

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

**ΕΞΕΛΙΚΤΙΚΗ
ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΕ
ΠΡΟΒΛΗΜΑΤΑ
ΠΟΛΥΚΡΙΤΗΡΙΑΣ
ΟΜΑΔΟΠΟΙΗΣΗΣ**

ΝΙΚΟΛΑΚΑΚΗ ΚΩΝΣΤΑΝΤΙΝΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΔΟΥΜΠΟΣ ΜΙΧΑΛΗΣ

Χανιά, 2009

ΕΥΧΑΡΙΣΤΙΕΣ

Καταρχήν θα ήθελα να ευχαριστήσω τον καθηγητή μου κο Μιχάλη Δούμπο για τις πραγματικά πολύτιμες συμβουλές και παρεμβάσεις του κατά τη διάρκεια εκπόνησης της παρούσας εργασίας. Ακόμα, ευχαριστώ τα μέλη της εξεταστικής επιτροπής, κο Ματσατσίνη και κο Γρηγορούδη, καθώς και το σύνολο των καθηγητών του τμήματος Μηχανικών Παραγωγής και Διοίκησης, γιατί με την καθοδήγηση τους μου δόθηκε η δυνατότητα να έρθω σε επαφή και να με συνεπάρει ένας νέος, για μένα, κλάδος της επιστήμης.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την υπομονή που έχουν όλοι δείξει όλο αυτό το διάστημα, αλλά και γιατί δεν παύουν να με στηρίζουν και να με βοηθάνε σε ό,τι στόχο βάζω.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΕΧΟΜΕΝΑ.....	3
ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ.....	1
ΚΕΦΑΛΑΙΟ 2 ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ	3
2.1 ΕΙΣΑΓΩΓΗ.....	3
2.2 ΤΡΟΠΟΙ ΜΕΤΡΗΣΗΣ ΑΠΟΣΤΑΣΗΣ	4
2.2.1 Μήτρες Γειτνίασης (Proximity Matrices).....	5
2.2.2 Διαφορές Βάσει Χαρακτηριστικών (Dissimilarities Based on Attributes).....	5
2.2.3 Μέτρα ομοιότητας (Similarity measures).....	6
2.3 ΑΞΙΟΛΟΓΗΣΗ ΟΜΑΔΟΠΟΙΗΣΗΣ	8
2.3.1 Μη Επιτηρούμενοι Δείκτες.....	8
2.3.2 Επιτηρούμενοι Δείκτες.....	10
2.4 ΙΕΡΑΡΧΙΚΕΣ ΜΕΘΟΔΟΙ (HIERARCHICAL CLUSTERING).....	11
2.5 ΔΙΑΧΩΡΙΣΤΙΚΗ ΟΜΑΔΟΠΟΙΗΣΗ (PARTITIONAL CLUSTERING).....	13
2.6 ΕΞΕΛΙΚΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ	14
ΚΕΦΑΛΑΙΟ 3 ΠΟΛΥΚΡΙΤΗΡΙΑ ΟΜΑΔΟΠΟΙΗΣΗ.....	18
3.1 ΕΙΣΑΓΩΓΗ.....	18
3.2 ΔΙΑΤΥΠΩΣΗ ΠΡΟΒΛΗΜΑΤΟΣ.....	19
3.3 ΜΙΑ ΠΡΟΕΚΤΑΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ K-MEANS	20
3.4 PROMETHEE CLUSTER	22
3.5 ΔΙΑΤΕΤΑΓΜΕΝΗ ΠΟΛΥΚΡΙΤΗΡΙΑ ΟΜΑΔΟΠΟΙΗΣΗ.....	25
ΚΕΦΑΛΑΙΟ 4 ΕΞΕΛΙΚΤΙΚΟΣ ΑΛΓΟΡΙΘΜΟΣ ΠΟΛΥΚΡΙΤΗΡΙΑΣ ΟΜΑΔΟΠΟΙΗΣΗΣ.....	29
4.1 ΕΙΣΑΓΩΓΗ.....	29
4.2 ΔΙΑΦΟΡΙΚΗ ΕΞΕΛΙΞΗ	29
4.2.1 Δομή Πληθυσμού	29
4.2.2 Αναπαράσταση παραμέτρων	30
4.2.3 Αρχικοποίηση.....	30
4.2.4 Μετάλλαξη.....	31
4.2.5 Διασταύρωση.....	35
4.2.6 Επιλογή.....	36
4.2.7 Κριτήρια τερματισμού	36
4.2.8 Υλοποίηση σε περιβάλλον MATLAB	37
4.3 ΠΟΛΥΚΡΙΤΗΡΙΑ ΟΜΑΔΟΠΟΙΗΣΗ	38
4.4 ΑΝΤΙΚΕΙΜΕΝΙΚΗ ΣΥΝΑΡΤΗΣΗ	39
ΚΕΦΑΛΑΙΟ 5 ΠΕΙΡΑΜΑΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ.....	42
5.1 ΔΕΔΟΜΕΝΑ.....	42

5.2	ΜΕΘΟΔΟΛΟΓΙΑ.....	45
5.3	ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ.....	46
5.3.1	<i>Αποτελέσματα Διαχωρισμού σε 2 Ομάδες</i>	46
5.3.2	<i>Αποτελέσματα Διαχωρισμού σε 3 Ομάδες</i>	52
5.3.3	<i>Αποτελέσματα Διαχωρισμού σε 4 Ομάδες</i>	58
5.4	ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ.....	63
ΚΕΦΑΛΑΙΟ 6.....		67
ΣΥΜΠΕΡΑΣΜΑΤΑ		67
ΒΙΒΛΙΟΓΡΑΦΙΑ		69

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Η ανάλυση ομάδων (cluster analysis) ή ομαδοποίηση (clustering) είναι μια από τις πιο ευρέως χρησιμοποιούμενες τεχνικές για διερευνητική ανάλυση δεδομένων. Κύριος στόχος της είναι ο διαχωρισμός ενός συνόλου αντικειμένων σε ομογενή υποσύνολα ή «ομάδες» (clusters). Μπορεί να θεωρηθεί δηλαδή σαν ένα πρόβλημα βελτιστοποίησης όπου έχει στόχο τη μεγιστοποίηση των ομοιοτήτων μεταξύ των αντικειμένων μέσα στο ίδιο σύνολο και της ελαχιστοποίησής τους ανάμεσα σε διαφορετικά σύνολα. Οι ομάδες που θα σχηματιστούν δεν είναι γνωστές πριν την πραγματοποίηση της ανάλυσης.

Στην «παραδοσιακή» ομαδοποίηση ο διαχωρισμός των αντικειμένων σε ομάδες γίνεται ως προς ένα χαρακτηριστικό των αντικειμένων. Στην πραγματικότητα όμως, έχουμε να κάνουμε με πολλά και αντικρουόμενα κριτήρια όταν αντιμετωπίζουμε σύνθετα προβλήματα. Παρόλο όμως που οι τεχνικές της «παραδοσιακής» ομαδοποίησης εφαρμόζονται με επιτυχία εδώ και πολλές δεκαετίες, λίγη προσοχή έχει δοθεί στην ομαδοποίηση από την οπτική της πολυκριτήριας ανάλυσης. Γι' αυτό και η παρούσα εργασία θα ασχοληθεί με την πολυκριτήρια ομαδοποίηση.

