	91431,0	297,5	0,08	23,64
20	75	1	150		11184,7	77,6	180632,4	415,7	0,16	64,49
21	75	1	300		11148,0	130,7	290685,9	462,8	0,26	120,98
22	75	3	50		11094,3	92,7	209898,9	468,3	0,19	86,83
23	75	3	150		11030,8	243,1	554130,9	423,1	0,49	205,72
24	75	3	300		10997,9	474,5	1085055,0	379,1	0,95	359,73
25	75	5	50		11048,7	141,5	321191,1	419,7	0,3	118,8
26	75	5	150		10997,1	457,1	1046224,8	348,2	0,9	318,3
27	75	5	300		10996,8	920,5	2111314,5	341,9	1,8	629,4

Πίνακας 2 – Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Β-Α

Και σε αυτήν την περίπτωση επιλέγουμε τον 14^ο συνδυασμό παραμέτρων, παρόλο που η μέση τιμή της λύσης είναι κατά 0,005% χειρότερη από την καλύτερη που βρέθηκε με τον 15^ο συνδυασμό παραμέτρων, αλλά ο χρόνος είναι αισθητά μικρότερος (λιγότερο από μισός).



Γράφημα 2 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Β-Α



Εικόνα 12 - Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος HBMO περ.Β-Α

Αποτελέσματα HBMO (Τεχνική Γ – Εργατήρια Α)

Αντικαθιστούμε τον τελεστή διασταύρωσης και εφαρμόζουμε την «Τεχνική Γ» (βλ. κεφ. 7). Εκτελούμε εκ νέου τις απαραίτητες προσομοιώσεις και παραθέτουμε τα αποτελέσματα στον παρακάτω πίνακα.

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]			Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	25	1	50	11204,3	32,2	46460,4	483,4	0,0	19,5
2	25	1	150	11097,1	152,4	187518,9	407,0	0,2	77,5
3	25	1	300	11023,3	290,2	363611,4	329,7	0,4	119,6
4	25	3	50	11020,1	156,3	184374,6	414,8	0,2	81,1
5	25	3	150	11020,8	576,0	662806,5	289,3	0,7	208,3
6	25	3	300	10996,8	1008,6	1309955,4	331,2	1,3	417,6
7	25	5	50	11050,0	266,2	317071,5	344,5	0,3	114,7
8	25	5	150	10996,8	988,9	1123099,8	364,8	1,2	450,9
9	25	5	300	10996,8	1375,6	2230925,1	287,7	1,7	494,7
10	50	1	50	11245,9	34,4	58878,3	486,3	0,0	20,9
11	50	1	150	11117,3	130,8	169647,6	435,2	0,2	71,2
12	50	1	300	10996,8	261,3	326200,5	397,0	0,3	129,7

13	50	3	50		11050,3	125,9	152024,7	474,7	0,2	74,7
14	50	3	150		10996,8	512,5	603032,7	427,6	0,6	273,9
15	50	3	300		10996,8	992,1	1263333,9	326,1	1,2	404,4
16	50	5	50		11020,8	240,7	305346,0	423,4	0,3	127,4
17	50	5	150		10996,8	943,5	1085613,6	435,3	1,2	513,4
18	50	5	300		10996,8	914,5	2166785,4	345,4	1,1	394,9
19	75	1	50		11303,3	33,8	74754,0	479,7	0,0	20,2
20	75	1	150		11023,3	73,6	172614,3	441	0,1	40,6
21	75	1	300		11046,6	134,7	305856,3	417,5	0,2	70,3
22	75	3	50		10998,3	64,5	142441,2	470,1	0,1	37,9
23	75	3	150		10996,8	260,5	570529,5	398,1	0,3	129,7
24	75	3	300		10996,8	539,3	1187455,2	308,6	0,7	208,0
25	75	5	50		10997,1	111,1	249143,4	461,5	0,1	64,1
26	75	5	150		10996,8	482,8	1064526	373,6	0,6	225,5
27	75	5	300		10996,8	960,3	2128561,8	406,1	1,2	487,5

Πίνακας 3 - Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Γ-Α

Επιλέγουμε τον 23^ο συνδυασμό παραμέτρων μιας και δίνει την βέλτιστη λύση σε σχετικά μικρό χρονικό διάστημα. Η τμηματοποίηση των δεδομένων είναι ίδια με τις προηγούμενες περιπτώσεις και για εξοικονόμηση χώρου παραλείπεται.



Γράφημα 3 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Γ-Α

Στη συνέχεια εξετάζουμε την επίδραση των εργατριών στην επίδοση του αλγορίθμου. Αρχικά εξετάζουμε την «Εργάτρια Β» με τον ίδιο τρόπο όπως και πριν χρησιμοποιώντας την «τεχνική Β» ως τελεστή διασταύρωσης.

Αποτελέσματα ΗΒΜΟ (Τεχνική Β – Εργάτρια Β)

	Συνδυασμός παραμέτρων [πληθυσμός, βασίλισσες, επαναλήψεις τοπικής αναζήτησης]			Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	25	1	50	12058,2	37,6	79661,4	167,2	0,1	12,6
2	25	1	150	11869,3	98,0	212841,6	226,8	0,2	44,4
3	25	1	300	11653,3	196,9	422396,7	175,1	0,4	69,0
4	25	3	50	11346,3	108,2	230978,4	170,7	0,2	37,0
5	25	3	150	11354,7	301,9	674720,4	146,7	0,6	88,6
6	25	3	300	11232,1	604,1	1340205,9	145,5	1,2	175,8
7	25	5	50	11237,9	176,7	386487,6	263,2	0,4	93,0
8	25	5	150	11191,5	511,2	1132069,2	117,1	1,0	119,7
9	25	5	300	11138,1	993,3	2248082,1	115,5	2,0	229,5
10	50	1	50	11859,0	36,7	85179,0	195,9	0,1	14,4
11	50	1	150	11806,0	107,9	203894,4	315,8	0,22	68,18
12	50	1	300	11406,3	198,1	382908,9	275,2	0,40	109,04
13	50	3	50	11490,4	120,5	226780,5	249,0	0,24	59,99
14	50	3	150	11197,2	347,9	654674,7	214,0	0,70	148,89
15	50	3	300	11144,1	656,5	1287173,1	178,5	1,31	234,36
16	50	5	50	11207,7	196,3	374196,0	322,3	0,39	126,51
17	50	5	150	11127,4	573,2	1113065,4	176,2	1,15	202,01
18	50	5	300	11062,6	1124,9	2218617,6	228,2	2,25	513,41
19	75	1	50	11856,9	44,8	93358,8	346,9	0,09	31,11
20	75	1	150	11591,0	93,7	196759,2	356,6	0,19	66,81
21	75	1	300	11380,6	189,9	364009,5	306,8	0,38	116,54
22	75	3	50	11410,7	117,8	224479,8	338,4	0,24	79,76
23	75	3	150	11216,1	333,0	629917,8	262,2	0,67	174,63
24	75	3	300	11176,6	621,7	1217344,5	272,8	1,24	339,22
25	75	5	50	11289,0	198,0	374940,0	287,0	0,4	113,6
26	75	5	150	11122,3	566,2	1089350,4	188,2	1,1	213,1
27	75	5	300	11074,6	1111,5	2170370,7	237,3	2,2	527,5

Πίνακας 4 - Αποτελέσματα διαδικασίας ρύθμισης ΗΒΜΟ περ.Β-Β



Γράφημα 4 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Β-Β

Με χρήση της συγκεκριμένης εργάτριας παρατηρούμε ότι η λύση είναι αρκετά μακριά από τη βέλτιστη λύση των προηγούμενων περιπτώσεων, οπότε καταλήγουμε ότι ο συγκεκριμένος τελεστής δεν λειτουργεί αποδοτικά στην περίπτωσή μας.

Αποτελέσματα HBMO (Τεχνική Γ– Εργάτρια Β)

Στη συνέχεια εξετάζουμε την επίδραση των εργατριών στην επίδοση του αλγορίθμου. Τώρα, εξετάζουμε την «Εργάτρια Β» με τον ίδιο τρόπο όπως και πριν χρησιμοποιώντας την «τεχνική Γ» ως τελεστή διασταύρωσης.

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]			Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	25	1	50	11367,5	29,1	52886,4	495,6	0,1	28,8
2	25	1	150	11212,0	78,8	141494,1	494,7	0,2	77,9
3	25	1	300	11187,1	151,5	272047,2	495,6	0,3	150,2
4	25	3	50	11246,6	60,5	101984,1	471,0	0,1	57,0
5	25	3	150	11107,1	284,1	526589,4	486,6	0,6	276,4
6	25	3	300	11022,1	654,7	1182000,3	484,9	1,3	634,9
7	25	5	50	11306,2	75,8	117238,2	419,0	0,2	63,5
8	25	5	150	11020,5	520,9	936509,1	490,3	1,0	510,8
9	25	5	300	11022,9	913,0	2064773,1	488,0	1,8	891,1
10	50	1	50	11247,5	34,8	60989,7	495,4	0,1	34,5
11	50	1	150	11096,0	82,1	147224,1	496,8	0,2	81,5
12	50	1	300	11153,8	165,6	294414,9	497,0	0,3	164,7

13	50	3	50		11193,9	69,9	124959,0	465,8	0,1	65,1
14	50	3	150		11145,7	231,4	425955,0	478,6	0,5	221,5
15	50	3	300		11046,2	597,7	1072168,8	490,2	1,2	586,0
16	50	5	50		11323,7	81,7	141865,5	442,8	0,2	72,4
17	50	5	150		11076,9	434,0	755286	485,0	0,9	421,0
18	50	5	300		11038,0	823,4	1863558,0	489,6	1,6	806,3
19	75	1	50		11252,1	40,9	73224,0	495,3	0,1	40,6
20	75	1	150		11142,6	83,3	149828,4	497,2	0,2	82,8
21	75	1	300		11223,4	161,4	287254,5	497,3	0,3	160,5
22	75	3	50		11243,9	71,7	129696,3	475,4	0,1	68,2
23	75	3	150		11128,8	226,6	416962,5	496,6	0,5	225,1
24	75	3	300		11086,0	500,5	893889,9	483,7	1,0	484,2
25	75	5	50		11334,4	86,5	154512,9	467,5	0,2	80,9
26	75	5	150		11117,0	348,5	590597,4	467,8	0,7	326,0
27	75	5	300		11030,3	745,3	1681305,9	477,4	1,5	711,6

Πίνακας 5 Αποτελέσματα διαδικασίας ρύθμισης HBMO περ.Γ-Β



Γράφημα 5 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO περ.Γ-Β

Αλγόριθμος ABC

Οι παράμετροι που καθορίζει ο χρήστης στον αλγόριθμο τεχνητής αποικίας μελισσών είναι το μέγεθος του πληθυσμού (τοποθεσιών τροφής) και ο αριθμός μέγιστων προσπαθειών για βελτίωση μίας λύσης. Προκειμένου να εντοπίσουμε τις βέλτιστες τιμές παραμέτρων, εκτελέστηκαν 10 προσομοιώσεις για 9 διαφορετικούς συνδυασμούς παραμέτρων. Το μέγεθος του πληθυσμού λάμβανε τιμές από το σύνολο {10,40,60} ενώ ο αριθμός μέγιστων προσπαθειών από το σύνολο {50,100,150}. Ο αριθμός επαναλήψεων ήταν σταθερός στις 2500. Για κάθε προσομοίωση καταγράφουμε το χρόνο εκτέλεσης, τον

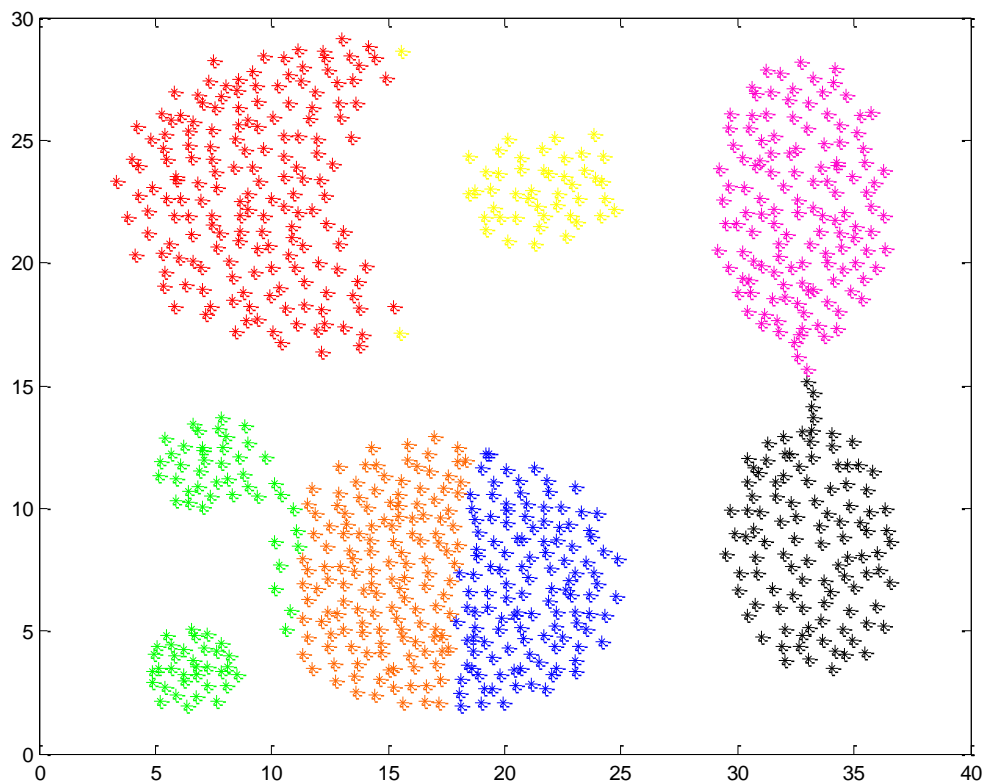
αριθμό φορών που κλήθηκε η συνάρτηση κόστους, την επανάληψη στην οποία βρέθηκε η βέλτιστη τιμή και τέλος την βέλτιστη λύση.

Στον παρακάτω πίνακα παραθέτουμε τη μέση τιμή των αποτελεσμάτων για τα 10 τρεξίματα για κάθε συνδυασμό παραμέτρων.

Συνδυασμός παραμέτρων [πληθυσμός, όριο επαναλήψεων χωρίς βελτίωση]	Βέλτιστη λύση	Χρόνος	Evaluation	Επανάληψη
[20, 50]	10997,9	22,74	50069,1	1146,3
[20, 100]	10997,1	21,98	50032,0	1875,4
[20, 150]	10996,8	21,74	50021,2	2087,8
[40, 50]	10997,9	43,94	100142,9	1465,0
[40, 100]	10996,8	43,45	100069,5	1564,3
[40, 150]	10996,7	43,87	100048,1	2109,5
[60, 50]	10997,9	66,38	150215,7	1440,2
[60, 100]	10996,8	66,42	150108,8	1794,1
[60, 150]	10996,8	66,02	150071,8	1864,3

Πίνακας 6 - Αποτελέσματα διαδικασίας ρύθμισης ABC

Επιλέγουμε τον τρίτο συνδυασμό παραμέτρων, δηλαδή το [20,150] καθώς είναι αφενός ο πιο γρήγορος και αφετέρου, έδωσε την μέση τιμή αποτελεσμάτων που είναι μόνο 0,0005% χειρότερη από τη βέλτιστη. Επίσης παρατηρούμε ότι κατά μέσο όρο η βέλτιστη λύση προκύπτει στην επανάληψη 2088, οπότε σαν μέγιστο αριθμό επαναλήψεων επιλέγουμε τις 2500. Η λύση που προκύπτει, οπτικά μπορεί να αναπαρασταθεί στην Εικόνα 13.



Εικόνα 13 - Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος ABC

Όπως παρατηρούμε, το κριτήριο που έχουμε επιλέξει στην συνάρτηση κόστους πιθανότατα δεν είναι κατάλληλο για να περιγράψει τις ομοιότητες και τις διαφορές των δεδομένων. Επίσης, παρατηρούμε ότι το σχήμα των τμημάτων είναι σχεδόν σφαιρικό, πράγμα που είναι γνωστή ιδιότητα του συγκεκριμένου κριτηρίου. Παρόλα αυτά, η επίδοση του αλγορίθμου είναι επαρκής, καθώς κατάφερε να ελαχιστοποιήσει τη συνάρτηση κόστους με ταχύτητα και στατιστική αξιοπιστία.