Εκτός από τους κλασσικούς αλγόριθμους ομαδοποίησης, έχουν χρησιμοποιηθεί και εξελικτικοί αλγόριθμοι. Οι εξελικτικοί αλγόριθμοι, όπως και οι γενετικοί (Holland, 1975), είναι ευρετικοί αλγόριθμοι στοχαστικής αναζήτησης εμπνευσμένοι από τη θεωρία εξέλιξης του Δαρβίνου και τη γενετική. Η κύρια ιδέα είναι η δημιουργία ενός πληθυσμού υποψήφιων λύσεων σε ένα πρόβλημα βελτιστοποίησης, ο οποίος βελτιώνεται επαναληπτικά μέσω τροποποίησης και επιλογής των καλών λύσεων για να χρησιμοποιηθούν στην επόμενη επανάληψη. Οι υποψήφιες λύσεις επιλέγονται σύμφωνα με την αντικειμενική συνάρτηση, η οποία εκτιμά την ποιότητα των λύσεων με γνώμονα το δεδομένο πρόβλημα βελτιστοποίησης.

Σημαντικό πλεονέκτημα αυτών των αλγορίθμων είναι η ικανότητά τους να αντεπεξέρχονται στην τοπική αναζήτηση διατηρώντας, συνδυάζοντας και συγκρίνοντας διαφορετικές λύσεις ταυτόχρονα.

Στόχος της συγκεκριμένης εργασίας είναι η ομαδοποίηση πολυκριτήριων δεδομένων με συνδυασμό του αλγορίθμου της Διαφορικής Εξέλιξης με μεθόδους πολυκριτήριας ομαδοποίησης, όπως είναι η PROMETHEE Cluster. Χρησιμοποιούνται διαφορετικοί τύποι υπολογισμού της αντικειμενικής συνάρτησης, τρόποι διαχωρισμού των εναλλακτικών, βάρη, παράμετροι και κριτήρια αξιολόγησης της ομαδοποίησης, προκειμένου να διαπιστωθεί σε ποιες περιπτώσεις επιτυγχάνεται καλύτερη ομαδοποίηση.

Στο κεφάλαιο δύο, γίνεται μια σύντομη περιγραφή του κλασσικού προβλήματος της ομαδοποίησης. Στη συνέχεια αναλύονται μερικοί χαρακτηριστικοί τρόποι μέτρησης της απόστασης και μέθοδοι αξιολόγησης της ομαδοποίησης. Ακόμα γίνεται αναφορά στην περίπτωση της ιεραρχικής ομαδοποίησης και στην επίλυση των συγκεκριμένων προβλημάτων με μεθόδους εξελικτικής βελτιστοποίησης.

Στο τρίτο κεφάλαιο αναλύονται οι τεχνικές ανάλυσης ομάδων με πολυκριτήριο χαρακτήρα. Αφού γίνει πρώτα η διατύπωση του συγκεκριμένου προβλήματος, παρουσιάζονται επιλεκτικά μερικές ενδιαφέρουσες τεχνικές επίλυσής του, οι οποίες χρησιμοποιούν τις πληροφορίες προτίμησης του αποφασίζοντας. Συγκεκριμένα παρουσιάζεται μια παραλλαγή του αλγόριθμου k-means, η μέθοδος PROMETHEE CLUSTER και μεθόδους επίλυσης της διατεταγμένης πολυκριτήριας ομαδοποίησης.

Στο κεφάλαιο τέσσερα περιγράφεται ο τρόπος λειτουργίας του αλγόριθμου της Διαφορικής Εξέλιξης, με έμφαση στα χαρακτηριστικά του εκείνα που τον κάνουν να διαφέρει και σε πολλές περιπτώσεις να είναι πιο αποτελεσματικός από άλλους εξελικτικούς αλγορίθμους. Στη συνέχεια περιγράφονται οι τρόποι ομαδοποίησης που χρησιμοποιήθηκαν στην παρούσα εργασία, καθώς και δώδεκα διαφορετικές εκδοχές της αντικειμενικής συνάρτησης βάσει της οποίας γίνεται η επιλογή στον αλγόριθμο Δ.Ε.

Τέλος, στο πέμπτο κεφάλαιο πραγματοποιείται ανάλυση και παρουσίαση των αποτελεσμάτων που προέκυψαν από την εφαρμογή των παραλλαγών του αλγορίθμου σε πραγματικά δεδομένα του τραπεζικού κλάδου.

ΚΕΦΑΛΑΙΟ 2

ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ

2.1 ΕΙΣΑΓΩΓΗ

Η ομαδοποίηση (clustering) είναι μια από τις πιο διαδεδομένες τεχνικές ανάλυσης δεδομένων. Διαχωρίζει τα δεδομένα σε κατηγορίες με σκοπό την καλύτερη κατανόησή τους ή προκειμένου να χρησιμοποιηθούν τα αποτελέσματα στη συνέχεια σε άλλες αναλύσεις. Είτε για κατανόηση είτε για παραπέρα χρήση, η ανάλυση ομάδων έχει παίξει σημαντικό ρόλο σε πολλά επιστημονικά πεδία. Από τις κοινωνικές επιστήμες ως τη βιολογία, και από την επιστήμη υπολογιστών ως τις οικονομικές επιστήμες, οι ερευνητές αναζητούν μια πρώτη αίσθηση των δεδομένων τους διαχωρίζοντάς τα σε ομάδες.

Για παράδειγμα, στη βιολογία η ομαδοποίηση μπορεί να χρησιμοποιηθεί στη δημιουργία ομάδων γονιδίων με παρόμοια δομή. Συχνά τέτοιες ομάδες περιέχουν πρωτεΐνες που συσχετίζονται λειτουργικά. Στην ανάλυση αλληλουχιών, η ομαδοποίηση μπορεί να χρησιμοποιηθεί για το διαχωρισμό ομόλογων αλληλουχιών σε οικογένειες γονιδίων. Ακόμα, η ανάλυση ομάδων είναι ευρέως γνωστή στην έρευνα αγοράς, όπου διαχωρίζεται ο γενικός πληθυσμός των καταναλωτών σε τμήματα με βάση τις απαντήσεις τους σε έρευνες αγοράς. Αυτό βοηθάει στην καλύτερη κατανόηση των σχέσεων μεταξύ των διαφορετικών ομάδων των εν δυνάμει καταναλωτών. Σε κοινωνιολογικές μελέτες, η ομαδοποίηση μπορεί να χρησιμοποιηθεί για την αναγνώριση κοινοτήτων μέσα σε ένα μεγαλύτερο σύνολο πληθυσμού. Είναι εμφανές λοιπόν ότι η ομαδοποίηση έχει μια πληθώρα εφαρμογών σε πρακτικά προβλήματα

Ανάλυση ομάδων (cluster analysis) ή ομαδοποίηση, είναι ο διαχωρισμός αντικειμένων ενός πληθυσμού με σκοπό να βρεθεί η δομή των δεδομένων (Webb, 2002). Βασικός στόχος είναι η δημιουργία τέτοιων συνόλων ώστε τα στοιχεία κάθε συνόλου να βρίσκονται πιο κοντά μεταξύ τους απ' ό,τι με τα στοιχεία των άλλων συνόλων. Εκτός αυτού η ανάλυση ομάδων, σύμφωνα με τους Hastie et al. (2001), μπορεί να έχει και άλλους στόχους.

Ένα στόχος, για παράδειγμα είναι η διεύθυνση των ομάδων σε μια φυσική ιεραρχία. Αυτό περιλαμβάνει τη διαδοχική ταξινόμηση των ίδιων των ομάδων έτσι ώστε σε κάθε επίπεδο ιεραρχίας οι ομάδες μέσα στην ίδια κατηγορία να είναι πιο όμοιες μεταξύ τους απ' ό,τι αυτές σε διαφορετικές κατηγορίες.

Η ανάλυση ομάδων χρησιμοποιείται επίσης για περιγραφική στατιστική (descriptive statistics) για να εξακριβωθεί κατά πόσο τα δεδομένα αποτελούνται από ένα σύνολο ευδιάκριτων υποκατηγοριών, όπου κάθε κατηγορία αντιπροσωπεύει αντικείμενα με σημαντικά διαφορετικές ιδιότητες. Αυτός ο στόχος απαιτεί μια εκτίμηση του βαθμού διαφορετικότητας των αντικειμένων που έχουν καταχωρηθεί σε κάθε κατηγορία. Τέλος, κεντρικός όλων των στόχων είναι η αναζήτηση του βαθμού ομοιότητας (ή διαφορετικότητας) μεταξύ των αντικειμένων που ομαδοποιούνται.

Είναι δύσκολο να δοθεί ένα ενιαίος ορισμός για την ανάλυση ομάδων. Συνήθως εννοείται ο διαχωρισμός του συνόλου των δεδομένων σε αμοιβαία μη αλληλεπικαλυπτόμενα υποσύνολα. Ωστόσο, ανάλογα με τη μέθοδο μπορεί να προκύψουν διαφορετικές ομαδοποιήσεις, αφού κάθε μέθοδος προϋποθέτει διαφορετική δομή στα δεδομένα. Βασική σε όλες τις τεχνικές ομαδοποίησης είναι η επιλογή του τρόπου μέτρησης της απόστασης ή της ανομοιότητας μεταξύ δυο αντικειμένων.

2.2 ΤΡΟΠΟΙ ΜΕΤΡΗΣΗΣ ΑΠΟΣΤΑΣΗΣ

Κάθε αλγόριθμος ομαδοποίησης χρησιμοποιεί την κατάλληλη αντικειμενική συνάρτηση που βασίζεται σε κάποιο μέγεθος ομοιότητας ή ανομοιότητας. Υπάρχουν διαφορετικοί τρόποι μέτρησης αφού οι μεταβλητές μπορεί να είναι αριθμητικές ή μη και αφού μπορεί να μετρηθεί η ομοιότητα ή η ανομοιότητα. Υπάρχουν περιπτώσεις που αλγόριθμοι που χρησιμοποιούν διαφορετικά μεγέθη μέτρησης δίνουν ίδια αποτελέσματα (τέτοια μεγέθη μέτρησης ονομάζονται ισοδύναμα).

Ένα μέγεθος μέτρησης της ανομοιότητας έχει τις ακόλουθες ιδιότητες (Nemery, 2006): $\forall x_i, x_j \in A : d(x_i, x_j) \mapsto R$ με:

1. $d(x_i, x_i) = 0$, κανονικοποίηση
2. $d(x_i, x_j) \geq 0$, μη αρνητικότητα
3. $d(x_i, x_j) = d(x_j, x_i)$, συμμετρία

Όταν ισχύουν επιπλέον και οι ακόλουθες ανισότητες, το μέτρο ανομοιότητας ονομάζεται απόσταση.

4. $d(x_i, x_j) = 0 \Rightarrow x_i = x_j$, ταυτότητα
5. $\forall x_z : d(x_i, x_j) \leq d(x_i, x_z) + d(x_z, x_j)$, τριγωνική ανισότητα

Σύμφωνα με τους Hastie et al. (2001), υπάρχουν οι παρακάτω μέθοδοι μέτρησης απόστασης ή ανομοιότητας δυο αντικειμένων:

2.2.1 Μήτρες Γειτνίασης (Proximity Matrices)

Κάποιες φορές τα δεδομένα αναπαριστώνται άμεσα με όρους γειτνίασης (ομοιότητας ή συνάφειας) μεταξύ ζευγών αντικειμένων. Αυτοί μπορεί να είναι είτε ομοιότητες είτε διαφορές.

Για παράδειγμα, στα πειράματα των κοινωνικών επιστημών, ζητούν στους συμμετέχοντες να κρίνουν κατά πόσο συγκεκριμένα αντικείμενα διαφέρουν μεταξύ τους, οπότε οι διαφορές των αντικειμένων μπορούν να υπολογισθούν από τον μέσο όρο αυτών των κρίσεων. Αυτού του τύπου τα δεδομένα μπορούν να αναπαρασταθούν από έναν $N \times N$ πίνακα D , όπου N είναι ο αριθμός των αντικειμένων, και κάθε στοιχείο d_{ij} , αναπαριστά την ομοιότητα μεταξύ των αντικειμένων i και j . Αυτός ο πίνακας είναι το εισερχόμενο σε έναν αλγόριθμο ομαδοποίησης.

Οι περισσότεροι αλγόριθμοι χρειάζονται έναν πίνακα διαφορών με μη αρνητικές εγγραφές και μηδενικά στοιχεία διαγωνίου: $d_{ij}=0, i=1,2,...N$. Αν τα αρχικά δεδομένα είχαν συλλεχθεί ως ομοιότητες, μπορεί να χρησιμοποιηθεί μια μονότονη φθίνουσα συνάρτηση για τη μετατροπή τους σε διαφορές. Επίσης, οι περισσότεροι αλγόριθμοι θεωρούν ως δεδομένο ότι οι πίνακες διαφορών είναι συμμετρικοί, έτσι αν ο αρχικός πίνακας D δεν είναι συμμετρικός πρέπει να αντικατασταθεί από $(D+D^T)/2$.

Οι διαφορές που έχουν εξαχθεί από υποκειμενικές κρίσεις, σπάνια μπορούν να θεωρηθούν σαν αποστάσεις με την αυστηρή έννοια του όρου, καθώς δεν ισχύει η τριγωνική ανισότητα: $d_{ij} \leq d_{ik} + d_{jk}$, για κάθε $k \in \{1,2,...N\}$. Γι' αυτό κάποιοι αλγόριθμοι που προϋποθέτουν αποστάσεις δεν μπορούν να χρησιμοποιηθούν.