Αλγόριθμος Bumble Bees

Αποτελέσματα BB (Τεχνική Α – Εργασία Α)

Σε αυτήν την περίπτωση χρησιμοποιούμε ως τελεστή διασταύρωσης την «Τεχνική Α» ενώ στην φάση της τοπικής αναζήτησης αλλάζουμε κάθε φορά μόνο μία από τις διαστάσεις του διανύσματος λύσης.

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]				Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	30	1	20		11623,0	20,0	40830,0	606,7	0,0	15,2
2	30	1	50		11460,4	31,7	64830,0	616,6	0,0	24,4

3	30	1	80		11622,7	44,1	88830,0	555,5	0,1	30,6
4	30	3	20		11428,8	36,8	74430,0	517,5	0,0	23,8
5	30	3	50		11135,0	73,5	146430,0	565,0	0,1	51,9
6	30	3	80		11311,6	111,9	218430,0	575,7	0,1	80,5
7	30	5	20		11093,9	55,4	108030,0	574,6	0,1	39,8
8	30	5	50		11237,9	120,2	228030,0	373,9	0,2	56,2
9	30	5	80		11231,2	181,9	348030,0	354,2	0,2	80,6
10	30	10	20		11037,7	96,3	192030,0	447,3	0,1	53,8
11	30	10	50		11025,6	213,5	432030,0	470,7	0,3	125,6
12	30	10	80		11004,8	339,6	672030,0	506,6	0,4	215,0
13	50	1	20		11420,0	28,5	56850,0	788,1	0,0	28,1
14	50	1	50		11208,8	40,5	80850,0	737,0	0,1	37,3
15	50	1	80		11068,4	53,2	104850,0	693,3	0,1	46,1
16	50	3	20		11432,9	45,6	90450,0	724,5	0,1	41,3
17	50	3	50		11208,8	83,5	162450,0	699,9	0,1	73,1
18	50	3	80		11308,0	119,1	234450,0	605,0	0,1	90,1
19	50	5	20		11071,6	62,7	124050,0	550,7	0,1	43,1
20	50	5	50		11251,7	122,6	244050,0	566,5	0,2	86,8
21	50	5	80		11143,8	184,5	364050,0	549,8	0,2	126,8
22	50	10	20		11230,1	105,1	208050,0	476,4	0,1	62,6
23	50	10	50		11221,6	226,6	448050,0	370,1	0,3	104,8
24	50	10	80		11128,1	340,9	688050,0	326,1	0,4	139,0
25	75	1	20		11528,9	37,3	76875,0	766,9	0,0	35,7
26	75	1	50		11147,7	49,1	100875,0	738,0	0,1	45,3
27	75	1	80		11127,6	61,1	124875,0	770,3	0,1	58,9
28	75	3	20		11145,2	53,8	110475,0	766,9	0,1	51,6
29	75	3	50		11247,0	90,3	182475,0	704,0	0,1	79,4
30	75	3	80		11196,2	125,0	254475,0	629,9	0,2	98,4
31	75	5	20		11068,1	70,3	144075,0	665,4	0,1	58,5
32	75	5	50		11207,0	124,2	264075,0	595,6	0,2	92,4
33	75	5	80		11149,5	179,6	384075,0	614,7	0,2	138,0
34	75	10	20		11444,6	106,0	228075,0	439,8	0,1	58,3
35	75	10	50		11132,2	219,2	468075,0	388,9	0,3	106,5
36	75	10	80		11188,2	331,4	708075,0	465,5	0,4	192,8

Πίνακας 7 - Αποτελέσματα διαδικασίας ρύθμισης BBMO περ.Α-Α

Επιλέγουμε τον 11^ο συνδυασμό παραμέτρων, διότι δίνει την δεύτερη καλύτερη λύση σε σχετικά μικρό χρόνο.

Αποτελέσματα BBMO – (Τεχνική Α Εργάτρια Β)

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]				Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	30	1	20		11170,1	25,0	40830,0	662,7	0,0	20,7
2	30	1	50		11198,5	39,0	64830,0	603,7	0,0	29,5
3	30	1	80		11215,5	53,5	88830,0	562,3	0,1	37,6

4	30	3	20		11278,6	45,0	74430,0	619,9	0,1	34,9
5	30	3	50		11198,1	91,1	146430,0	567,1	0,1	64,6
6	30	3	80		11170,5	133,0	218430,0	572,4	0,2	95,2
7	30	5	20		11146,9	65,1	108030,0	432,6	0,1	35,2
8	30	5	50		11348,2	142,1	228030,0	314,9	0,2	55,9
9	30	5	80		11153,7	216,0	348030,0	469,6	0,3	126,8
10	30	10	20		11146,3	117,9	192030,0	552,6	0,1	81,4
11	30	10	50		11014,6	268,2	432030,0	505,1	0,3	169,4
12	30	10	80		11006,6	422,5	672030,0	617,5	0,5	326,1
13	50	1	20		11073,2	35,8	56850,0	721,2	0,0	32,2
14	50	1	50		11148,8	50,6	80850,0	710,1	0,1	44,9
15	50	1	80		11194,3	63,8	104850,0	701,9	0,1	56,0
16	50	3	20		11274,6	54,1	90450,0	688,6	0,1	46,6
17	50	3	50		11179,5	99,6	162450,0	641,9	0,1	80,0
18	50	3	80		11254,2	143,0	234450,0	597,6	0,2	106,9
19	50	5	20		11250,3	75,3	124050,0	461,6	0,1	43,5
20	50	5	50		11235,9	149,6	244050,0	485,6	0,2	90,8
21	50	5	80		11186,1	221,8	364050,0	520,6	0,3	144,3
22	50	10	20		11234,4	128,8	208050,0	349,7	0,2	56,3
23	50	10	50		11117,3	278,9	448050,0	390,7	0,3	136,2
24	50	10	80		11170,0	377,9	688050,0	482,8	0,5	228,1
25	75	1	20		11193,3	40,4	76875,0	763,3	0,1	38,5
26	75	1	50		11280,0	53,4	100875,0	657,1	0,1	43,9
27	75	1	80		11345,1	66,4	124875,0	690,5	0,1	57,3
28	75	3	20		11076,3	42,0	110475,0	723,1	0,1	38,0
29	75	3	50		11067,1	62,5	182475,0	648,0	0,1	50,6
30	75	3	80		11120,1	88,2	254475,0	639,1	0,1	70,5
31	75	5	20		11172,8	51,4	144075,0	595,8	0,1	38,3
32	75	5	50		11141,6	89,7	264075,0	596,8	0,1	66,9
33	75	5	80		11152,5	136,2	384075,0	534,3	0,2	91,0
34	75	10	20		11204,2	74,8	228075,0	515,1	0,1	48,2
35	75	10	50		11250,2	243,9	468075,0	580,5	0,3	177,0
36	75	10	80		11147,2	366,2	708075,0	488,7	0,5	223,7

Πίνακας 8 - Αποτελέσματα διαδικασίας ρύθμισης ΒΒΜΟ περ.Α-Β



Γράφημα 6 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Α-Β

Επιλέγουμε τον 11^ο συνδυασμό παραμέτρων μιας και έδωσε την καλύτερη λύση σε σχετικά μικρό χρόνο.

Αποτελέσματα BBMO (Τεχνική Β - Εργάτρια Α)

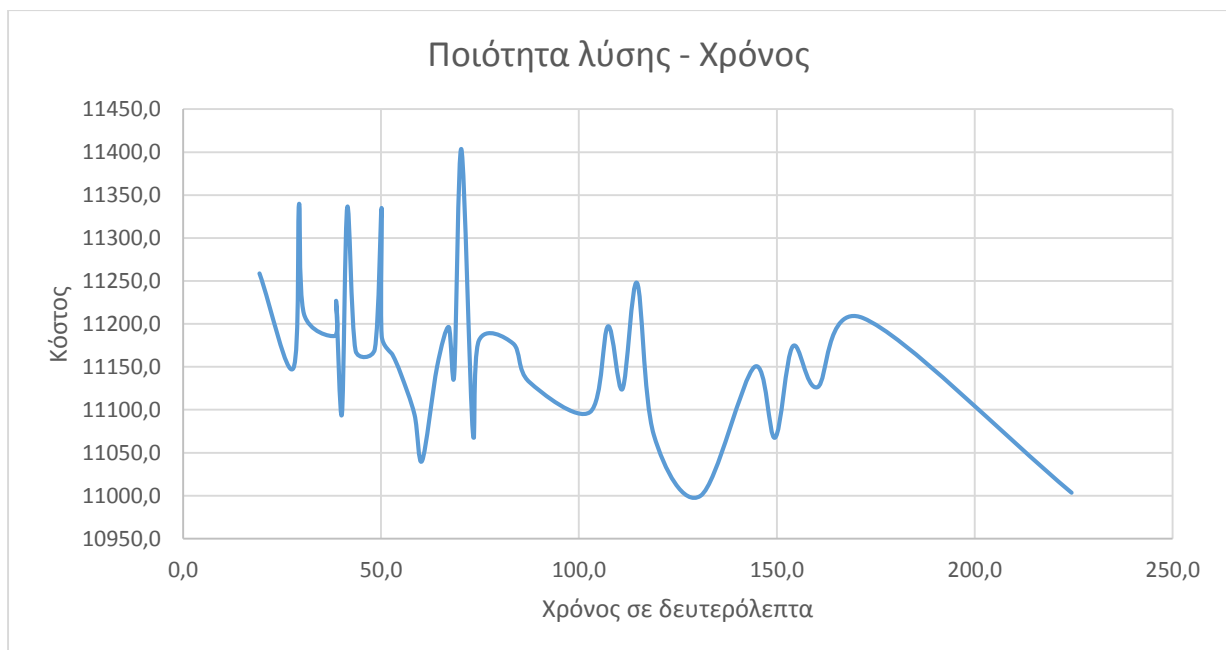
Αντικαθιστούμε τον τελεστή διασταύρωσης και εφαρμόζουμε την «Τεχνική Β» (βλ. κεφ. 7). Ενώ διατηρούμε τον ίδιο τρόπο τοπικής αναζήτησης. Εκτελούμε εκ νέου τις απαραίτητες προσομοιώσεις και παραθέτουμε τα αποτελέσματα στον παρακάτω πίνακα.

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]				Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	30	1	20		11258,7	21,5	40830,0	713,1	0,0	19,3
2	30	1	50		11211,5	35,1	64830,0	684,6	0,0	30,5
3	30	1	80		11335,9	45,3	88830,0	732,0	0,1	41,5
4	30	3	20		11339,1	37,9	74430,0	540,3	0,0	29,3
5	30	3	50		11094,2	74,5	146430,0	666,5	0,1	58,4
6	30	3	80		11132,3	110,4	218430,0	617,2	0,1	87,5
7	30	5	20		11223,7	54,8	108030,0	446,9	0,1	38,6
8	30	5	50		11140,6	116,0	228030,0	458,3	0,1	68,5
9	30	5	80		11247,6	175,7	348030,0	612,0	0,2	114,7
10	30	10	20		11040,5	97,7	192030,0	437,4	0,1	60,4
11	30	10	50		10999,7	218,2	432030,0	438,7	0,3	130,8
12	30	10	80		11003,3	340,2	672030,0	562,3	0,4	224,5

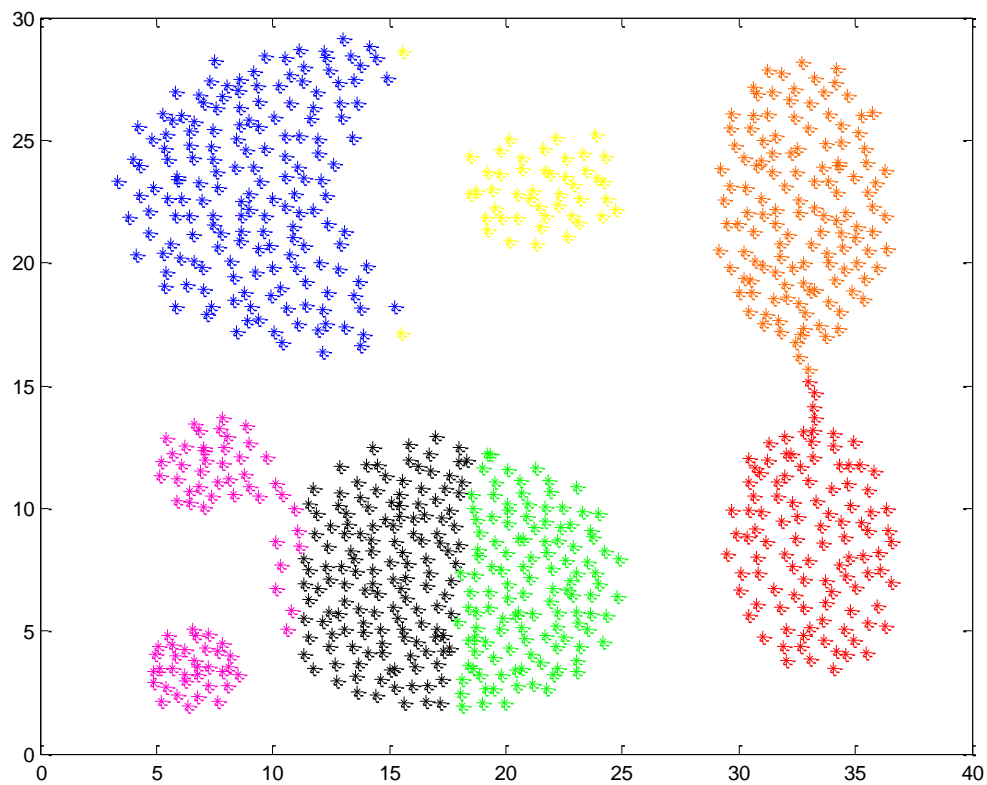
13	50	1	20		11148,2	29,0	56850,0	785,8	0,0	27,8
14	50	1	50		11094,9	41,3	80850,0	764,0	0,1	40,1
15	50	1	80		11185,8	53,3	104850,0	748,0	0,1	50,2
16	50	3	20		11168,7	46,2	90450,0	742,4	0,1	43,6
17	50	3	50		11403,2	82,3	162450,0	740,2	0,1	70,3
18	50	3	80		11097,1	118,9	234450,0	708,1	0,1	102,8
19	50	5	20		11170,3	63,0	124050,0	721,7	0,1	48,4
20	50	5	50		11180,5	124,3	244050,0	731,7	0,2	74,7
21	50	5	80		11149,7	185,4	364050,0	662,6	0,2	144,4
22	50	10	20		11150,1	109,0	208050,0	379,1	0,1	64,2
23	50	10	50		11070,6	259,3	448050,0	451,1	0,3	118,9
24	50	10	80		11206,7	349,9	688050,0	503,0	0,4	171,8
25	75	1	20		11185,9	39,2	76875,0	768,5	0,0	38,3
26	75	1	50		11334,9	51,2	100875,0	753,4	0,1	50,1
27	75	1	80		11141,1	64,7	124875,0	729,0	0,1	55,0
28	75	3	20		11164,0	56,5	110475,0	770,8	0,1	52,9
29	75	3	50		11177,1	97,2	182475,0	711,6	0,1	83,3
30	75	3	80		11197,0	133,3	254475,0	664,7	0,2	107,3
31	75	5	20		11196,3	76,5	144075,0	729,4	0,1	67,1
32	75	5	50		11124,0	139,6	264075,0	673,3	0,2	111,0
33	75	5	80		11173,8	194,8	384075,0	651,5	0,2	154,1
34	75	10	20		11073,1	114,7	228075,0	652,5	0,1	73,2
35	75	10	50		11067,3	235,9	468075,0	553,2	0,3	149,5
36	75	10	80		11126,3	351,5	708075,0	544,9	0,4	160,3

Πίνακας 9 - Αποτελέσματα διαδικασίας ρύθμισης BBMO περ.Β-Α

Σε αυτήν την περίπτωση επιλέγουμε τον 11^ο συνδυασμό παραμέτρων μιας και έδωσε την καλύτερη επίδοση σε σχετικά μικρό χρόνο.