2.2.2 Διαφορές Βάσει Χαρακτηριστικών (Dissimilarities Based on Attributes)

Πιο συχνά τα δεδομένα είναι μετρήσεις x_{ij} για $i=1,2,...n$ αντικείμενα, σε $j=1,2,...p$ χαρακτηριστικά. Αφού οι περισσότεροι από τους πιο δημοφιλείς αλγόριθμους ομαδοποίησης βασίζονται σε κάποιον πίνακα διαφορών, πρέπει πρώτα να βρεθούν οι διαφορές ανά ζεύγος αντικειμένων. Στην πιο συνηθισμένη περίπτωση ορίζεται μια διαφορά $d_j(x_{ij}, x_{il})$ μεταξύ τιμών του χαρακτηριστικού j , και ορίζεται:

$$D(x_i, x_l) = \sum_{k=1}^p d_k(x_{ik}, x_{lk}), \text{ η διαφορά μεταξύ των αντικειμένων } i \text{ και } l.$$

Στις ποσοτικές μεταβλητές, οι οποίες αναπαρίστανται από συνεχείς πραγματικούς αριθμούς, μπορεί να χρησιμοποιηθεί η απόσταση L_p :

$$d(x_i, x_l) = \|x_i - x_l\|_p$$
$$\|x_i - x_l\|_p = \left(\sum_{k=1}^p |x_{ik} - x_{lk}|^p \right)^{1/p} \text{ όπου } 1 \leq p \leq \infty$$

όσο μεγαλύτερη η τιμή του p , τόσο μεγαλύτερη η σημασία των διαφορών. Η πιο συνηθισμένη επιλογή είναι για $p=2$, η γνωστή Ευκλείδεια ή τετραγωνική απόσταση:

$$d_j(x_{ij}, x_{lj}) = (x_{ij} - x_{lj})^2$$

Η απόσταση Manhattan αντιστοιχεί σε $p=1$, ενώ η περίπτωση $p=\infty$ ισοδυναμεί στη μέγιστη απόλυτη απόσταση.

Σε ορισμένες εφαρμογές, τα σημεία μπαίνουν σε κλίμακα ώστε το μέγεθος η απόσταση να μετράται από την απόσταση μεταξύ των σημείων:

$$d(x_i, x_l) = \arccos(x_i^T x_l / \|x_i\| \cdot \|x_l\|)$$

Ενώ, όταν τα χαρακτηριστικά δεν είναι αναγκαστικά ανεξάρτητα είναι δυνατόν να χρησιμοποιηθεί η απόσταση Mahalanobis (Nemery, 2006):

$$d(x_i, x_l) = ((x_i - x_l)^T S^{-1} (x_i - x_l))^{1/2}$$

Για μη ποσοτικά δεδομένα η τετραγωνική απόσταση και το απόλυτο σφάλμα δεν είναι κατάλληλα.

Οι τακτικές μεταβλητές συνήθως αναπαρίστανται με συνεχομένους ακέραιους και οι εφικτές τιμές θεωρούνται ως ένα διατεταγμένο σύνολο. Παραδείγματα είναι οι βαθμοί, ο βαθμός προτίμησης κ.α. Η μέτρηση του σφάλματος σε αυτού του είδους τις μεταβλητές γενικά ορίζεται αντικαθιστώντας τις M αρχικές τιμές τους με:

$$\frac{i - 1/2}{M}, i = 1, \dots, M$$

Μετά αντιμετωπίζονται όπως προηγουμένως οι ποσοτικές μεταβλητές.

Στις ονομαστικές μεταβλητές, που προσδιορίζουν μη διατεταγμένες κατηγορίες, ο βαθμός απόκλισης μεταξύ δύο τιμών πρέπει να περιγράφει αναλυτικά. Αν η μεταβλητή μπορεί να περιγραφεί από M διακριτές τιμές, αυτές μπορούν να καταγραφούν σε έναν $M \times M$ πίνακα με στοιχεία $L_{rd} = L_{dr}, L_{rr} = 0, L_{rd} \geq 0$. Η πιο συνηθισμένη επιλογή είναι $L_{rd} = 1$ για όλα τα $r \neq d$.

Επιπλέον, είναι πολλές φορές επιθυμητό να σταθμίζονται οι ιδιότητες διαφορετικά και όχι να τους αποδίδεται το ίδιο βάρος. Αυτό σχεδόν πάντα γίνεται με τη χρήση σταθμισμένου μέσου όρου. Έτσι, ο υπολογισμός της συνολικής ανομοιότητας $D(x_i, x_l)$ μεταξύ δύο αντικειμένων (x_i, x_l) , μπορεί να γίνει ως εξής:

$$D(x_i, x_l) = \sum_{j=1}^p w_j \cdot d_j(x_{ij}, x_{lj}) \quad \sum_{j=1}^p w_j = 1$$

Όπου w_j είναι ένα βάρος που αποδίδεται στο χαρακτηριστικό j και το οποίο ρυθμίζει τη σχετική επιρροή αυτής της μεταβλητής στον καθορισμό της συνολικής ανομοιότητας μεταξύ αντικειμένων.

2.2.3 Μέτρα ομοιότητας (Similarity measures)

Αντίθετα με τα μεγέθη που βασίζονται στις αποστάσεις, οι δείκτες ομοιότητας μπορούν να χρησιμοποιηθούν σε όλα τα είδη μεταβλητών, ακόμα και ποιοτικές. Οι

δείκτες ομοιότητας ορίζονται χρησιμοποιώντας ως αναφορά τους δείκτες παρατήρησης και παίρνουν τιμές στο κλειστό διάστημα $[0,1]$. Το συμπληρωματικό ενός δείκτη ομοιότητας ονομάζεται δείκτης ανομοιότητας και αντιπροσωπεύει μια κατηγορία δεικτών γειτνίασης πιο ευρύ από αυτούς που σχετίζονται με τις αποστάσεις, αφού οι ιδιότητες που πρέπει να ικανοποιούν, όπως φαίνεται στην αρχή της ενότητας, είναι λιγότερες (δεν χρειάζεται να ικανοποιούν την ιδιότητα της ταυτότητας και της τριγωνικής ανισότητας).

Οι δείκτες ομοιότητας μπορούν να υπολογιστούν, κατ' αρχήν, για ποσοτικές μεταβλητές. Αλλά θα είχαν περιορισμένη χρήση, αφού θα διέκριναν μόνο αν δύο παρατηρήσεις έχουν ίδιες ή διαφορετικές τιμές, χωρίς όμως να λένε τίποτα για το μέγεθος της διαφοράς. Οι κύριοι δείκτες ομοιότητας έχουν ως αναφορά μήτρες δεδομένων που περιέχουν δυαδικές μεταβλητές. Σε πιο γενικές περιπτώσεις, με μεταβλητές που έχουν πάνω από δύο επίπεδα, μπορούν να μετασχηματιστούν σε αυτό το πλαίσιο (Giudici, 2003).

Για την παρουσίαση μερικών χαρακτηριστικών δεικτών ομοιότητας μπορεί να βοηθήσει η χρήση συγκεκριμένου παραδείγματος. Για n επισκέπτες δικτυακών τόπων, καθένας από τους οποίους έχει P σελίδες, αναζητείται ο βαθμός ομοιότητας μεταξύ των επισκεπτών όσον αφορά τις σελίδες που επισκέπτεται ο καθένας. Συνεπώς, υπάρχουν P δυαδικές μεταβλητές, που λαμβάνουν τιμή 1 όταν η συγκεκριμένη σελίδα έχει δεχθεί επίσκεψη, και 0 σε διαφορετική περίπτωση.