Γράφημα 7 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Β-Α



Εικόνα 14 - Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος BBMO περ.Β-Α

Αποτελέσματα BBMO (Τεχνική Β, Εργάτρια Β)

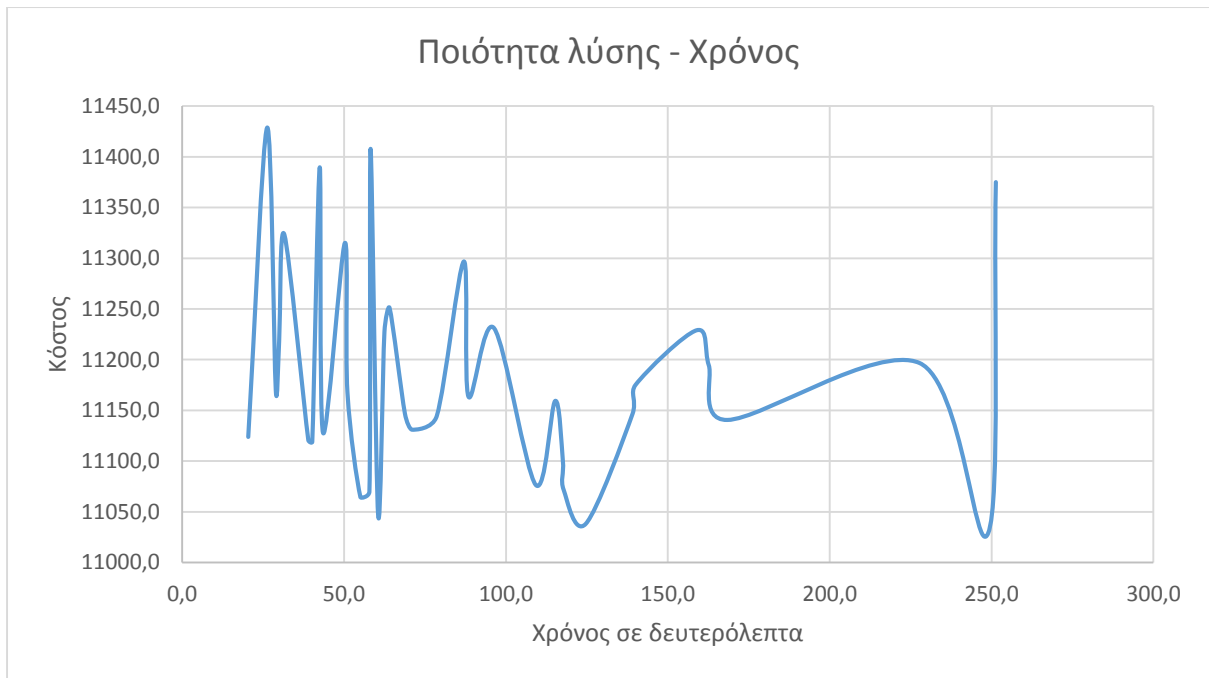
Αντικαθιστούμε την διαδικασία τοπικής αναζήτησης, όπου πλέον αντί να αλλάζουμε μόνο τη μία διάσταση, αλλάζουμε όλες τις διαστάσεις του συγκεκριμένου κέντρου του τμήματος. Εκτελούμε εκ νέου τις απαραίτητες προσομοιώσεις και παραθέτουμε τα αποτελέσματα στον παρακάτω πίνακα.

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]				Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	30	1	20		11123,9	22,9	40830,0	713,1	0,0	20,4
2	30	1	50		11218,5	35,0	64830,0	684,6	0,0	30,0
3	30	1	80		11389,6	46,5	88830,0	732,0	0,1	42,5
4	30	3	20		11428,6	38,9	74430,0	540,3	0,0	26,2
5	30	3	50		11250,9	76,9	146430,0	666,5	0,1	64,1
6	30	3	80		11163,6	114,6	218430,0	617,2	0,1	88,4
7	30	5	20		11323,1	56,5	108030,0	446,9	0,1	31,6
8	30	5	50		11146,9	120,1	228030,0	458,3	0,2	68,8

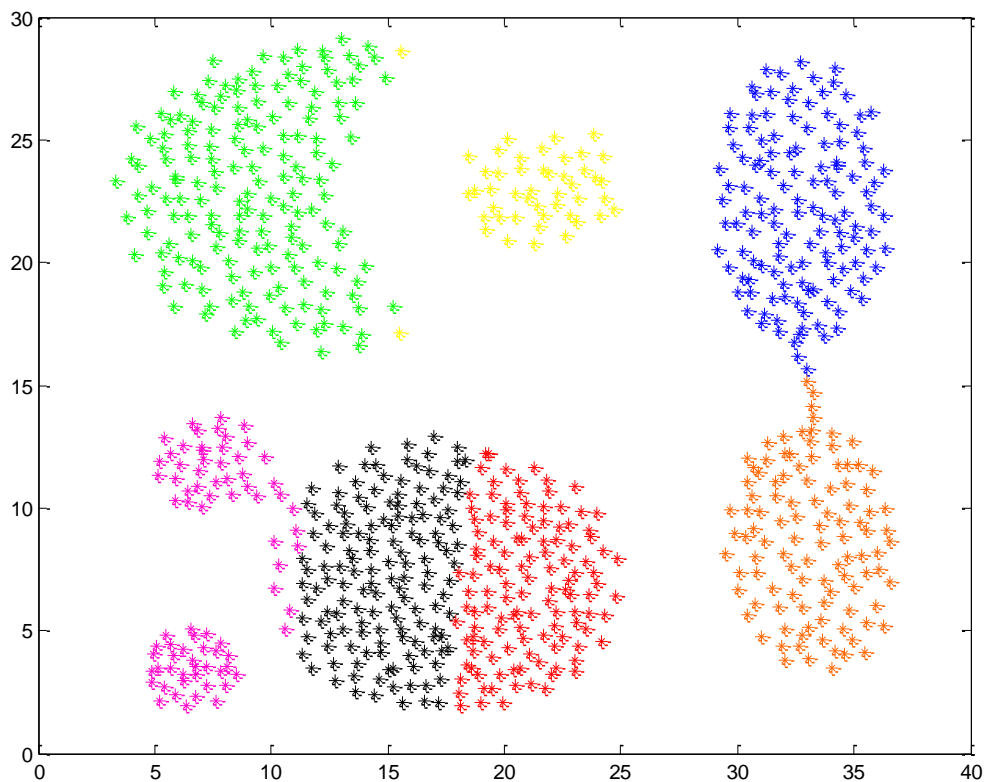
9	30	5	80		11175,3	183,1	348030,0	612,0	0,2	140,1
10	30	10	20		11064,5	100,7	192030,0	437,4	0,1	55,1
11	30	10	50		11038,3	227,4	432030,0	438,7	0,3	124,7
12	30	10	80		11028,7	354,2	672030,0	562,3	0,4	248,9
13	50	1	20		11170,1	29,4	56850,0	785,8	0,0	28,9
14	50	1	50		11119,1	42,1	80850,0	764,0	0,1	40,2
15	50	1	80		11167,9	54,6	104850,0	748,0	0,1	51,1
16	50	3	20		11127,9	46,9	90450,0	742,4	0,1	43,5
17	50	3	50		11144,5	85,0	162450,0	740,2	0,1	78,7
18	50	3	80		11076,6	123,4	234450,0	708,1	0,2	109,2
19	50	5	20		11407,8	64,5	124050,0	721,7	0,1	58,2
20	50	5	50		11100,6	128,6	244050,0	731,7	0,2	117,6
21	50	5	80		11229,0	191,9	364050,0	662,6	0,2	159,0
22	50	10	20		11227,5	131,9	208050,0	379,1	0,2	62,5
23	50	10	50		11146,2	246,7	448050,0	451,1	0,3	139,1
24	50	10	80		11196,8	362,2	688050,0	503,0	0,5	227,7
25	75	1	20		11120,7	40,6	76875,0	768,5	0,1	39,0
26	75	1	50		11314,9	53,4	100875,0	753,4	0,1	50,2
27	75	1	80		11047,0	66,4	124875,0	729,0	0,1	60,5
28	75	3	20		11070,0	60,0	110475,0	770,8	0,1	57,8
29	75	3	50		11296,8	97,8	182475,0	711,6	0,1	87,0
30	75	3	80		11159,2	138,7	254475,0	664,7	0,2	115,2
31	75	5	20		11131,0	77,9	144075,0	729,4	0,1	71,0
32	75	5	50		11071,8	140,0	264075,0	673,3	0,2	117,8
33	75	5	80		11194,3	199,8	384075,0	651,5	0,2	162,7
34	75	10	20		11231,0	118,1	228075,0	652,5	0,1	96,4
35	75	10	50		11140,6	243,5	468075,0	553,2	0,3	168,4
36	75	10	80		11375,0	369,0	708075,0	544,9	0,5	251,4

Πίνακας 10 - Αποτελέσματα διαδικασίας ρύθμισης BBMO περ.Β-Β

Επιλέγουμε τον 27^ο συνδυασμό παραμέτρων, καθώς έδωσε την καλύτερη λύση σε σχέση με τους υπολοίπους σε σχετικά μικρό χρονικό διάστημα.



Γράφημα 8 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Β-Β



Εικόνα 15 -Τμηματοποίηση των δεδομένων που αντιστοιχεί στην καλύτερη λύση που βρήκε ο αλγόριθμος BBMO περ.Β-Β

ΒΒΜΟ Τεχνική Γ, εργάτρια Α

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]				Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	30	1	20		11095,2	24,7	40830,0	534,5	0,0	16,5
2	30	1	50		11594,6	39,4	64830,0	479,9	0,0	23,6
3	30	1	80		11388,9	55,7	88830,0	549,7	0,1	38,2
4	30	3	20		11304,8	46,7	74430,0	358,6	0,1	20,9
5	30	3	50		11045,2	92,1	146430,0	495,7	0,1	57,1
6	30	3	80		11020,8	135,9	218430,0	374,3	0,2	63,6
7	30	5	20		11022,4	67,3	108030,0	506,7	0,1	42,6
8	30	5	50		11002,0	142,6	228030,0	429,9	0,2	76,6
9	30	5	80		10998,9	217,7	348030,0	379,5	0,3	103,3
10	30	10	20		11163,8	109,2	192030,0	459,5	0,1	62,7
11	30	10	50		11044,4	274,2	432030,0	479,5	0,3	164,3
12	30	10	80		11018,8	525,5	672030,0	537,3	0,7	352,9
13	50	1	20		11299,9	34,4	56850,0	399,7	0,0	17,2
14	50	1	50		11318,9	49,4	80850,0	415,6	0,1	25,7
15	50	1	80		11163,2	64,7	104850,0	402,4	0,1	32,6
16	50	3	20		11619,4	55,1	90450,0	498,5	0,1	34,3
17	50	3	50		11389,2	100,2	162450,0	467,6	0,1	58,5
18	50	3	80		11090,2	143,6	234450,0	338,6	0,2	60,8
19	50	5	20		11392,7	75,7	124050,0	412,3	0,1	39,0
20	50	5	50		11066,2	150,6	244050,0	321,4	0,2	60,5
21	50	5	80		11020,5	223,4	364050,0	416,2	0,3	116,2
22	50	10	20		11024,8	115,8	208050,0	344,7	0,1	49,9
23	50	10	50		11003,3	290,8	448050,0	528,3	0,4	192,0
24	50	10	80		10999,3	524,6	688050,0	463,9	0,7	304,2
25	75	1	20		11090,4	37,9	76875,0	543,2	0,0	25,7
26	75	1	50		11600,8	51,8	100875,0	465,6	0,1	30,1
27	75	1	80		11873,5	61,8	124875,0	408,2	0,1	31,5
28	75	3	20		11160,8	54,6	110475,0	450,8	0,1	30,8
29	75	3	50		11090,3	91,1	182475,0	370,8	0,1	42,2
30	75	3	80		11138,1	127,3	254475,0	492,9	0,2	78,5
31	75	5	20		11092,0	72,7	144075,0	548,3	0,1	49,8
32	75	5	50		11319,3	131,0	264075,0	542,3	0,2	88,8
33	75	5	80		11090,2	190,5	384075,0	330,2	0,2	78,6
34	75	10	20		11141,2	113,3	228075,0	372,9	0,1	52,8
35	75	10	50		11024,0	237,6	468075,0	380,7	0,3	113,1
36	75	10	80		11119,5	354,9	708075,0	337,0	0,4	149,5

Πίνακας 11 - Αποτελέσματα διαδικασίας ρύθμισης ΒΒΜΟ περ.Γ-Α



Γράφημα 9 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Γ-Α

Επιλέγουμε το συνδυασμό 9 μιας και έδωσε την καλύτερη λύση σε σχετικά μικρό χρονικό διάστημα.

Αποτελέσματα BBMO (Τεχνική Γ, Εργατρία Β)

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσεις, επαναλήψεις τοπικής αναζήτησης]				Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	30	1	20		11123,5	21,0	40830,0	389,0	0,0	10,2
2	30	1	50		11163,3	33,7	64830,0	366,3	0,0	15,4
3	30	1	80		11299,4	47,0	88830,0	468,1	0,1	27,5
4	30	3	20		11096,0	38,7	74430,0	354,4	0,0	17,1
5	30	3	50		11181,0	78,5	146430,0	296,7	0,1	29,1
6	30	3	80		11117,0	113,1	218430,0	405,7	0,1	57,4
7	30	5	20		11197,2	55,5	108030,0	332,2	0,1	23,0
8	30	5	50		11143,3	117,9	228030,0	468,1	0,1	69,0
9	30	5	80		11094,7	179,9	348030,0	360,1	0,2	81,0
10	30	10	20		11131,5	100,4	192030,0	542,8	0,1	68,1
11	30	10	50		11049,7	226,9	432030,0	437,2	0,3	124,0
12	30	10	80		11027,3	347,8	672030,0	421,6	0,4	183,3
13	50	1	20		11227,3	28,7	56850,0	574,6	0,0	20,6
14	50	1	50		11073,1	41,5	80850,0	400,7	0,1	20,8
15	50	1	80		11258,4	54,7	104850,0	593,2	0,1	40,6
16	50	3	20		11139,4	47,9	90450,0	557,0	0,1	33,3
17	50	3	50		11164,6	84,6	162450,0	439,1	0,1	46,4
18	50	3	80		11340,5	119,6	234450,0	505,9	0,1	75,7

19	50	5	20		11152,1	64,9	124050,0	279,6	0,1	22,7
20	50	5	50		11073,7	130,5	244050,0	481,5	0,2	78,5
21	50	5	80		11107,7	188,7	364050,0	352,2	0,2	83,1
22	50	10	20		11146,8	107,5	208050,0	420,0	0,1	56,5
23	50	10	50		11139,0	232,7	448050,0	503,1	0,3	146,3
24	50	10	80		11133,4	362,2	688050,0	319,4	0,5	144,6
25	75	1	20		11227,4	38,0	76875,0	352,7	0,0	16,8
26	75	1	50		11138,1	50,3	100875,0	481,3	0,1	30,3
27	75	1	80		11221,2	62,4	124875,0	514,5	0,1	40,1
28	75	3	20		11225,3	77,9	110475,0	406,5	0,1	39,6
29	75	3	50		11121,2	179,2	182475,0	449,7	0,2	100,7
30	75	3	80		11203,6	241,5	254475,0	348,5	0,3	105,2
31	75	5	20		11073,6	136,4	144075,0	500,0	0,2	85,2
32	75	5	50		11120,6	249,4	264075,0	478,7	0,3	149,2
33	75	5	80		11020,2	376,1	384075,0	436,7	0,5	205,3
34	75	10	20		11094,9	210,7	228075,0	371,4	0,3	97,8
35	75	10	50		11150,1	413,4	468075,0	582,2	0,5	300,8
36	75	10	80		11219,8	495,6	708075,0	387,8	0,6	240,3

Πίνακας 12 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ.Γ-Β



Γράφημα 10 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο BBMO περ. Γ-Β

Σε αυτήν την περίπτωση επιλέγουμε τον 33^ο συνδυασμό μας και μας δίνει την καλύτερη λύση.

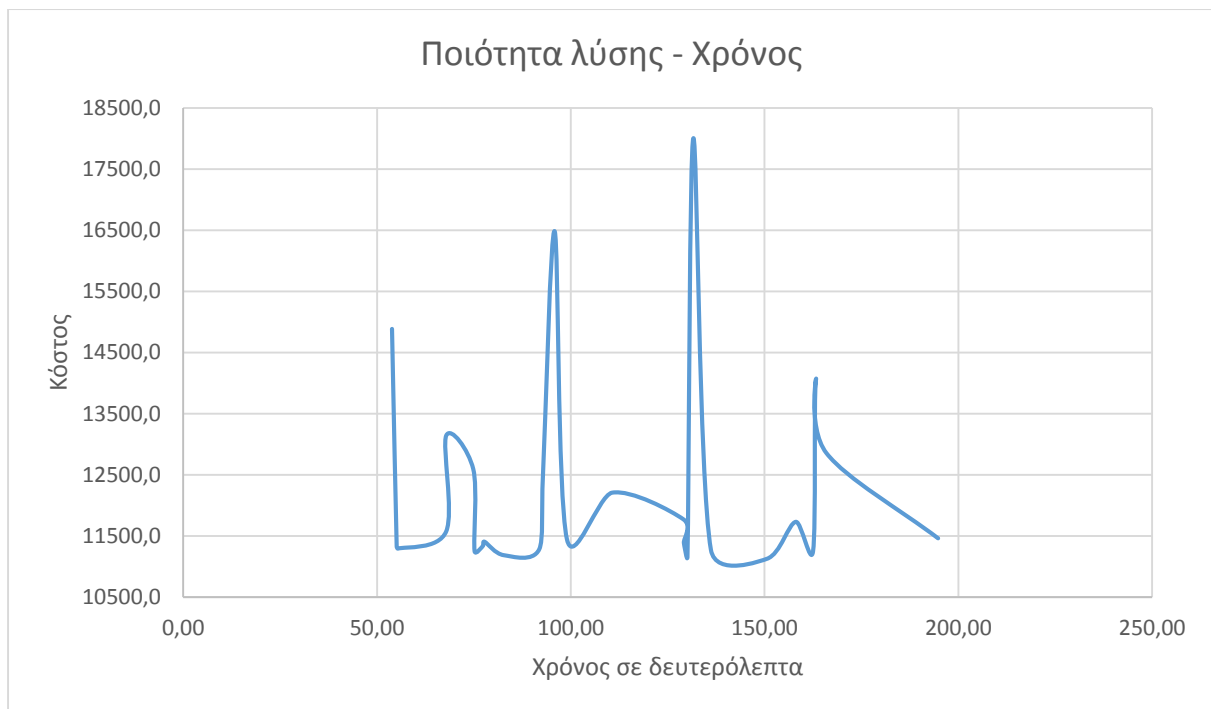
8.3. Διακριτή αναπαράσταση

Προκειμένου να ρυθμίσουμε τις παραμέτρους των αλγορίθμων με την νέα αναπαράσταση λύσης, εφαρμόζουμε την ίδια διαδικασία με προηγούμενως.

Αλγόριθμος HBMO

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίτσες, επαναλήψεις τοπικής αναζήτησης]			Βέλτιστη λύση	Χρόνος	Evaluation	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	15	1	100	14886,3	53,9	108515,0	499,9	0,11	53,89
2	15	1	250	11292,2	55,7	257962,8	496,9	0,11	55,36
3	15	1	500	11333,0	63,1	496992,0	436,3	0,13	55,10
4	15	3	100	13141,5	67,9	225816,4	499,9	0,14	67,87
5	15	3	250	11573,4	71,2	463230,6	476,9	0,14	67,86
6	15	3	500	11337,7	89,5	867431,4	432,0	0,18	77,34
7	15	5	100	12595,7	74,9	253551,0	499,9	0,15	74,84
8	15	5	250	11185,0	87,8	581903,4	471,4	0,18	82,75
9	15	5	500	11279,7	112,8	987471,0	407,3	0,23	91,85
10	25	1	100	16481,1	95,9	113525,0	500,0	0,19	95,86
11	25	1	250	11405,6	77,8	263424,6	499,4	0,16	77,73
12	25	1	500	11249,9	81,5	497593,2	461,3	0,16	75,20
13	25	3	100	12493,1	92,8	264580,6	500,0	0,19	92,83
14	25	3	250	11450,8	103,8	627826,4	476,8	0,21	98,99
15	25	3	500	11158,4	130,7	1247389,8	497,9	0,26	130,10
16	25	5	100	12207,9	110,8	344673,6	499,6	0,22	110,67
17	25	5	250	11385,2	144,0	754531,2	448,4	0,29	129,18
18	25	5	500	11248,9	176,1	1525945,8	461,5	0,35	162,49
19	40	1	100	18007,8	131,6	121040,0	500,0	0,26	131,64
20	40	1	250	11771,7	129,1	271040,0	499,9	0,26	129,06
21	40	1	500	11271,7	136,5	520939,8	498,6	0,27	136,16
22	40	3	100	14034,0	158,2	296719,4	499,7	0,32	158,07
23	40	3	250	11730,9	154,9	707880,4	486,2	0,31	150,58
24	40	3	500	11125,3	173,0	1357108,8	479,5	0,35	165,95
25	40	5	100	12855,5	163,3	361298,8	499,9	0,33	163,22
26	40	5	250	11461,06202	196,2	962444,6	496,2	0,39	194,75
27	40	5	500	11466,8	205,1	1848690	472,5	0,41	193,84

Πίνακας 13 - Αποτελέσματα διαδικασίας ρύθμισης διακριτού HBMO περ.Α-Δ



Γράφημα 11 - Ποιότητα λύσης με το χρόνο για τον διακριτό αλγόριθμο HBMO περ.Α-Δ

HBMO Με εφαρμογή του τελεστή Mutation αντί τοπικής αναζήτησης

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]			Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	15	1	100	11621,2	66,9	108494,8	489,4	0,1	65,5
2	15	1	250	11328,9	84,6	243204,0	327,5	0,2	55,4
3	15	1	500	11231,0	134,4	507813,6	164,5	0,3	44,2
4	15	3	100	11397,8	107,6	224200,4	480,7	0,2	103,4
5	15	3	250	11195,6	152,7	533058,8	312,2	0,3	95,4
6	15	3	500	11226,2	215,2	876249,0	207,4	0,4	89,3
7	15	5	100	11552,9	109,2	244299,4	465,3	0,2	101,6
8	15	5	250	11356,6	161,4	587927,4	330,8	0,3	106,8
9	15	5	500	11366,5	267,2	1177650,6	219,6	0,5	117,3
10	25	1	100	12271,5	76,9	113525,0	497,6	0,2	76,5
11	25	1	250	11327,9	108,6	263223,8	377,6	0,2	82,0
12	25	1	500	11091,5	156,3	512523,0	256,5	0,3	80,2
13	25	3	100	11414,8	143,7	274842,2	476,5	0,3	137,0
14	25	3	250	11413,3	201,3	643388,4	353,8	0,4	142,4
15	25	3	500	11333,9	331,1	1328652,0	244,5	0,7	161,9
16	25	5	100	12059,0	152,5	317363,2	487,9	0,3	148,8
17	25	5	250	11345,6	232,5	752573,4	307,7	0,5	143,1
18	25	5	500	11225,8	402,2	1644783,0	204,5	0,8	164,5

19	40	1	100		12701,8	118,9	121040,0	497,8	0,2	118,4
20	40	1	250		11153,6	128,6	270939,6	437,9	0,3	112,7
21	40	1	500		11172,8	181,8	520238,4	300,6	0,4	109,3
22	40	3	100		11282,1	176,6	293810,6	486,6	0,4	171,8
23	40	3	250		11294,0	233,7	645331,2	355,2	0,5	166,0
24	40	3	500		11355,4	366,3	1349794,2	260,9	0,7	191,1
25	40	5	100		11822,4	205,6	388972,8	488,4	0,4	200,8
26	40	5	250		11132,2	309,2	977404,2	295,9	0,6	183,0
27	40	5	500		11348,5	447,2	1707007,2	241,9	0,9	216,4

Πίνακας 14 – Αποτελέσματα διαδικασίας ρύθμισης διακριτού HBMO με χρήση του τελεστή mutation



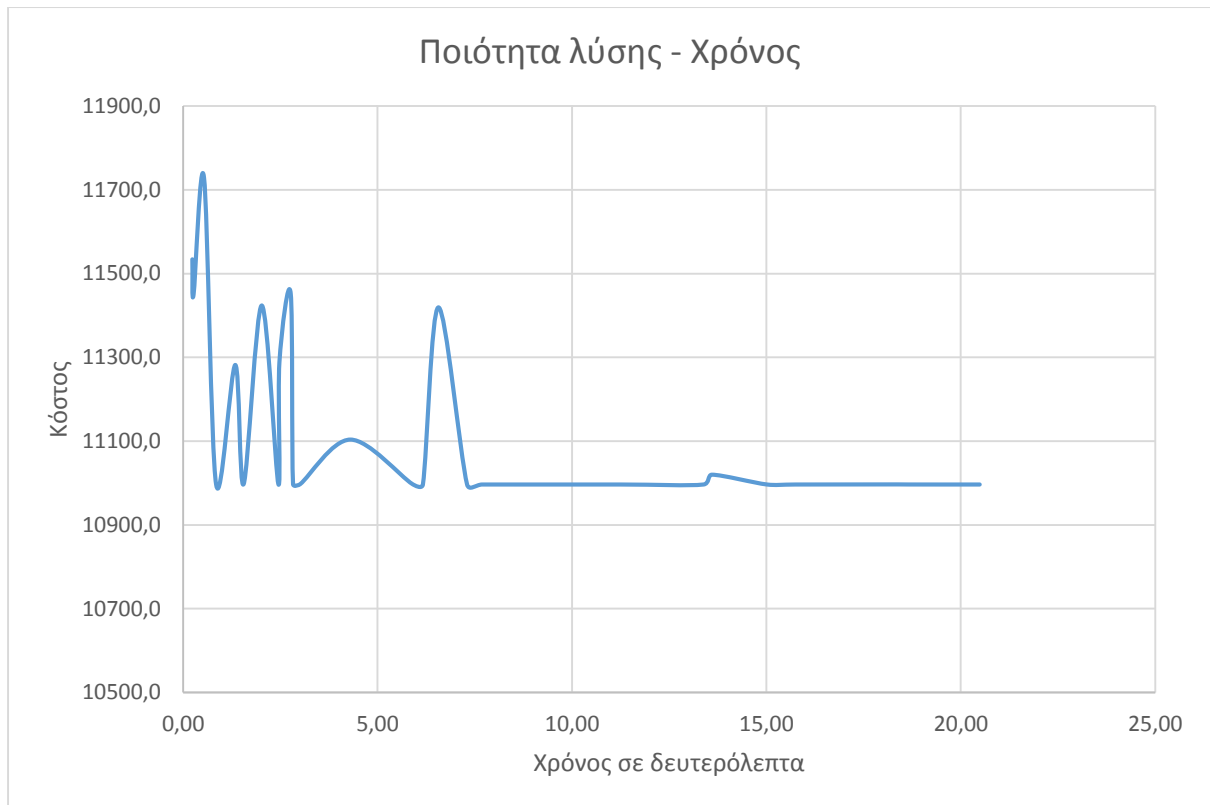
Γράφημα 12 – Ποιότητα λύσης με το χρόνο για τον διακριτό αλγόριθμο HBMO με χρήση του τελεστή mutation

HBMO με εφαρμογή των τελεστών mutation και KMO

Εφαρμόζουμε τον τελεστή KMO στα δεδομένα μας

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]			Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	15	1	100	11728,6	16,4	16817,6	16,4	0,03	0,54
2	15	1	250	11534,1	10,9	26060,4	10,9	0,02	0,24
3	15	1	500	11450,1	11,6	48884,2	11,6	0,02	0,27
4	15	3	100	10997,8	37,6	52280,3	37,6	0,08	2,83
5	15	3	250	10997,7	38,8	331655,6	38,8	0,08	3,01
6	15	3	500	10997,7	27,9	256226,1	27,9	0,06	1,56
7	15	5	100	10997,8	54,3	226938,8	54,3	0,11	5,90
8	15	5	250	10997,7	35,0	181975,2	35	0,07	2,45
9	15	5	500	10997,0	20,6	287189,7	20,6	0,04	0,85
10	25	1	100	11282,1	25,9	28035,4	25,9	0,05	1,34
11	25	1	250	11424,1	59,2	44301,9	17,1	0,12	2,02
12	25	1	500	11452,9	68,8	79257,7	20,1	0,14	2,77
13	25	3	100	11020,2	93,8	138158,6	72,5	0,19	13,60
14	25	3	250	11049,8	156,2	368977,5	78,0	0,31	24,37
15	25	3	500	10996,8	328,7	913995,9	11,7	0,66	7,69
16	25	5	100	10996,9	123,2	199790,2	75,7	0,25	18,64
17	25	5	250	10996,8	160,1	338072,5	35,1	0,32	11,24
18	25	5	500	10996,8	182,7	421341,8	41,1	0,37	15,02
19	40	1	100	11104,0	90,4	41163,3	23,8	0,18	4,30
20	40	1	250	11419,5	98,3	78673,5	33,4	0,20	6,57
21	40	1	500	11.288,6	99,9	105329,8	12,4	0,20	2,48
22	40	3	100	10.996,8	113,0	124495,3	27,3	0,23	6,17
23	40	3	250	10.997,0	144,3	248831,7	25,3	0,29	7,30
24	40	3	500	10.997,1	306,3	749069,6	30	0,61	18,38
25	40	5	100	10996,8	170,7	256169,0	60	0,34	20,49
26	40	5	250	10996,75605	243,3	477255,4	27,5	0,49	13,38
27	40	5	500	10996,8	305,2	792811,7	25,8	0,61	15,75

Πίνακας 15 - Αποτελέσματα διαδικασίας ρύθμισης διακριτού HBMO με χρήση του τελεστή mutation και KMO



Γράφημα 13 - Ποιότητα λύσης με το χρόνο για τον διακριτό αλγόριθμο HBMO με χρήση του τελεστή mutation και KMO

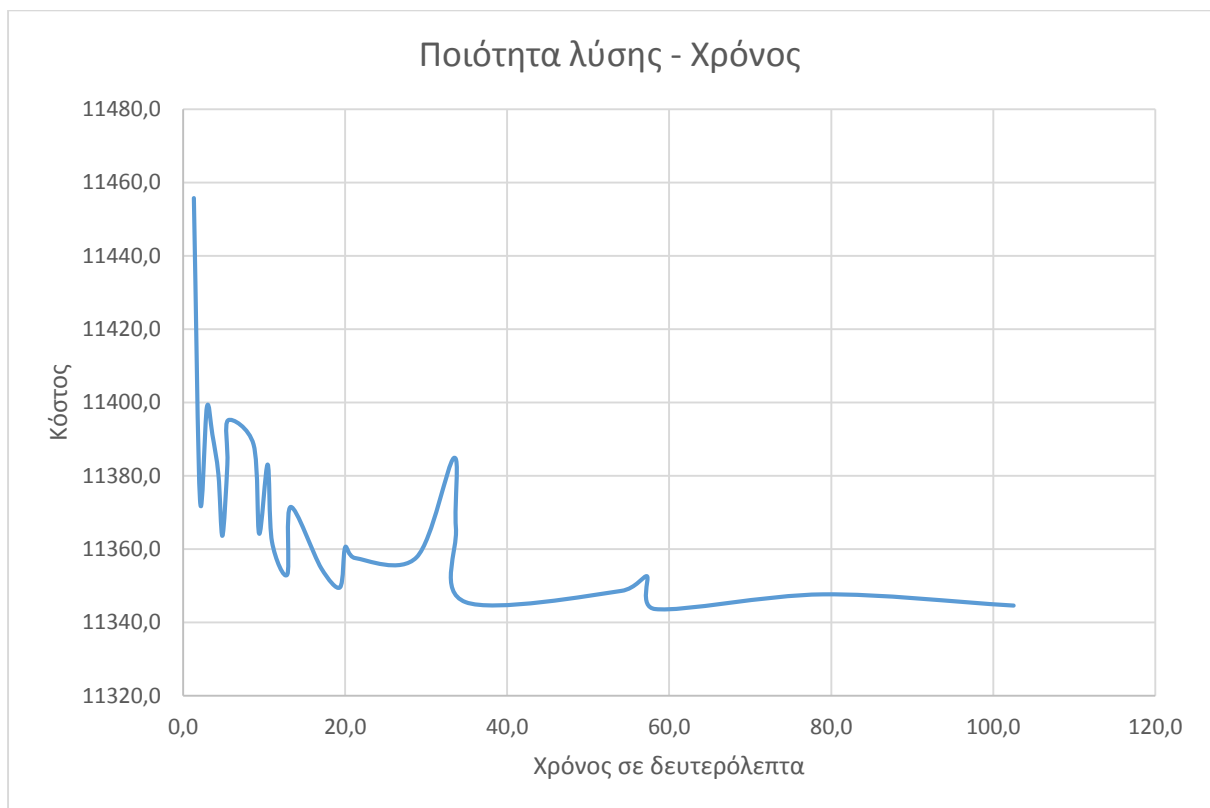
HBMO - Αναπαράσταση λύσης μέσω medoid

Ένας τρίτος τρόπος αναπαράστασης λύσης, όπως προαναφέρθηκε, είναι η αναπαράσταση με ακέραια μορφή της «ταυτότητας» του αντικειμένου που θεωρούμε ως κέντρο του τμήματος. Με αυτόν τον τρόπο, το μήκος ενός διανύσματος λύσης είναι ίσο με τον αριθμό των τμημάτων. Εκτός αυτού, εξασφαλίζουμε ότι τα κέντρα των τμημάτων είναι αντικείμενα που υπάρχουν στο σετ δεδομένων και επομένως θα έχουν φυσική σημασία.