Θεωρώντας, για διευκόλυνση, ότι υπάρχουν 2 επισκέπτες, με CP συμβολίζεται ο αριθμός των σελίδων που έχουν επισκεφτεί και οι δύο επισκέπτες, με CA ο αριθμός των σελίδων που δεν έχει πάει κανένας επισκέπτης και με AP και PA ο αριθμός των σελίδων που έχει πάει μόνο ο ένα ή μόνο ο άλλος επισκέπτης αντίστοιχα.

Μερικοί λοιπόν χαρακτηριστικοί δείκτες ομοιότητας είναι οι ακόλουθοι:

- Δείκτης ομοιότητας των Rassel και Rao:

$$S_{ij} = \frac{CP}{P}$$

Αυτός ο δείκτης ισούται με το κλάσμα του αριθμού των σελίδων που υπάρχει ταυτόχρονη παρουσία με τον συνολικό αριθμό των δυαδικών μεταβλητών.

- Δείκτης ομοιότητας του Jaccard:

$$S_{ij} = \frac{CP}{CP + PA + AP}$$

Αυτός ο δείκτης είναι το κλάσμα του αριθμού των σελίδων που υπάρχει ταυτόχρονη παρουσία με τον συνολικό αριθμό των δυαδικών μεταβλητών εξαιρουμένων των αριθμών των σελίδων που δεν έχουν καμία παρουσία. Σημειώνεται ότι αυτός ο δείκτης δεν ορίζεται όταν υπάρχουν μόνο ταυτόχρονες απουσίες ($CA=P$)

- Δείκτης ομοιότητας των Sokkar και Michener:

$$S_{ij} = \frac{CP + CA}{P}$$

Αυτός αναπαριστά το κλάσμα μεταξύ του αριθμού των σελίδων που υπάρχει ταυτόχρονη παρουσία και ταυτόχρονη απουσία με τον συνολικό αριθμό των μεταβλητών.

2.3 ΑΞΙΟΛΟΓΗΣΗ ΟΜΑΔΟΠΟΙΗΣΗΣ

Οποιαδήποτε μέθοδος ομαδοποίησης και να εφαρμοστεί θα καταλήξει σε κάποιο διαχωρισμό σε ομάδες, χωρίς όμως αυτό να σημαίνει ότι οι ομάδες έχουν φυσική σημασία, γι' αυτό είναι απαραίτητη η διατύπωση κριτηρίων μέτρησης της καταλληλότητας των αποτελεσμάτων.

Σύμφωνα με τους Pang- Ning Tan et al. (2006), μπορούν να διακριθούν οι παρακάτω τύποι μεγεθών ή δεικτών αξιολόγησης της ομαδοποίησης:

1. Μη επιτηρούμενοι δείκτες:
2. Επιτηρούμενοι δείκτες

2.3.1 Μη Επιτηρούμενοι Δείκτες

Οι μη επιτηρούμενοι δείκτες, ή αλλιώς εσωτερικοί δείκτες, μετράνε την ποιότητα της ομαδοποίησης χωρίς να λαμβάνονται υπ' όψιν εξωτερικές πληροφορίες. Αυτού του τύπου οι δείκτες συχνά διαχωρίζονται σε δύο παραπέρα κατηγορίες, *δείκτες συνοχής των ομάδων*, όπου προσδιορίζουν πόσο κοντά είναι τα αντικείμενα που έχουν καταταγεί σε κάθε ομάδα, και *δείκτες διαχωρισμού ομάδων*, όπου μετράνε πόσο καλά είναι διαχωρισμένη η μία ομάδα από την άλλη.

Γενικά, η συνολική αξία μιας ομαδοποίησης μπορεί να εκφραστεί σαν το σταθμισμένο άθροισμα της αξίας των επιμέρους ομάδων που έχουν προκύψει:

$$\text{Συνολική Αξία} = \sum_{h=1}^K w_h \cdot \text{αξία}(C_h)$$

Η συνάρτηση αξίας μπορεί να αναπαριστά το βαθμό συνοχής, διαχωρισμού ή κάποιο συνδυασμό των δύο. Τα βάρη εξαρτώνται από το δείκτη μέτρησης της αξίας. Σε κάποιες περιπτώσεις είναι απλά 1 ή το μέγεθος της ομάδας, ενώ σε άλλες περιπτώσεις μπορεί να αντανακλούν σε πιο πολύπλοκες ιδιότητες όπως την τετραγωνική ρίζα της συνοχής.

Η συνοχή μίας ομάδας μπορεί να οριστεί σαν το άθροισμα των αποστάσεων μεταξύ των σημείων που έχουν καταταγεί σε μια ομάδα:

$$\text{Συνοχή}(C_h) = \sum_{\substack{x_i \in C_h \\ x_l \in C_h}} \text{απόσταση}(x_i, x_l)$$

Ή το άθροισμα των αποστάσεων των σημείων που έχουν καταταγεί σε μία ομάδα από το σημείο αναφοράς της ομάδας:

$$\text{Συνοχή}(C_h) = \sum_{x \in C_h} \text{απόσταση}(x, r_h)$$

Αντίστοιχα, ο διαχωρισμός μεταξύ δύο ομάδων μπορεί να μετρηθεί από την απόσταση των σημείων που έχουν καταταγεί σε κάθε ομάδα:

$$\Delta\chi\omega\rho\iota\sigma\mu\acute{o}\varsigma(C_h, C_z) = \sum_{\substack{x_i \in C_h \\ x_l \in C_z}} \alpha\pi\acute{o}\sigma\tau\alpha\sigma\eta(x_i, x_l)$$

Ή από την απόσταση των κέντρων αναφοράς των ομάδων ή από την απόσταση του σημείου αναφοράς κάθε ομάδας από το κέντρο όλων των σημείων:

$$\Delta\chi\omega\rho\iota\sigma\mu\acute{o}\varsigma(C_h, C_z) = \alpha\pi\acute{o}\sigma\tau\alpha\sigma\eta(r_h, r_z)$$

$$\Delta\chi\omega\rho\iota\sigma\mu\acute{o}\varsigma(C_h) = \alpha\pi\acute{o}\sigma\tau\alpha\sigma\eta(r_h, r)$$

Για παράδειγμα, οι Maulik και Bandyopadhyay (2000), για την μέτρηση της καταλληλότητας των ομάδων χρησιμοποιούν το μέγεθος μέτρησης της ομαδοποίησης M , το οποίο υπολογίζεται ως εξής:

$$M = \sum_{h=1}^k M_h$$

$$M_h = \sum_{x_i \in C_h} \|x_i - r_h\|$$

Δηλαδή, το μέγεθος μέτρησης της ομαδοποίησης κάθε ομάδας M_h εκφράζεται από το άθροισμα της απόστασης όλων των σημείων από το σημείο αναφοράς της ομάδας, ενώ το συνολικό M είναι το άθροισμα των επιμέρους M_h . Όσο καλύτερη η ομαδοποίηση τόσο μικρότερη η τιμή του συγκεκριμένου δείκτη.