	Συνδυασμός παραμέτρων [πληθυσμός, βασιλίσσες, επαναλήψεις τοπικής αναζήτησης]			Βέλτιστη λύση	Χρόνος	Evaluations	Επανάληψη	Χρόνος ανά επανάληψη	Μέσος χρόνος για εύρεση βέλτιστου
1	15	1	100	11455,7	11,3	24240,6	58,8	0,0	1,3
2	15	1	250	11373,4	19,9	43659,0	52,2	0,0	2,1
3	15	1	500	11391,3	34,5	76352,4	52,2	0,1	3,6
4	15	3	100	11380,4	26,9	59752,2	81,0	0,1	4,4
5	15	3	250	11364,2	56,7	129576,3	82,7	0,1	9,4
6	15	3	500	11354,7	107,2	247995,0	79,6	0,2	17,1

7	15	5	100	11371,5	49,0	110383,5	135,4	0,1	13,3
8	15	5	250	11353,1	74,6	170765,4	86,1	0,1	12,9
9	15	5	500	11349,6	127,2	296241,3	76,1	0,3	19,4
10	25	1	100	11398,7	17,0	38431,5	86,0	0,0	2,9
11	25	1	250	11363,6	29,4	67494,0	82,5	0,1	4,8
12	25	1	500	11383,0	40,9	94739,1	66,7	0,1	5,5
13	25	3	100	11383,1	39,9	91395,9	130,8	0,1	10,4
14	25	3	250	11360,6	82,0	190835,4	122,0	0,2	20,0
15	25	3	500	11357,6	113,2	264728,4	93,8	0,2	21,2
16	25	5	100	11357,8	77,3	179296,2	186,3	0,2	28,8
17	25	5	250	11345,5	135,9	317640,6	128,3	0,3	34,9
18	25	5	500	11343,7	215,2	505058,1	135,4	0,4	58,3
19	40	1	100	11395,1	23,5	54218,4	118,0	0,0	5,5
20	40	1	250	11388,5	38,1	88864,2	114,0	0,1	8,7
21	40	1	500	11362,0	55,9	130811,1	98,0	0,1	11,0
22	40	3	100	11384,8	71,0	165510,3	235,3	0,1	33,4
23	40	3	250	11348,6	127,9	299704,2	211,5	0,3	54,1
24	40	3	500	11352,6	192,9	454256,7	148,4	0,4	57,2
25	40	5	100	11365,8	85,8	198204,0	196,1	0,2	33,7
26	40	5	250	11347,7	193,9	455499,9	203,7	0,4	79,0
27	40	5	500	11344,6	294,6	694285,8	174,0	0,6	102,5

Πίνακας 16 - Αποτελέσματα διαδικασίας ρύθμισης HBMO medoid περ. Α-Δ



Γράφημα 14 - Ποιότητα λύσης με το χρόνο για τον αλγόριθμο HBMO medoid περ.Α-Δ

8.4. Σύνοψη αποτελεσμάτων διαδικασίας ρύθμισης

Σε αυτήν την παράγραφο, θα παρατεθούν τα βέλτιστα αποτελέσματα για κάθε περίπτωση του αλγορίθμου, με σκοπό να επιλέξουμε την καλύτερη εκδοχή. Στο επόμενο κεφάλαιο, θα εφαρμόσουμε την συγκεκριμένη εκδοχή του κάθε αλγορίθμου σε διάφορα γνωστά datasets, προκειμένου να εξετάσουμε συγκριτικά την επίδοσή τους.

Συνεχής αναπαράσταση

Αλγόριθμος HBMO

		Βέλτιστη λύση	Μέσος χρόνος για εύρεση βέλτιστου
1	HBMO-A-A	10998,0	236,9
2	HBMO-B-A	10997,4	172,04
3	HBMO - Γ -A	10996,8	129,7
4	HBMO – B -B	11122,3	213,1
5	HBMO – B –Γ	11020,5	510,8

Αλγόριθμος ABC

	Βέλτιστη λύση	Μέσος χρόνος για εύρεση βέλτιστου
ABC	10996,8	18,15

Αλγόριθμος BBMO

		Βέλτιστη λύση	Μέσος χρόνος για εύρεση βέλτιστου
1	BBMO-A-A	11025,6	125,6
2	BBMO-A-B	11014,6	169,4
3	BBMO-B-A	10999,7	130,8
4	BBMO-B-B	11047,0	60,5
5	BBMO-Γ-A	10998,9	103,3
6	BBMO-Γ-B	11020,2	205,3

Διακριτή Αναπαράσταση

		Βέλτιστη λύση	Μέσος χρόνος για εύρεση βέλτιστου
1	HBMO-A-Δ	11185,0	82,75
2	HBMO-μετ.	11091	80,2
3	HBMO - μετ. & KMO	10996,8	6,17
4	HBMO - medoid	11343,7	58,3

9. Αξιολόγηση σε γνωστά σετ δεδομένων

Η υλοποίηση και εκτέλεση των αλγορίθμων έγινε σε περιβάλλον Matlab 2013b σε έναν υπολογιστή Intel i5 με 6 GB RAM. Επιλέχθηκε μία πληθώρα από σετ δεδομένων προκειμένου να αξιολογηθεί η επίδοση των αλγορίθμων όσον αφορά τη βέλτιστη λύση, το χρόνο εκτέλεσης αλλά και το πόσες φορές καλείται η συνάρτηση ποιότητας. Συγκεκριμένα χρησιμοποιήθηκαν τα εξής σετ δεδομένων: Iris, Wine και το Wholesale Customer Data τα οποία αντλήθηκαν από το Machine Learning Laboratory [51].

- Το σετ δεδομένων Iris, είναι πιθανώς το πιο γνωστό στο τομέα της αναγνώρισης προτύπων. Πρωτοεμφανίστηκε στην ερευνητική δουλειά του Fisher και χρησιμοποιείται μέχρι και σήμερα. Περιλαμβάνει 50 μετρήσεις από 3 διαφορετικές κλάσεις φυτών. Η μία κλάση είναι γραμμικά διαχωρίσιμη από τις άλλες δύο. Οι τελευταίες όμως, δεν είναι γραμμικά διαχωρίσιμες μεταξύ τους. Οι διαστάσεις των δεδομένων είναι τέσσερις και αφορούν στο μήκος του φύλλου κάλυκος του ανθού σε εκατοστά, το πλάτος του σε εκατοστά, το μήκος του πετάλου και το πλάτος του πετάλου σε εκατοστά.
- Το σετ δεδομένων Wine αποτελείται από εργαστηριακές μετρήσεις μίας χημικής ανάλυσης από κρασιά τα οποία προέρχονται από την ίδια γεωγραφική τοποθεσία, αλλά από διαφορετικούς καλλιεργητές. Η ανάλυση έχει να κάνει με μετρήσεις 13 διαφορετικών συστατικών, όπως αυτά μετρήθηκαν στους τρεις τύπους κρασιών. Περιλαμβάνονται 178 μετρήσεις, η κάθε μία από τις οποίες αποτελείται από τις επιμέρους μετρήσεις των 13 συστατικών.
- Το σετ δεδομένων Wholesale customers αποτελείται από 440 αντικείμενα 8 διαστάσεων τα οποία αφορούν τα ποσά σε χρηματικές μονάδες που καταναλώνουν οι καταναλωτές σε συγκεκριμένα αγαθά.

Σε αυτό το σημείο πρέπει να σημειωθεί πως η ποιότητα της τμηματοποίησης μπορεί να οριστεί σε δύο επίπεδα. Αρχικά στο επίπεδο της συνάρτησης κόστους και αφετέρου στην εύρεση του βέλτιστου.

Το πρώτο έχει να κάνει με το εάν το βέλτιστο της συνάρτησης κόστους που έχουμε επιλέξει συμπίπτει όντως με την βέλτιστη (κατά τον χρήστη) τμηματοποίηση των δεδομένων. Για παράδειγμα στο σετ δεδομένων με το οποίο ρυθμίσαμε τις παραμέτρους η συνάρτηση κόστους μπορεί μεν να παρουσιάζει ελάχιστο, αλλά παρατηρώντας την τμηματοποίηση που αντιστοιχεί σε αυτό το ελάχιστο, γίνεται εύκολα αντιληπτό ότι το κριτήριο που επιλέξαμε δεν είναι το κατάλληλο.

Το δεύτερο έχει να κάνει με το κατά πόσο ο εκάστοτε αλγόριθμος καταφέρει να βρει το βέλτιστο της δοθείσας συνάρτησης κόστους. Σε αυτό οι αλγόριθμοι παρουσίασαν επαρκή επίδοση, μιας και κατάφεραν να βρουν το βέλτιστο της συνάρτησης στην πλειοψηφία των περιπτώσεων. Καθώς λοιπόν σκοπός της εργασίας είναι η προσαρμογή των αλγορίθμων στο πρόβλημα και η αξιολόγησή τους, δεν θα μελετηθεί το κατά πόσο «σωστή» ήταν η τμηματοποίηση, μιας και αυτό όπως προαναφέρθηκε έχει να κάνει με την συνάρτηση κόστους και όχι με τον αλγόριθμο. Αντίθετα θα μελετήσουμε την επίδοση των αλγορίθμων ως προς την εύρεση λύσης.

Εφόσον οι παράμετροι των αλγορίθμων έχουν ρυθμιστεί, τους εφαρμόζουμε σε διάφορα γνωστά σετ δεδομένων προκειμένου να ελέγξουμε την αποτελεσματικότητά τους.

9.1. Αποτελέσματα στο Iris Dataset

Εκτελέστηκαν δέκα επαναλήψεις του κάθε αλγορίθμου στο σετ δεδομένων «Iris» με παραμέτρους αυτές που αναφέρθηκαν στο προηγούμενο κεφάλαιο. Αρχικά παραθέτουμε τα αποτελέσματα για τη συνεχή αναπαράσταση λύσης.

	Καλύτερη λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	78.9408	78.9408	78.9408		45.4	543710
ABC	78.9408	78.9408	78.9409		4.69	50021
BBMO	78.9428	78.9455	78.9479		49.93	348030
K-means	78.9408	91.7230	142.8516		0.004	-

Πίνακας 17 – Αποτελέσματα στο Iris Dataset

	Καλύτερη λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	78.9408	78.9408	78.9408		5.66	32720
HBMO medoid	83,9599	83,9599	83,9599		8.4351	57916

Πίνακας 18 - Αποτελέσματα στο Iris dataset για την διακριτή αναπαράσταση λύσης

9.2. Αποτελέσματα στο Wine Dataset

	Καλύτερη λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	2.370.689,6	2.370.693,1	2.370.723,6		59,47	438710
ABC	2.370.690,0	2.370.690,0	2.370.690,0		8,11	50012
BBMO	2.370.689,8	2.370.690,0	2.370.690,1		63,34	348030
K-means	2.370.689,6	2.396.552,2	2.629.315,1		0.004	-

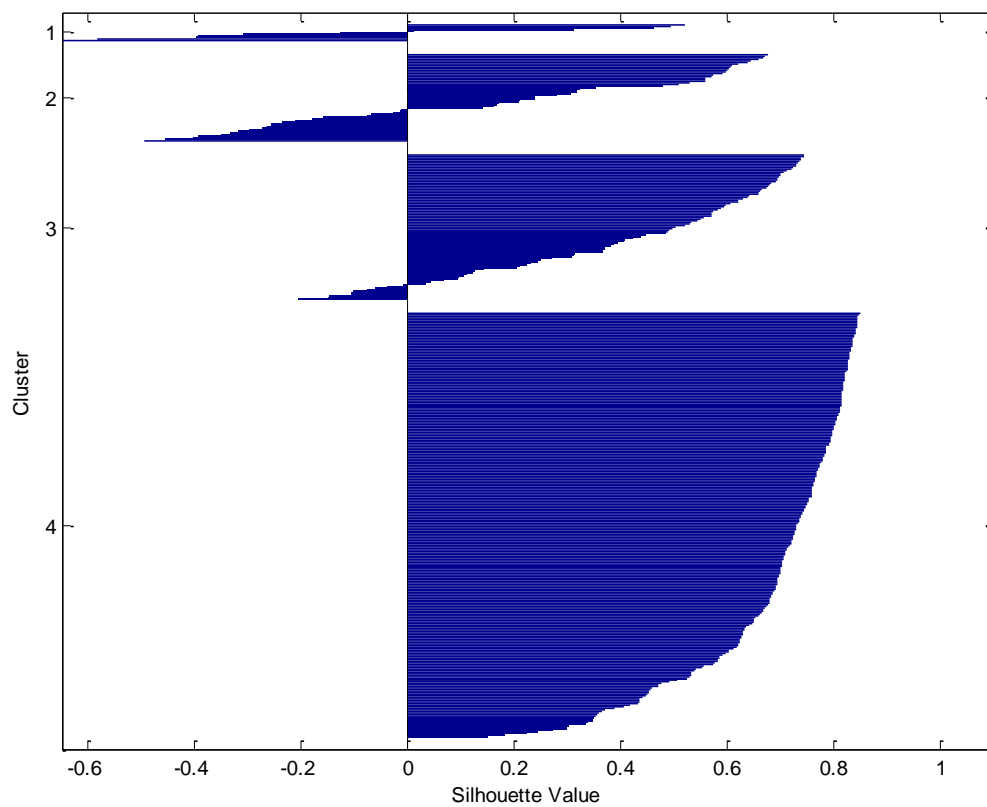
Πίνακας 19 – Αποτελέσματα στο Wine Dataset

	Καλύτερη λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	2370689,6	2370689,6	2370689,6		4.58	16650
HBMO medoid	2388935,3	2388935,3	2388935,3		9.38	50250

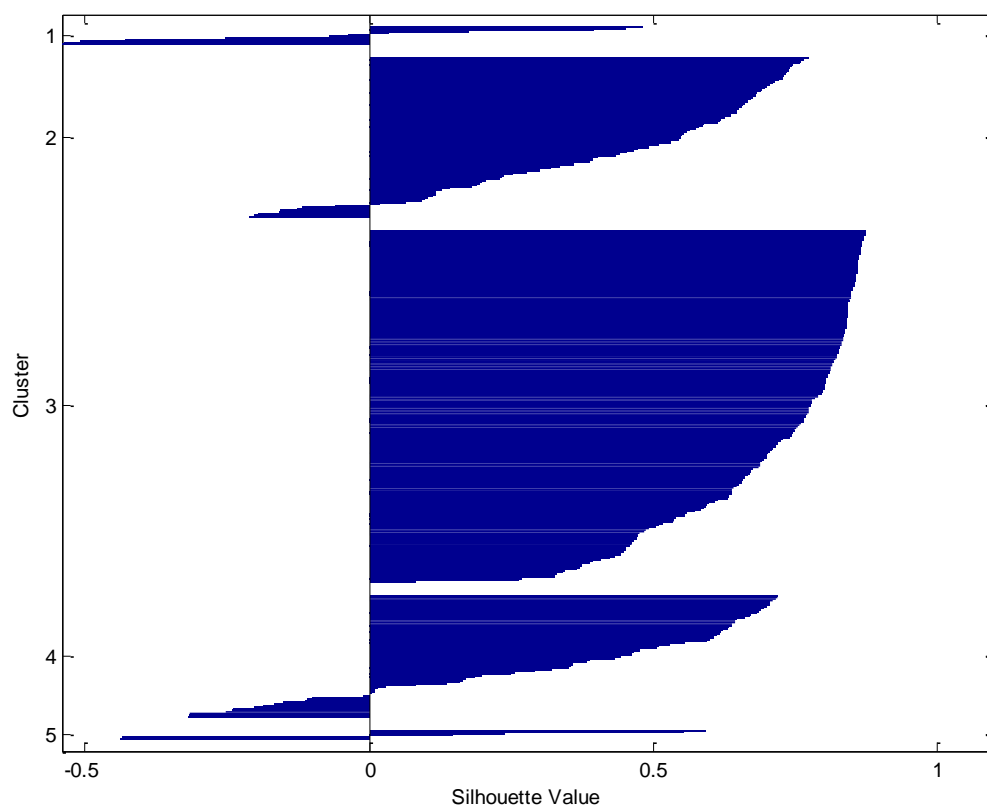
Πίνακας 20-Αποτελέσματα στο Wine dataset για την διακριτή αναπαράσταση λύσης

9.3. Αποτελέσματα στο Wholesale Customer Dataset

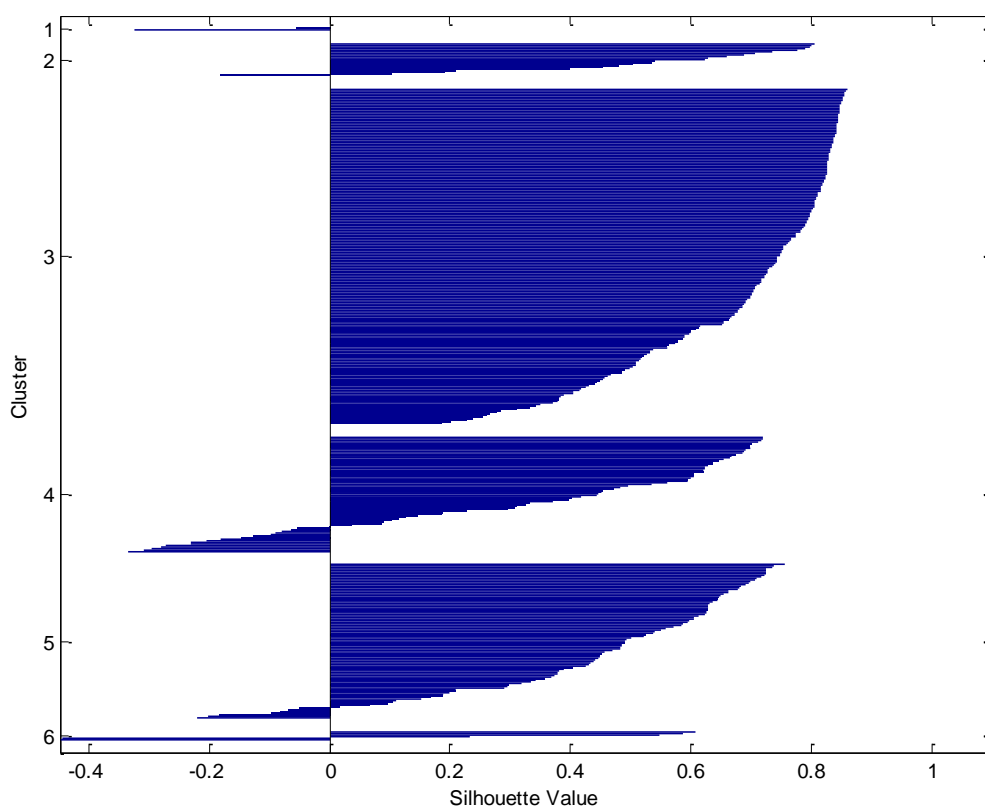
Στο συγκεκριμένο σετ δεδομένων, σε αντίθεση με τα προηγούμενα, δεν είναι γνωστός ο αριθμός των τμημάτων. Για αυτόν τον λόγο προτού προχωρήσουμε στην εφαρμογή των αλγορίθμων, πρέπει να ορίσουμε τον κατάλληλο αριθμό τμημάτων. Όπως έχει αναφερθεί και σε προηγούμενο κεφάλαιο, ένας απλός τρόπος εκτίμησης των αριθμών των τμημάτων είναι χρησιμοποιώντας τη συνάρτηση Silhouette. Στην παρούσα εργασία χρησιμοποιήθηκε η υλοποίηση της Matlab για την εν λόγω συνάρτηση για 3 έως 10 τμήματα, βασιζόμενα στην καλύτερη λύση από 10 τρεξίματα του K-means. Με άλλα λόγια για k τμήματα εκτελέσαμε τον K-means 10 φορές, επιλέξαμε την καλύτερη λύση και πάνω σε αυτή εφαρμόσαμε τη συνάρτηση.



Εικόνα 16 – Διάγραμμα Silhouette για 4 τμήματα



Εικόνα 17 - Διάγραμμα Silhouette για 5 τμήματα



Εικόνα 18 - Διάγραμμα Silhouette για 6 τμήματα

Παρατηρούμε ότι στα 4 και στα 6 τμήματα η συνάρτηση silhouette λαμβάνει τιμές μεγαλύτερες του 0.5 για κάθε τμήμα (αν και υπάρχουν και αρνητικές τιμές), πράγμα που αποτελεί ένδειξη ότι τα 4 ή 6 τμήματα είναι αρκετά διαχωρίσιμα. Παρόλα αυτά, δοκιμάζουμε τις τιμές 4, 5 και 6 μιας και εξετάζουμε την επίδοση των αλγορίθμων και όχι το πόσο καλά εκτιμούμε τον αριθμό των cluster.

Για 4 τμήματα λάβαμε τα ακόλουθα αποτελέσματα:

	Καλύτερη Λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	64.855.378.336	65.347.557.616	67.315.940.100		171,78	694425
ABC	64.855.545.654	64.857.752.829	64.863.435.058		14,55	50016
BBMO	64.855.847.653	65.854.504.717	67.458.884.827		102,2	348030
K-means	64.855.545.528	66.856.114.758	67.496.056.746		0,05	-

Πίνακας 21 – Αποτελέσματα στο WC dataset για 4 clusters

	Καλύτερη λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	64.855.545.528	65.101.584.985	67.315.940.100		29.52	54893

HBMO medoid	68.713.543.237	68.741.547.111	68.993.581.977		39.82	155160
-------------	----------------	----------------	----------------	--	-------	--------

Πίνακας 22 - Αποτελέσματα στο WC dataset για την διακριτή αναπαράσταση λύσης για 4 clusters

Για 5 τμήματα τα ακόλουθα:

	Καλύτερη Λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	52.928.148.942	52.987.170.263	53.205.171.749		190,36	649170
ABC	52.930.137.089	52.972.101.556	53.006.584.003		17,47	50017
BBMO	52.928.509.860	53.035.228.697	53.206.418.880		117,01	348030
K-means	52.928.148.942	54.520.506.637	61.069.479.432		0,02	-

Πίνακας 23 - Αποτελέσματα στο WC dataset για 5 clusters

	Καλύτερη λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	52.928.148.942	52.989.575.973	53.004.932.730		31.51	66836
HBMO medoid	55.914.172.877	55.934.256.000	56014588493		94.52	228510

Πίνακας 24-Αποτελέσματα στο WC dataset για την διακριτή αναπαράσταση λύσης για 5 clusters

Για 6 τμήματα τα ακόλουθα:

	Καλύτερη Λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	46.489.839.347	47.127.119.787	47.559.633.578		217,30	653473
ABC	46.493.237.688	46.686.120.349	47.160.980.667		20,41	50015
BBMO	46.490.713.725	47.088.078.114	48.150.617.260		130,87	348030
K-means	47.257.645.388	47.391.052.152	48.120.758.727		0,02	-

Πίνακας 25 - Αποτελέσματα στο WC dataset για 6 clusters

	Καλύτερη λύση	Μέση τιμή λύσεων	Χειρότερη λύση		Μέσος Χρόνος	Μέσος αριθμός Evaluations
HBMO	46.489.839.347	46.585.693.073	46.986.441.596		36.48	88339
HBMO medoid	50.545.876.829	50.599.773.745	50.955.904.231		102.21	301850

Πίνακας 26 - Αποτελέσματα στο WC dataset για την διακριτή αναπαράσταση λύσης για 6 clusters

10. Συμπεράσματα

Στην παρούσα εργασία μελετήσαμε τρεις αλγορίθμους βελτιστοποίησης οι οποίοι βασίζονται σε συμπεριφορά μελισσών στο πρόβλημα της τμηματοποίησης δεδομένων. Αρχικά παρουσιάσαμε συνοπτικά την βιβλιογραφία γύρω από την τμηματοποίηση, τα κριτήρια και τις μεθόδους που χρησιμοποιούνται τόσο από ερευνητικές ομάδες όσο και από εταιρίες. Μελετήσαμε πολλές παραλλαγές των αλγορίθμων με διαφορετικές διαδικασίες τοπικής αναζήτησης και εξέλιξης. Μελετήσαμε επίσης πολλούς τρόπους αναπαράστασης της λύσης. Κατόπιν, αναπτύξαμε ένα απλό σύστημα ρύθμισης παραμέτρων των αλγορίθμων μιας και αυτή η διαδικασία μπορεί να είναι ιδιαίτερα επίπονη για τον ερευνητή και στη συνέχεια για τις βέλτιστες τιμές παραμέτρων μελετήσαμε την επίδοση των αλγορίθμων σε άλλα σετ δεδομένων.

Πιο συγκεκριμένα, στο σετ δεδομένων Iris, από τους αλγόριθμους συνεχούς αναπαράστασης ο αλγόριθμος ABC έδειξε την καλύτερη επίδοση, καθώς βρήκε την χαμηλότερη τιμή από όλους σε μικρό χρονικό διάστημα. Ακόμη και σε σύγκριση με τον K-means ο αλγόριθμος κατάφερε να βρει το βέλτιστο περισσότερες φορές από τον K-means. Στην διακριτή αναπαράσταση ο HBMO είχε καλύτερη επίδοση από τον ABC καθώς βρήκε το βέλτιστο και τις 10 φορές και με λιγότερα function evaluations. Με άλλα λόγια ο διακριτός αλγόριθμος HBMO λειτουργεί πολύ αποδοτικά καθώς βρήκε και τις 10 φορές την βέλτιστη τιμή σε σχετικά μικρό μέσο χρόνο. Ο αλγόριθμος BB λειτουργεί και αυτός αρκετά αποδοτικά, παρόλο που βρίσκει χειρότερη μέση λύση από τους υπόλοιπους και χρειάζεται περισσότερο χρόνο. Πρέπει να σημειωθεί σε αυτό το σημείο ότι ο αλγόριθμος K-means είχε χειρότερη μέση λύση και από τους 3, αλλά ο χρόνος που χρειάζεται είναι εξαιρετικά μικρότερος από τους υπόλοιπους. Τέλος η αναπαράσταση λύσης μέσω medoid έδωσε και αυτή μία «κακή» μέση τιμή λύσης όσον αφορά την αντικειμενική συνάρτηση. Παρόλα αυτά πρέπει να τονίσουμε ότι μέσω της συγκεκριμένης αναπαράστασης εφαρμόζουμε έναν πολύ αυστηρό περιορισμό που επιβάλλει το κέντρο του τμήματος να είναι ένα από τα αντικείμενα. Αυτός είναι και ο λόγος που δεν είναι εφικτό να πιάσει την χαμηλότερη τιμή όπως οι υπόλοιποι αλγόριθμοι, στους οποίους επιτρέπεται το κέντρο να βρίσκεται οπουδήποτε στο χώρο των λύσεων. Επομένως η τιμή στην οποία συγκλίνει ο HBMO – medoid ίσως να είναι και η βέλτιστη.

Στο σετ δεδομένων Wine την καλύτερη μέση τιμή λύσεων για τη συνεχή αναπαράσταση είχε ο αλγόριθμος ABC ο οποίος μάλιστα λειτουργεί σε εξαιρετικά χαμηλότερο χρόνο από τους υπόλοιπους. Δεύτερος έρχεται ο HBMO ο οποίος βρίσκει πολύ κοντινή λύση με τον ABC αλλά χρειάστηκε πολύ περισσότερο χρόνο. Πρέπει να τονιστεί όμως, ότι ο HBMO βρήκε καλύτερη λύση από τον ABC, αν και κατά μέση τιμή ήταν λίγο λιγότερο αποδοτικός. Τέλος ο αλγόριθμος BBMO παρόλο που βρήκε κάποιες φορές τη βέλτιστη λύση που βρήκε και ο ABC, η μέση τιμή των λύσεων ήταν αρκετά χαμηλότερη. Όσον αφορά

τώρα την διακριτή αναπαράσταση, ο HBMO βρήκε την καλύτερη λύση σε πολύ μικρό χρονικό διάστημα και με πολύ λίγα function evaluations. Πολύ καλή επίδοση είχε και ο HBMO-medoid, αν και ισχύουν και εδώ όσα αναφέρθηκαν στην προηγούμενη παράγραφο.

Τέλος, στο σετ δεδομένων Wholesale Customers τρέξαμε τους αλγόριθμους για τρεις διαφορετικούς αριθμούς τμημάτων. Για 4 τμήματα ο αλγόριθμος ABC είχε καλύτερη μέση λύση από όλους τους αλγόριθμους και σε σχετικά μικρό χρόνο. Η καλύτερη λύση βρέθηκε από τον αλγόριθμο HBMO του οποίου η απόδοση ήταν καλή και σαν μέση τιμή λύσεων. Ο αλγόριθμος BB βρήκε μία σχετικά καλή λύση αλλά η μέση τιμή ήταν αρκετά χειρότερη από τους υπόλοιπους. Τέλος, ο K-means είχε την χειρότερη μέση τιμή λύσεων, αλλά σε πολύ μικρότερο χρόνο από τους υπόλοιπους. Στην διακριτή αναπαράσταση ο HBMO δεν είχε τόσο καλή επίδοση, ενώ ακόμη χειρότερη είχε ο HBMO-medoid.

Για την περίπτωση των 5 τμημάτων παρατηρούμε παρόμοια συμπεριφορά των αλγορίθμων. Δηλαδή, την καλύτερη μέση τιμή λύσεων την είχε ο ABC ενώ την χειρότερη ο K-means. Την καλύτερη λύση την βρήκε ο HBMO αλλά και ο K-means. Ο αλγόριθμος BBMO λειτούργησε χειρότερα από τους άλλους κατά μέση τιμή αλλά βρήκε πολύ καλή λύση σε κάποιο ή κάποια από τα τρεξίματά του. Ο K-means είχε την χειρότερη μέση τιμή λύσεων από όλους τους αλγορίθμους. Όσον αφορά τους διακριτούς, ο HBMO είχε σχετικά καλή επίδοση μιας και βρήκε την καλύτερη λύση, αλλά είχε χειρότερη μέση τιμή λύσεων από τον ABC.