Μια άλλη περίπτωση δείκτη που έχει χρησιμοποιηθεί είναι αυτός που πρότειναν οι Calinski και Harabasz (1974):

$$VRC = \frac{\text{trace}(\mathbf{B})/(k-1)}{\text{trace}(\mathbf{W})/(n-k)}, \text{ όπου}$$

$$\mathbf{B} = \sum_{h=1}^k n_h (r_h - \bar{x})(r_h - \bar{x})' \text{ με } \bar{x} = \left(\sum_{i=1}^n x_i \right) / n, \text{ και}$$

$$\mathbf{W} = \sum_{h=1}^k \mathbf{W}_h \text{ με } \mathbf{W}_h = \sum_{l=1}^{n_h} (x_l^{(h)} - r_h)(x_l^{(h)} - r_h)'$$

Τα $(k-1)$ και $(n-k)$ είναι αντίστοιχα, οι βαθμοί ελευθερίας της απόστασης των «κέντρων» των ομάδων από το συνολικό «κέντρο» των σημείων και της απόστασης των σημείων από το «κέντρο» της ομάδας που έχει καταχωρηθεί το καθένα. Ο συγκεκριμένος δείκτης είναι προς μεγιστοποίηση, δηλαδή όσο πιο μεγάλη η τιμή του, τόσο καλύτερη η ομαδοποίηση. Υπονοεί χαμηλή συσχέτιση μεταξύ των μετρήσεων και δίνει ίση σημασία σε όλα τα σημεία κάθε ομάδας. Η χρήση του δείκτη VRC οδηγεί στην δημιουργία σφαιρικών ομάδων, ενώ μπορεί να χρησιμοποιηθεί για την αναγνώριση του βέλτιστου αριθμού ομάδων συγκρίνοντας τα αποτελέσματα της μεγιστοποίησης για διάφορες τιμές του k .

Ένας άλλος δείκτης αξιολόγησης, που ονομάζεται και δείκτης Marriott (Marriott, 1971, 1982) είναι ο εξής:

$$MC = k^2 \frac{\det(\mathbf{W})}{\det(\mathbf{T})}, \text{ όπου } \mathbf{T} = \mathbf{B} + \mathbf{W}.$$

Ο συγκεκριμένος δείκτης είναι προς ελαχιστοποίηση και λαμβάνει υπόψη τις συσχετίσεις μεταξύ των μεταβλητών, γεγονός που δίνει τη δυνατότητα διαμόρφωσης ελλειπτικών ομάδων. Ο δείκτης Marriott χρησιμοποιείται συχνά για την ομαδοποίηση δεδομένων που συσχετίζονται ισχυρά μεταξύ τους..

2.3.2 Επιτηρούμενοι Δείκτες

Οι επιτηρούμενοι δείκτες, ή αλλιώς εξωτερικοί δείκτες, υπολογίζουν κατά πόσο η δομή της ομαδοποίησης που καταλήγει ένας αλγόριθμος ταιριάζει με κάποια εξωτερική ταξινόμηση των ίδιων στοιχείων. Υπάρχουν αρκετοί δείκτες που χρησιμοποιούνται για αυτού του είδους την αξιολόγηση, όπως η εντροπία, η καθαρότητα, η ακρίβεια, η ανάκληση και άλλες.

Η *εντροπία* (entropy) δείχνει το βαθμό που κάθε ομάδα αποτελείται από αντικείμενα μίας προκαθορισμένης κατηγορίας. Για κάθε ομάδα, υπολογίζεται πρώτα η κατανομή των αντικειμένων διαφορετικών κατηγοριών κάθε ομάδας. Για την ομάδα h υπολογίζεται η πιθανότητα p_{ih} ένα μέλος της ομάδας i να ανήκει στην κατηγορία h : $p_{ih} = m_{ih}/m_i$, όπου m_i είναι ο αριθμός των αντικειμένων στην ομάδα i και m_{ih} είναι ο αριθμός των αντικειμένων της κατηγορίας h που βρίσκονται στην ομάδα i . Έτσι η εντροπία κάθε ομάδας μπορεί να υπολογιστεί από τον τύπο:

$$e_i = -\sum_{h=1}^L p_{ih} \log_2 p_{ih},$$

όπου L είναι ο αριθμός των τάξεων. Η συνολική εντροπία μπορεί να εντοπιστεί από το άθροισμα των επιμέρους εντροπιών σταθμισμένο από το μέγεθος κάθε ομάδας:

$$e = \sum_{i=1}^K \frac{m_i}{m} e_i,$$

όπου K είναι ο αριθμός των ομάδων και m ο συνολικός αριθμός των αντικειμένων.

Ένα άλλο μέγεθος αξιολόγησης είναι η *καθαρότητα* (purity). Η καθαρότητα κάθε ομάδας είναι:

$$p_i = \max_h p_{ih},$$

ενώ η συνολική καθαρότητα:

$$\text{Καθαρότητα} = \sum_{i=1}^K \frac{m_i}{m} p_i.$$

Αντίστοιχα, η *ακρίβεια* (precision) εκφράζει το κλάσμα μιας ομάδας i που αποτελείται από αντικείμενα μιας συγκεκριμένης τάξης h :

$$\text{Ακρίβεια}(i, h) = p_{ih}.$$

Τέλος, η *ανάκληση* (recall) εξετάζει κατά πόσο μία ομάδα i περιέχει όλα τα στοιχεία μίας τάξης h :

$$\text{Ανάκληση}(i, h) = m_{ih} / m_h,$$

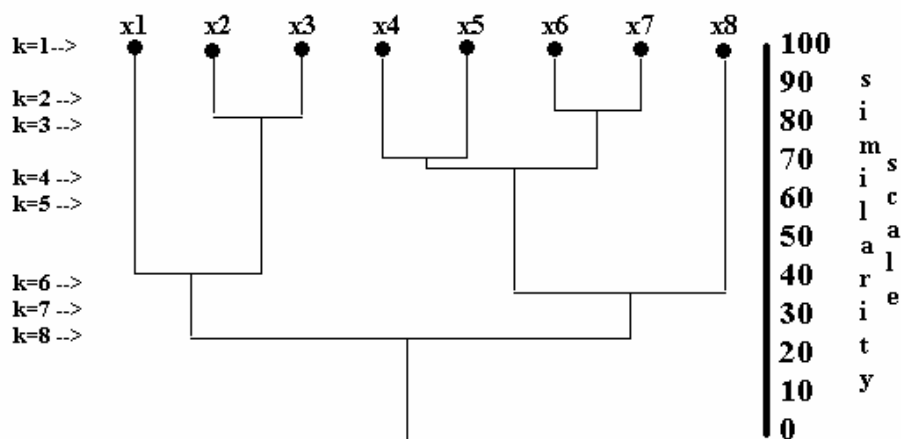
όπου m_h ο αριθμός των αντικειμένων της τάξης h .

Στη συνέχεια παρουσιάζονται μερικές βασικές μέθοδοι ομαδοποίησης.

2.4 ΙΕΡΑΡΧΙΚΕΣ ΜΕΘΟΔΟΙ (HIERARCHICAL CLUSTERING)

Η ιεραρχική ομαδοποίηση είναι η πιο συχνά χρησιμοποιούμενη μέθοδος για την παρουσίαση της δομής των δεδομένων. Ένα χαρακτηριστικό παράδειγμα ιεραρχικής ταξινόμησης είναι στο ζωικό βασίλειο, όπου κάθε είδος ανήκει σε μια σειρά ομάδων αυξανόμενου μεγέθους με μειούμενο αριθμό κοινών χαρακτηριστικών. Αυτές οι μέθοδοι παράγουν ιεραρχικές αναπαραστάσεις στις οποίες οι ομάδες κάθε επιπέδου ιεραρχίας δημιουργούνται από την συγχώνευση ομάδων του αμέσως προηγούμενου επιπέδου. Στο πρώτο επίπεδο, κάθε ομάδα περιέχει ένα μόνο αντικείμενο. Στο υψηλότερο επίπεδο υπάρχει μια μόνο ομάδα που περιέχει όλα τα δεδομένα (Hastie et al., 2001).

Η πιο φυσική απεικόνιση της ιεραρχικής ομαδοποίησης είναι ένα αντίστοιχο δέντρο που ονομάζεται δενδρόγραμμα, το οποίο δείχνει πως είναι ομαδοποιημένα τα δεδομένα. Το σχήμα 2.1 παρουσιάζει ένα δενδρόγραμμα για ένα απλό πρόβλημα που περιλαμβάνει οκτώ αντικείμενα. Στο επίπεδο $k=1$ φαίνονται τα 8 αντικείμενα σε 8 ομάδες. Στο επίπεδο 2, τα αντικείμενα x_6 και x_7 έχουν καταταχθεί σε μία ομάδα, και είναι μαζί σε όλα τα επόμενα επίπεδα. Εάν είναι δυνατόν να υπολογισθεί η ομοιομορφία μεταξύ των ομάδων, τότε το δενδρόγραμμα σχεδιάζεται υπό κλίμακα για να φανεί η ομοιότητα μεταξύ των ομάδων που έχουν συνενωθεί. Στο σχήμα 2.1, για παράδειγμα, η ομοιότητα μεταξύ των δύο ομάδων αντικειμένων που προκύπτουν στο επίπεδο 5 έχει περίπου τιμή 60 (Duda et al., 2001).



Σχήμα 2.1: Παράδειγμα δένδρογραμματος

Υπάρχουν αρκετοί διαφορετικοί αλγόριθμοι για την ανεύρεση ενός ιεραρχικού δένδρου. Ένας συσσωρευτικός αλγόριθμος (*agglomerative algorithm*) ξεκινάει με n υποομάδες, κάθε μία από τις οποίες περιέχει ένα στοιχείο, και σε κάθε βήμα συγχωνεύει τα δυο πιο όμοια σύνολα για το σχηματισμό μιας νέας ομάδας, μειώνοντας έτσι τον αριθμό των ομάδων κατά μία. Ο αλγόριθμος προχωράει μέχρι όλα τα αντικείμενα να ανήκουν στην ίδια ομάδα (Webb, 2002). Στο σχήμα 2.2 παρουσιάζεται η βασική μορφή ενός συσσωρευτικού ιεραρχικού αλγόριθμου.

Αντίθετα ένας επιμεριστικός αλγόριθμος (*divisive algorithm*) λειτουργεί διαχωρίζοντας διαδοχικά τα σύνολα. Ξεκινάει με ένα σύνολο και συνεχίζει μέχρι να προκύψουν n ομάδες του ενός αντικειμένου. Γενικά, αυτοί οι αλγόριθμοι είναι υπολογιστικά μη εφικτοί (Webb, 2002).

Βασικός συσσωρευτικός ιεραρχικός αλγόριθμος

1. Υπολογισμός του πίνακα γειτνίασης
 2. Επανάληψη
 3. Ένωση των δύο κοντινότερων ομάδων
 4. Ενημέρωση του πίνακα γειτνίασης ώστε να ανταποκρίνεται στις αποστάσεις μεταξύ της νέας ομάδας και των υπαρχόντων ομάδων
 5. Μέχρι να παραμείνει μια μόνο ομάδα
-

Σχήμα 2.2: Βασικός συσσωρευτικός ιεραρχικός αλγόριθμος (Tan et al., 2006)

Ανάλογα με το κριτήριο που χρησιμοποιείται για να καθοριστεί ποιες ομάδες πρέπει να συνδυαστούν σε κάθε στάδιο, και με τον τρόπο που εκτιμώνται οι αποστάσεις μεταξύ των ομάδων, μπορούν τα τελικά αποτελέσματα να είναι πολύ διαφορετικά. Όλα

τα κριτήρια που χρησιμοποιούνται στηρίζονται σε έναν πίνακα αποστάσεων μεταξύ ζευγών παρατηρήσεων:

- *Κριτήριο πλησιέστερου γείτονα (nearest neighbor method)*: Δύο αντικείμενα a και b ανήκουν στην ίδια ομάδα στο επίπεδο d αν υπάρχει μια αλυσίδα ενδιάμεσων αντικειμένων $i_1, \dots, i_{m-1}$, που τα ενώνει έτσι ώστε όλες οι αποστάσεις $d_{ik, ik+1} \leq d$ για $k=0, \dots, m-1$, όπου $i_0=a$ και $i_m=b$. (Webb, 2002) Σε ένα συσσωρευτικό αλγόριθμο ο παραπάνω ορισμός σημαίνει ότι, σε κάθε επίπεδο του αλγόριθμου οι δύο πιο κοντινές ομάδες ενώνονται σε μια νέα ομάδα. Ως απόσταση μεταξύ των δύο ομάδων θεωρείται η απόσταση μεταξύ των πιο κοντινών μελών τους.
- *Κριτήριο απώτερου γείτονα (furthest neighbor method)*: Η απόσταση δύο ομάδων A και B είναι η απόσταση μεταξύ των δύο πιο απομακρυσμένων σημείων από κάθε ομάδα.
- *Μέθοδος αθροίσματος τετραγώνων (sum of squares method)*: Σε κάθε στάδιο του αλγόριθμου τα δύο σύνολα που έχουν τη μικρότερη αύξηση στη μέση τιμή των αποστάσεων μεταξύ όλων των ζευγών των αντικειμένων τους συνενώνονται.

2.5 ΔΙΑΧΩΡΙΣΤΙΚΗ ΟΜΑΔΟΠΟΙΗΣΗ (PARTITIONAL CLUSTERING)

Η διαχωριστική ομαδοποίηση αφορά το διαχωρισμό του συνόλου των δεδομένων σε υποσύνολα (ομάδες) έτσι ώστε κάθε αντικείμενο να ανήκει σε ένα ακριβές υποσύνολο. Μία από τις πιο δημοφιλείς μεθόδους ομαδοποίησης αυτή της κατηγορίας είναι ο αλγόριθμος K-means, καθώς και οι παραλλαγές του.