Για την περίπτωση των 6 τμημάτων, την καλύτερη μέση τιμή την παρουσιάζει και πάλι ο ABC ενώ την καλύτερη λύση την βρήκε και πάλι ο HBMO. Ο αλγόριθμος K-means λειτούργησε χειρότερα ως προς την καλύτερη λύση που βρήκε, ενώ ο BB λειτούργησε χειρότερα ως προς τη μέση τιμή λύσεων. Στην διακριτή αναπαράσταση ο HBMO λειτούργησε εξαιρετικά καθώς είχε την καλύτερη λύση, την καλύτερη μέση τιμή λύσεων καθώς και την μικρότερη χειρότερη λύση. Αυτό αποτελεί ένδειξη ότι για μεγάλο αριθμό τμημάτων, ο συγκεκριμένος αλγόριθμος μπορεί να λειτουργήσει πολύ αποδοτικά. Ο HBMO-medoid δεν είχε τόσο καλή επίδοση όσον αφορά την τιμή της αντικειμενικής συνάρτησης,

11. Προτάσεις για μελλοντική έρευνα

Η παρούσα εργασία αποτελεί μία πρώτη ανάλυση της επίδοσης των αλγορίθμων μελισσών στο πρόβλημα της τμηματοποίησης. Μελετήθηκαν πολλές παραλλαγές των αλγορίθμων, εξετάστηκε σε βάθος η επίδραση των παραμέτρων και αξιολογήθηκαν τρεις διαφορετικοί τρόποι αναπαράστασης της λύσης. Επίσης εφαρμόστηκαν οι τελεστές μετάλλαξης και ο τελεστής ΚΜΟ σε αλγόριθμους μελισσών το οποίο δεν υπήρχε στην αντίστοιχη βιβλιογραφία. Παρόλα αυτά, υπάρχουν προτάσεις για περαιτέρω έρευνα, η οποία θα εμβραθύνει ακόμη περισσότερο στον συγκεκριμένο τομέα.

Πιο συγκεκριμένα, δεν μελετήθηκε η επίδοση των αλγορίθμων όταν η αντικειμενική συνάρτηση είναι διαφορετική από αυτήν που βασίζεται στην Ευκλείδεια απόσταση. Μιας και η συγκεκριμένη αντικειμενική συνάρτηση καλύπτει μόνο ένα μέρος από τις πιθανές τοπολογίες των δεδομένων, θα ήταν ιδιαίτερα χρήσιμο να αξιολογηθούν και άλλες αντίστοιχες συναρτήσεις.

Μία άλλη πτυχή του προβλήματος η οποία δεν μελετήθηκε στην παρούσα εργασία είναι το τι ακριβώς συμβαίνει στις περιπτώσεις όπου τα δεδομένα αποτελούνται από πραγματικές, κέραιες ή/και δυαδικές τιμές ταυτόχρονα. Δεδομένου ότι είναι πολύ πιθανόν τα δεδομένα της καταναλωτικής αγοράς να περιλαμβάνουν τέτοιου είδους δεδομένα (για παράδειγμα το φύλο, τον αριθμό μελών της οικογένειας κ.ά.) είναι ιδιαίτερα χρήσιμο να αναπτυχθεί μία μεθοδολογία η οποία να μπορεί να το χειριστεί με επαρκή αποτελεσματικότητα.

Η γνώση του αριθμού των τμημάτων στις υλοποιήσεις των αλγορίθμων στην παρούσα εργασία είναι ένας δύσκολος περιορισμός, ο οποίος περιορίζει τις δυνατότητες του αλγορίθμου. Προς αυτήν λοιπόν την κατεύθυνση, είναι ιδιαίτερα ενδιαφέρον το να γίνει ενοποίηση των αλγορίθμων που εντοπίζουν τον αριθμό των τμημάτων με τους αλγόριθμους που κάνουν την τμηματοποίηση σε ένα ενιαίο εργαλείο.

Τέλος, δεδομένου ότι η βιβλιογραφία είναι αρκετά μεγάλη στο εν λόγω θέμα, είναι απαραίτητη η ανάπτυξη ενός πιο γενικού πλαισίου γύρω από το οποίο θα αξιολογούνται οι αλγόριθμοι. Ένα τέτοιο παράδειγμα είναι ένα λογισμικό στο οποίο θα μπορεί ο χρήστης να επιλέγει την μέθοδο τμηματοποίησης, τον αλγόριθμο καθώς και την αντικειμενική συνάρτηση και να κρίνει ο ίδιος ποια αποτελέσματα είναι καλύτερα με βάση την απόφαση που πρέπει να ληφθεί.

Βιβλιογραφία

- [1] S. Dibb, “Market segmentation: strategies for success,” *Mark. Intell. Plan.*, vol. 16, no. 7, pp. 394–406, 1998.
- [2] D. S. Boone and M. Roehm, “Retail segmentation using artificial neural networks,” *Int. J. Res. Mark.*, vol. 19, no. 3, pp. 287–301, 2002.
- [3] Y. Wind, “Issues and advances in segmentation research,” *J. Mark. Res.*, pp. 317–337, 1978.
- [4] Π. Μάλλιαρης, *Εισαγωγή στο Μάρκετινγκ*, 3rd ed. Σταμούλης, 2001.
- [5] M. Wedel and W. A. Kamakura, “Introduction to the special issue on market segmentation,” *Intern. J. Res. Mark.*, vol. 19, pp. 181–183, 2002.
- [6] Γ. Σιώμκος, *Συμπεριφορά καταναλωτή & Στρατηγική Μάρκετινγκ*. 2002, p. 325.
- [7] S. Young, L. Ott, and B. Feigin, “Some practical considerations in market segmentation,” *J. Mark. Res.*, pp. 405–412, 1978.
- [8] P. Kotler, *Μάρκετινγκ-Μάνατζμεντ: Ανάλυση, Σχεδιασμός, Υλοποίηση και Έλεγχος*. 2000.
- [9] T. Dalgic and M. Leeuw, “Niche marketing revisited: concept, applications and some European cases,” *Eur. J. Mark.*, vol. 28, no. 4, pp. 39–55, 1994.
- [10] O. C. Walker, H. W. Boyd, and J.-C. Larreche, *Marketing strategy: Planning and implementation*. Irwin Chicago, IL, 1996.
- [11] G. J. Hooley, J. A. Saunders, and N. Piercy, *Marketing strategy and competitive positioning*. Pearson Education, 2004.
- [12] R. I. Haley, “Benefit segmentation: a decision-oriented research tool,” *J. Mark.*, pp. 30–35, 1968.
- [13] T. P. Beane and D. M. Ennis, “Market segmentation: a review,” *Eur. J. Mark.*, vol. 21, no. 5, pp. 20–42, 1987.
- [14] S.-I. Wu, “Benefit segmentation: an empirical study for on-line marketing,” *Asia Pacific J. Mark. Logist.*, vol. 13, no. 4, pp. 3–18, 2001.
- [15] S. Dibb and L. Simkin, “Market segmentation: diagnosing and treating the barriers,” *Ind. Mark. Manag.*, vol. 30, no. 8, pp. 609–625, 2001.
- [16] Π. Τομαράς, *Βιομηχανικό Μάρκετινγκ: B to B Marketing*. Σταμούλης, 2009.
- [17] O. Maimon and L. Rokach, *Data mining and knowledge discovery handbook*, vol. 2. Springer, 2005.
- [18] H. Duda, P. Hart, and others, “Stork, Pattern Classification.” John Wiley & Sons, 2001.
- [19] C. Fraley and A. E. Raftery, “How many clusters? Which clustering method? Answers via model-based cluster analysis,” *Comput. J.*, vol. 41, no. 8, pp. 578–588, 1998.

- [20] H. Jiawei and M. Kamber, "Data mining: concepts and techniques," *San Fr. CA, itd Morgan Kaufmann*, vol. 5, 2001.
- [21] P. H. A. Sneath, R. R. Sokal, and others, *Numerical taxonomy. The principles and practice of numerical classification*. 1973.
- [22] B. King, "Step-wise clustering procedures," *J. Am. Stat. Assoc.*, vol. 62, no. 317, pp. 86–101, 1967.
- [23] F. Murtagh, "A survey of recent advances in hierarchical clustering algorithms," *Comput. J.*, vol. 26, no. 4, pp. 354–359, 1983.
- [24] S. Guha, R. Rastogi, and K. Shim, "CURE: an efficient clustering algorithm for large databases," in *ACM SIGMOD Record*, 1998, vol. 27, no. 2, pp. 73–84.
- [25] I. S. Dhillon and D. S. Modha, "Concept decompositions for large sparse text data using clustering," *Mach. Learn.*, vol. 42, no. 1–2, pp. 143–175, 2001.
- [26] C. T. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters," *Comput. IEEE Trans.*, vol. 100, no. 1, pp. 68–86, 1971.
- [27] J. D. Banfield and A. E. Raftery, "Model-based Gaussian and non-Gaussian clustering," *Biometrics*, pp. 803–821, 1993.
- [28] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. R. Stat. Soc. Ser. B (Statistical Methodol.)*, vol. 63, no. 2, pp. 411–423, 2001.
- [29] S. Ray and R. H. Turi, "Determination of number of clusters in k-means clustering and application in colour image segmentation," in *Proceedings of the 4th international conference on advances in pattern recognition and digital techniques*, 1999, pp. 137–143.
- [30] K. Do-Jong, P. Yong-Woon, and P. Dong-Jo, "A novel validity index for determination of the optimal number of clusters," *IEICE Trans. Inf. Syst.*, vol. 84, no. 2, pp. 281–285, 2001.
- [31] C. L. (Chung L. Liu, *Introduction to combinatorial mathematics*. New York : McGraw-Hill, 1968.
- [32] H. A. Abbass, "An agent based approach to 3-SAT using marriage in honey-bees optimization," *Int. J. Knowl. BASED Intell. Eng. Syst.*, vol. 6, no. 2, pp. 64–71, 2002.
- [33] A. Afshar, O. B. Haddad, M. A. Mariño, and B. J. Adams, "Honey-bee mating optimization (HBMO) algorithm for optimal reservoir operation," *J. Franklin Inst.*, vol. 344, no. 5, pp. 452–462, 2007.
- [34] H. A. Abbass, "A single queen single worker honey bees approach to 3-sat," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2001, pp. 807–814.
- [35] M. Fathian, B. Amiri, and A. Maroosi, "Application of honey-bee mating optimization algorithm on clustering," *Appl. Math. Comput.*, vol. 190, no. 2, pp. 1502–1513, 2007.
- [36] O. B. Haddad, A. Afshar, and M. A. Mariño, "Honey-bees mating optimization (HBMO) algorithm: a new heuristic approach for water resources optimization," *Water Resour. Manag.*, vol. 20, no. 5, pp. 661–680, 2006.
- [37] M. Marinaki, Y. Marinakis, and C. Zopounidis, "Honey bees mating optimization algorithm for financial classification problems," *Appl. Soft Comput.*, vol. 10, no. 3, pp. 806–812, 2010.

- [38] Y. Marinakis, M. Marinaki, and G. Dounias, "Honey bees mating optimization algorithm for the vehicle routing problem," in *Nature inspired cooperative strategies for optimization (NICSO 2007)*, Springer, 2008, pp. 139–148.
- [39] Y. Marinakis, M. Marinaki, and G. Dounias, "Honey bees mating optimization algorithm for large scale vehicle routing problems," *Nat. Comput.*, vol. 9, no. 1, pp. 5–27, 2010.
- [40] Y. Marinakis, M. Marinaki, and N. Matsatsinis, "A Bumble bees mating optimization algorithm for global unconstrained optimization problems," in *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*, Springer, 2010, pp. 305–318.
- [41] Y. Marinakis, M. Marinaki, and N. Matsatsinis, "A hybrid clustering algorithm based on honey bees mating optimization and greedy randomized adaptive search procedure," in *Learning and Intelligent Optimization*, Springer, 2008, pp. 138–152.
- [42] Y. Marinakis, M. Marinaki, and N. Matsatsinis, "Honey bees mating optimization for the location routing problem," in *Engineering Management Conference, 2008. IEMC Europe 2008. IEEE International*, 2008, pp. 1–5.
- [43] J. Teo and H. A. Abbass, "A true annealing approach to the marriage in honey-bees optimization algorithm," *Int. J. Comput. Intell. Appl.*, vol. 3, no. 02, pp. 199–211, 2003.
- [44] Y. Marinakis, M. Marinaki, and N. Matsatsinis, "A hybrid bumble bees mating optimization-grasp algorithm for clustering," in *Hybrid Artificial Intelligence Systems*, Springer, 2009, pp. 549–556.
- [45] Y. Marinakis and M. Marinaki, "A bumble bees mating optimization algorithm for the open vehicle routing problem," *Swarm Evol. Comput.*, vol. 15, pp. 80–94, 2014.
- [46] D. Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 652–657, Jan. 2011.
- [47] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, Aug. 2009.
- [48] E. R. Hruschka, R. J. G. B. Campello, A. A. Freitas, and A. C. P. L. F. de Carvalho, "A survey of evolutionary algorithms for clustering," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 39, pp. 133–155, 2009.
- [49] E. Falkenauer, *Genetic algorithms and grouping problems*. John Wiley & Sons, Inc., 1998.
- [50] K. Krishna and M. N. Murty, "Genetic K-means algorithm," *Syst. Man, Cybern. Part B Cybern. IEEE Trans.*, vol. 29, no. 3, pp. 433–439, 1999.
- [51] "Machine Learning Laboratory." [Online]. Available: <http://archive.ics.uci.edu/ml/>.

Παράρτημα

Σετ δεδομένων Aggregator

15.55	28.65	2	9	25.85	2	5.6	20.75	2	9.85	20.7	2	6.1	11.75	7
14.9	27.55	2	8.6	25.65	2	5.8	21.95	2	9.35	20.6	2	7	12.35	7
14.45	28.35	2	9.4	25.55	2	6.4	21.95	2	9.25	19.65	2	7.05	12.45	7
14.15	28.8	2	8.45	25.05	2	6.55	21.15	2	9.95	19.8	2	7.9	12.5	7
13.75	28.05	2	8.85	24.6	2	7.45	21.95	2	10.7	20.35	2	8.55	12.1	7
13.35	28.45	2	9.65	24.7	2	7.4	21.55	2	11.3	20.7	2	7.85	11.85	7
13	29.15	2	10.55	24.35	2	7.75	21.2	2	12.35	21.6	2	7.1	11.95	7
13.45	27.5	2	11.05	23.9	2	7.65	20.65	2	13.1	21.3	2	6.9	11.5	7
13.6	26.5	2	10.55	23.55	2	6.95	19.8	2	12.85	20.75	2	6.85	10.9	7
12.8	27.35	2	9.45	23.35	2	6.6	20.1	2	12	20	2	6.4	10.7	7
12.4	27.85	2	9.2	23.9	2	6.05	20.2	2	11	19.85	2	5.9	10.3	7
12.3	28.4	2	8.35	23.9	2	5.4	19.65	2	10.35	19	2	6.4	10.25	7
12.2	28.65	2	7.35	24.75	2	5.35	19.05	2	9.9	18.65	2	7.05	10.05	7
13.4	25.1	2	7.4	25.45	2	5.8	18.25	2	10.6	18.15	2	7.35	10.5	7
12.95	25.95	2	6.6	25.75	2	6.3	19.1	2	11.4	18.3	2	7.65	11.1	7
12.9	26.5	2	6.1	26	2	7	18.9	2	11.4	19.25	2	8.1	11.2	7
11.85	27	2	5.8	26.95	2	7.15	17.9	2	12.35	18.8	2	8.8	11.4	7
11.35	28	2	5.65	25.8	2	7.35	18.2	2	12.8	19.75	2	8.3	10.55	7
11.15	28.7	2	5.3	26.1	2	8.2	20.05	2	12.15	18.1	2	9	10.9	7
11.25	27.4	2	6.4	25.4	2	8.3	19.45	2	11.05	17.5	2	9.35	10.5	7
10.75	27.7	2	5.4	25.25	2	8.3	18.5	2	11.95	17.25	2	10.15	11	4
10.5	28.35	2	5.35	24.7	2	8.75	18.8	2	12.25	17.5	2	10.4	10.55	4
9.65	28.45	2	4.8	25.05	2	9.05	18.2	2	13.05	17.4	2	10.9	10	4
10.25	27.25	2	4.2	25.55	2	9.35	17.7	2	13.75	18.15	2	11.55	10.2	4
10.75	26.55	2	6.4	24.8	2	8.9	17.65	2	13.5	18.65	2	11.75	10.85	4
11.7	26.35	2	6.55	24.3	2	8.45	17.2	2	13.65	19.25	2	10.1	8.65	4
11.6	25.9	2	7.4	24.25	2	10.05	17.2	2	14	19.9	2	11.05	9.1	4
11.9	25.05	2	5.45	24.2	2	10.4	16.75	2	15.2	18.2	2	11.85	9.8	4
12.6	24.05	2	4.3	24	2	8.6	20.9	2	15.5	17.15	2	12.85	10.65	4
11.9	24.5	2	4	24.25	2	8.65	21.3	2	13.9	17.1	2	12.9	11.7	4
11.1	25.2	2	3.35	23.3	2	8.65	21.9	2	13.75	16.6	2	13.6	11.1	4
10.55	25.15	2	4.85	23.05	2	8.65	22.5	2	12.15	16.4	2	14.05	11.75	4
10.05	25.95	2	4.3	22.75	2	8.95	22.8	2	7.8	13.7	7	14.5	11.8	4
9.35	26.6	2	5.85	23.4	2	9.95	22.65	2	8.85	13.35	7	14.3	12.45	4
9.3	27.25	2	5.9	23.55	2	8.95	22.2	2	9	12.7	7	17	12.9	4
9.2	27.8	2	7.55	23.7	2	9.65	21.9	2	9.7	12.1	7	15.8	12.6	4
7.5	28.25	2	6.85	23.25	2	10.55	22.3	2	8.05	12.9	7	15.85	12	4
8.55	27.45	2	7.65	23.1	2	10.9	22.85	2	7.7	13.25	7	16.7	12.2	4
8.5	27.05	2	6.95	22.55	2	11.35	23.45	2	6.8	13.2	7	16.25	11.7	4
8.05	27.2	2	6.1	22.6	2	12.05	23.4	2	6.6	13.45	7	15.55	11.15	4
7.85	26.8	2	5.5	22.6	2	12.3	22.75	2	6.2	12.55	7	14.8	11.35	4
7.3	27.4	2	4.7	22.1	2	11.7	22.15	2	5.4	12.85	7	14.45	10.75	4
6.8	26.85	2	3.8	21.85	2	11.15	22.05	2	5.7	12.25	7	13.75	10.45	4
7	26.5	2	4.65	21.2	2	10.85	21.5	2	5.2	11.9	7	12.8	10.1	4
7.55	26.3	2	4.15	20.35	2	10.85	21.05	2	5.15	11.35	7	13.15	9.8	4
8.55	26.3	2	5.3	20.4	2	9.6	21.3	2	5.85	11.2	7	12.45	9.3	4

11.8	8.95	4	13.65	5.9	4	17.05	3.7	4	18.7	5.75	4	18.85	9.05	4
11.1	8.45	4	13.9	5.3	4	17.25	3.05	4	19.25	5.95	4	18.75	9.55	4
10.35	7.7	4	13.1	5.1	4	16.65	2.8	4	18.4	6	4	18.6	10	4
10.1	6.75	4	12.55	4.9	4	16.55	2.15	4	18.45	6.6	4	16.95	10.35	4
11.3	7.95	4	11.5	4.75	4	17.2	2.05	4	17.65	7.05	4	17.35	10.85	4
12.35	8.45	4	11.35	4.05	4	18.15	1.95	4	16.7	7.4	4	18	10.65	4
13.1	8.95	4	12.4	4.35	4	18.05	2.45	4	18.65	7.3	4	18.5	10.55	4
13.2	9.35	4	11.75	3.45	4	18.15	3.05	4	18.05	7.35	4	18.1	11.1	4
14.1	10.05	4	12.65	3.7	4	18.6	3.45	4	17.85	7.75	4	17.55	11.3	4
11.5	7.5	4	13.4	4.35	4	18.4	3.6	4	17.5	8.25	4	17.95	11.9	4
11.35	6.9	4	13.9	4.95	4	18.85	3.2	4	17.15	8.6	4	18.3	12	4
11.95	6.75	4	12.75	3	4	19.1	2.65	4	17.05	9	4	18	12.5	4
12.4	7.1	4	13.55	3.15	4	19.45	2.65	4	16.4	8.7	4	19	11.65	4
12.25	7.6	4	13.7	3.65	4	19	2.1	4	16.05	8.95	4	19.5	11.05	4
12.95	7.6	4	14.1	4.1	4	19.9	2.05	4	16.05	9.6	4	19.45	10.55	4
13.45	7.95	4	14.65	5.05	4	20.45	2.8	4	16.5	9.75	4	19.4	9.65	4
13.35	8.25	4	14.35	5.75	4	19.8	3.25	4	17.25	9.6	4	20.1	9.4	4
13.75	9	4	14.5	6.55	4	19.45	3.9	4	17.6	9.9	4	20.05	9.95	4
14.3	9.3	4	15.15	7.1	4	18.65	4.2	4	17.8	9.3	4	20.05	10.2	4
14.85	9.55	4	13.6	2.55	4	18.4	4.6	4	18	8.55	4	19.35	12.2	4
15.1	10.25	4	14.45	2.4	4	18.65	4.75	4	18.8	8.1	4	19.2	12.25	4
15.45	10.55	4	14.6	3.05	4	18.75	5.15	4	18.8	8.35	4	20.05	11.6	4
16.35	10.85	4	15	3.4	4	19.1	4.55	4	19.4	7.6	4	20.6	11.15	4
16.75	11.5	4	15.25	3.5	4	17.9	5.4	4	19.25	6.6	4	20.7	10.65	4
16.25	10.2	4	14.7	4.1	4	17.65	5.7	4	20.05	6.95	4	21.3	11.65	4
15.4	10.1	4	14.7	4.5	4	17.05	6.05	4	19.8	7.5	4	21.8	11.15	4
15.45	9.7	4	15.25	2.7	4	17.4	6.5	4	20.05	6.35	4	21.85	10.7	4
15.15	9.3	4	15.65	2.05	4	16.6	6.85	4	21.15	5.7	4	21.65	10.05	4
15.25	8.65	4	15.95	2.8	4	15.7	7.15	4	21.65	4.85	4	20.95	10.2	4
15.55	8.2	4	16.1	3.55	4	15.75	7.75	4	22.15	4.35	4	20.9	9.7	4
14.25	8.7	4	15.9	4	4	16.6	7.95	4	23.05	3.35	4	21.65	9.45	4
14.25	8.25	4	15.6	4.75	4	20.4	3.4	4	23.05	3.8	4	21.2	9.25	4
15.05	7.8	4	15.55	5.05	4	20.7	3.45	4	23.15	4.4	4	20.75	8.75	4
14.3	7.5	4	15.35	5.5	4	21.15	2.85	4	22.5	4.75	4	20.55	8.75	4
13.55	7.45	4	15.15	5.95	4	21.75	2.65	4	22.15	5.2	4	21.1	8	4
14.3	6.95	4	15.5	6.75	4	22	3.25	4	24.15	4.55	4	21.65	8.65	4
13.95	6.7	4	15.7	6.35	4	22.2	3.5	4	23.5	5.05	4	21.75	8.2	4
13.05	6.95	4	16.2	5.9	4	21.45	3.75	4	23.1	5.3	4	21.95	7.55	4
13.05	6.2	4	16.35	5.35	4	21.1	4.05	4	23	5.75	4	22	6.75	4
11.55	6.3	4	16.2	4.55	4	20.15	4.3	4	22.2	5.75	4	22.8	6.45	4
10.8	5.85	4	16.55	4.2	4	20.8	4.7	4	21.85	6.2	4	22.65	6.65	4
10.6	5.05	4	16.95	4.75	4	20.7	5.15	4	20.75	6.55	4	22.75	7.05	4
11.35	5.55	4	17.05	5.1	4	19.75	5.05	4	21	7.15	4	23	7.35	4
12.15	5.4	4	17.3	4.8	4	19.85	5.5	4	20.75	7.65	4	22.55	7.9	4
12.4	5.8	4	17.3	4.15	4	20.4	5.65	4	20	8.2	4	22.2	8.7	4
12.8	5.7	4	17.6	4.3	4	20.55	5.75	4	19.5	8.65	4	22.9	8.45	4

22.35	9.2	4	35.95	6.05	3	31.95	11.35	3	30.7	20.05	6	34.85	18.85	6
22.75	9.35	4	34.8	5.85	3	31.65	11.05	3	31	19.9	6	35.3	18.55	6
22.4	10.05	4	33.7	6.15	3	32.95	11.15	3	31.55	19.65	6	35.4	19.35	6
23.05	10.9	4	33.95	6.6	3	32.65	11.7	3	31.5	18.55	6	34.55	19.75	6
23.3	9.85	4	33.7	7.05	3	32.25	12.25	3	32.05	18.6	6	35.05	20	6
23.95	9.8	4	32.75	7.1	3	32.05	12.25	3	31.95	19.1	6	35.95	19.85	6
23.65	9.1	4	32.3	7.65	3	31.3	12.7	3	32.6	18.35	6	36.35	20.6	6
23.7	8.85	4	33	7.9	3	31.95	12.95	3	32.85	17.95	6	35.5	20.55	6
24.25	8.25	4	31.95	8.15	3	32.75	13.1	3	33.45	17.45	6	34.45	20.65	6
24.85	7.95	4	31.15	8.65	3	33.15	13.2	3	33.7	17	6	34.4	21.25	6
23.5	7.85	4	30.35	8.85	3	33.1	12.75	3	34.25	17.35	6	35	21.05	6
23.85	7.35	4	29.85	9	3	33.15	12.1	3	34.3	18.05	6	35.75	21.3	6
23.95	6.9	4	30.7	9.15	3	34.3	11.75	3	33.85	18.4	6	35.05	21.5	6
23.65	6.5	4	29.7	9.9	3	34	10.85	3	33.05	18.85	6	34.6	22.05	6
23.6	5.7	4	30.45	9.95	3	34.65	11	3	33.25	18.95	6	34.2	21.75	6
24.3	5.65	4	30.95	9.85	3	34.8	10.1	3	32.8	19.15	6	36.25	21.95	6
24.8	6.4	4	31.8	9.45	3	35.65	9.85	3	32.3	19.85	6	35.7	22.3	6
34.05	3.5	3	32.45	8.8	3	36.35	10	3	32.8	20	6	35.5	22.9	6
33.05	3.85	3	33.55	8.6	3	35.55	10.75	3	31.75	20.25	6	35.85	23.25	6
32	3.8	3	34.35	7.7	3	35.8	11.55	3	31.75	20.75	6	36.3	23.8	6
31.9	4.4	3	34.7	8	3	35.2	11.75	3	32.1	21.15	6	35.45	24.1	6
31.05	4.75	3	34.6	7.25	3	34.7	11.75	3	31.55	21.6	6	34.9	23.5	6
30.4	5.65	3	35	6.8	3	34.95	12.75	3	30.65	21.3	6	34.2	22.9	6
30.75	6.1	3	35.5	7.35	3	34.05	12.55	3	29.95	21.6	6	33.85	23.3	6
30	6.7	3	36.1	7.5	3	34.05	13.05	3	29.5	21.6	6	33.25	23.35	6
30.1	7.4	3	36.55	7	3	33.25	13.7	3	30.35	22.05	6	32.45	23.7	6
29.5	8.15	3	36	8.2	3	33.2	14.15	3	31.05	22	6	33.6	23.9	6
30.75	8	3	35.35	8.05	3	33.25	14.7	6	31.55	22.2	6	34.25	23.95	6
30.85	7.35	3	36.55	8.65	3	33	15.15	6	30.95	22.65	6	34.25	24.1	6
31.5	6.75	3	36.4	9.1	3	32.95	15.65	6	30.3	23.1	6	35.4	24.7	6
31.75	5.95	3	35.5	9.1	3	32.6	16.15	6	29.6	23.15	6	35.15	25.3	6
32.35	6.45	3	34.55	8.85	3	32.45	16.75	6	29.35	22.55	6	34.4	24.9	6
32.8	6	3	35.25	9.4	3	32.65	17.05	6	29.2	23.85	6	33.7	24.85	6
32.05	5.1	3	34.4	9.5	3	32.75	17.3	6	30.75	24	6	32.25	24.45	6
32.8	4.8	3	33.5	9.3	3	31.75	17.2	6	30.95	24.15	6	32.5	24.7	6
32.65	4.4	3	33.85	9.8	3	31.7	17.65	6	31.45	23.7	6	31.45	24.45	6
33.65	4.6	3	32.5	9.65	3	31	17.5	6	31.95	23.15	6	31.55	25.2	6
33.05	5.15	3	32.3	10.25	3	31.15	17.9	6	32.55	22.05	6	31.05	25	6
33.6	5.45	3	33.3	10.3	3	30.45	18.05	6	32.6	22.55	6	30.25	24.3	6
34.5	5.05	3	31.6	10.5	3	30.05	18.8	6	33.25	22.25	6	29.8	24.8	6
34.9	4.65	3	30.6	10.5	3	30.55	18.8	6	33.65	21.9	6	29.6	25.5	6
35.45	4.1	3	30.4	11.1	3	30.5	19.3	6	33.5	21.3	6	29.7	26.05	6
34.6	4.05	3	30.9	11.45	3	30.25	19.4	6	33.1	20.75	6	30.5	25.5	6
34.2	4.2	3	30.7	11.65	3	29.6	19.85	6	33.8	20.4	6	30.65	26	6
36.3	5.2	3	30.4	12.05	3	29.15	20.55	6	33.85	20	6	31.25	26.05	6
35.55	5.35	3	31.2	12	3	30.25	20.45	6	34.15	19.3	6	31.45	26.95	6

30.75	26.9	6	20.15	20.9	1	19.7	24.6	1	24.7	22.2	1	5.8	4	5
30.65	27.15	6	19.2	21.35	1	20.15	25.05	1	24.3	22.6	1	4.95	4.05	5
31.25	27.85	6	19.1	21.85	1	22.15	25.1	1	24.15	23.3	1	5.1	4.35	5
31.85	27.75	6	18.45	22.8	1	21.6	24.65	1	23.9	23.45	1	5.7	4.45	5
32.7	28.2	6	18.75	22.95	1	21.7	23.8	1	5.2	2.15	5	5.45	4.85	5
33.25	27.55	6	19.4	23	1	21.9	23.65	1	6.35	1.95	5	6.7	4.8	5
32.4	27.1	6	19.55	22.25	1	22.55	23.5	1	6.75	2.3	5	6.55	5.05	5
32.15	26.65	6	19.8	21.85	1	22.55	24.3	1	5.9	2.4	5	7.2	4.9	5
32.35	25.95	6	20.5	21.85	1	23.3	24.45	1	5.4	2.7	5	6.2	4.25	5
32.95	25.5	6	21.45	21.45	1	24.25	24.35	1	4.85	2.9	5	7.1	4.3	5
33.85	26.05	6	21.7	21.9	1	23.8	25.25	1	4.85	3.35	5	7.85	4.5	5
33.05	26.5	6	21.4	22.3	1	23.4	23.8	1	5.15	3.45	5	7.6	4.15	5
33.65	27	6	21	22.6	1	22.9	23.2	1	5.7	3.45	5	7.25	3.55	5
34.1	27.35	6	21.15	22.95	1	22.3	22.8	1	6.2	3	5	7.8	3.35	5
34.2	27.95	6	20.5	22.85	1	22.2	22.4	1	6.2	3.2	5	8.05	2.75	5
34.65	26.85	6	19.75	23.65	1	23.1	21.7	1	7.65	2.15	5	8.5	3.25	5
35.25	26	6	19.2	23.7	1	22.85	21.9	1	7.2	2.75	5	8.1	3.55	5
35.7	26.15	6	18.45	24.35	1	22.65	21.1	1	6.75	3.2	5	8.15	4	5
34.4	25.6	6	20.65	23.85	1	23.15	22.6	1	6.75	3.55	5			
21.3	20.8	1	20.65	24.3	1	24.1	21.9	1	6.65	3.8	5			