Παρακάτω δίδεται μια σύντομη περιγραφή του βασικού αλγόριθμου (σχήμα 2.3): Αρχικά επιλέγονται K αρχικά «κέντρα» ομάδων, όπου το K είναι μια παράμετρος η οποία καθορίζεται από τον ερευνητή και έχει να κάνει με τον επιθυμητό αριθμό ομάδων. Κάθε σημείο κατατάσσεται στο κοντινότερο κέντρο, οπότε κάθε συλλογή σημείων που έχει καταταγεί σε ένα κέντρο αποτελεί μια ομάδα. Τότε το κέντρο κάθε ομάδας επαναπροσδιορίζεται με βάση τα σημεία που έχουν καταχωρηθεί σε κάθε ομάδα. Τα βήματα της καταχώρισης των σημείων και ο επαναπροσδιορισμός των κέντρων επαναλαμβάνονται μέχρι να μην αλλάζουν τα κέντρα (Tan et al., 2006).

Βασικός K-means αλγόριθμος

1. Επιλογή K σημείων ως αρχικά κέντρα
2. Επανάληψη
3. Σχηματισμός K ομάδων κατατάσσοντας κάθε σημείο στο κοντινότερο κέντρο
4. Επαναπροσδιορισμός των κέντρων κάθε ομάδας.
5. Μέχρι τα κέντρα να μην αλλάζουν

Σχήμα 2.3: Βασικός K-means αλγόριθμος (Tan et al., 2006)

Στη συνέχεια περιγράφεται πιο αναλυτικά κάθε βήμα του αλγόριθμου.

Η επιλογή των κατάλληλων αρχικών κέντρων είναι βήμα κλειδί του βασικού αλγορίθμου K-means. Μια συνηθισμένη προσέγγιση είναι η τυχαία επιλογή κέντρων, όμως συχνά τα αποτελέσματα δεν είναι ικανοποιητικά. Γι' αυτό έχουν προταθεί και άλλες μέθοδοι αρχικοποίησης. Μια άλλη μέθοδος, όπως την περιγράφουν οι Tan et al. (2006), είναι η επιλογή ενός δείγματος σημείων και η ομαδοποίησή τους χρησιμοποιώντας τεχνικές ιεραρχικής ομαδοποίησης. Στη συνέχεια K ομάδες λαμβάνονται από την ιεραρχική ομαδοποίηση και τα κέντρα αυτών των ομάδων χρησιμοποιούνται σαν αρχικά κέντρα στον αλγόριθμο K-means.

Για να καταταγεί ένα σημείο στο κοντινότερο κέντρο, χρειάζεται ένα μέτρο που να ποσοτικοποιεί την έννοια της απόστασης για τα δεδομένα που μελετώνται κάθε φορά. Ενώ μπορεί να υπάρξουν διάφοροι τύποι μέτρησης της εγγύτητας για τον ίδιο τύπο δεδομένων. Για παράδειγμα, για σημεία που βρίσκονται στον ευκλείδειο χώρο, μπορεί να χρησιμοποιηθεί η ευκλείδεια απόσταση ή η απόσταση Manhattan.

Ο επαναπροσδιορισμός των κέντρων εξαρτάται από τον τρόπο υπολογισμού της εγγύτητας δεδομένων και κέντρων, καθώς επίσης και από τον στόχο της ομαδοποίησης. Ο στόχος της ομαδοποίησης συνήθως εκφράζεται από μια αντικειμενική συνάρτηση, η οποία περιλαμβάνει την απόσταση των σημείων μεταξύ τους ή με τα κέντρα των ομάδων. Ένας στόχος, για παράδειγμα, είναι η ελαχιστοποίηση της τετραγωνικής απόστασης κάθε σημείου από το κοντινότερο κέντρο του.

Σε γενικές γραμμές, τα βασικά πλεονεκτήματα αυτού του αλγόριθμου είναι η απλότητά και η ταχύτητά του, που του επιτρέπουν να επεξεργάζεται μεγάλα σύνολα δεδομένων. Ενώ, το μειονέκτημά του είναι δεν βγάζει το ίδιο αποτέλεσμα πάντα, αφού το τελικό αποτέλεσμα εξαρτάται σε μεγάλο βαθμό από την αρχική τυχαία επιλογή κέντρου. Ανάλογα με την μέθοδο αρχικοποίησης, μπορεί να ελαττωθεί η διαφορετικότητα του αποτελέσματος, αλλά δεν μπορεί να διασφαλιστεί ότι θα υπάρξει ολικό ελάχιστο.

Εκτός από τον βασικό αλγόριθμο K-means, υπάρχει και η ασαφής K-means ομαδοποίηση (fuzzy k-means clustering), στην οποία κάθε σημείο έχει ένα ποσοστό συμμετοχής σε κάθε ομάδα, και δεν ανήκει εξολοκλήρου σε μία ομάδα. Έτσι, τα σημεία που βρίσκονται στις «άκρες» μιας ομάδας, μπορεί να ανήκουν στην ομάδα σε μικρότερο βαθμό από αυτά που βρίσκονται κοντά στο κέντρο της. (http://en.wikipedia.org/wiki/Data_clustering)

2.6 ΕΞΕΛΙΚΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ

Εκτός από τους κλασσικούς αλγόριθμους ομαδοποίησης, για την αντιμετώπιση προβλημάτων διαχωριστικής ομαδοποίησης (partitional clustering), έχουν χρησιμοποιηθεί και εξελικτικοί αλγόριθμοι. Οι εξελικτικοί αλγόριθμοι, όπως και οι γενετικοί (Holland, 1975), είναι ευρετικοί αλγόριθμοι στοχαστικής αναζήτησης εμπνευσμένοι από τη θεωρία εξέλιξης του Δαρβίνου και τη γενετική. Η κύρια ιδέα είναι η δημιουργία ενός πληθυσμού υποψηφίων λύσεων σε ένα πρόβλημα βελτιστοποίησης, ο οποίος βελτιώνεται επαναληπτικά μέσω τροποποίησης και επιλογής των καλών

λύσεων για να χρησιμοποιηθούν στην επόμενη επανάληψη. Οι υποψήφιες λύσεις επιλέγονται σύμφωνα με την αντικειμενική συνάρτηση, η οποία εκτιμά την ποιότητα την ποιότητά τους με γνώμονα το δεδομένο πρόβλημα βελτιστοποίησης.

Σημαντικό πλεονέκτημα αυτών των αλγορίθμων είναι η ικανότητά τους να αντεπεξέρχονται στην τοπική αναζήτηση διατηρώντας, επανασυνδιάζοντας και συγκρίνοντας διαφορετικές λύσεις ταυτόχρονα.

Όλοι οι εξελικτικοί αλγόριθμοι έχουν δύο σημαντικά χαρακτηριστικά που τους ξεχωρίζουν από άλλους αλγόριθμους αναζήτησης. Καταρχήν, όλοι βασίζονται σε πληθυσμούς λύσεων. Δεύτερον, γίνεται ανταλλαγή πληροφοριών και επικοινωνία μεταξύ των ατόμων ενός πληθυσμού. Αυτή ή επικοινωνία και η ανταλλαγή πληροφοριών είναι αποτέλεσμα επιλογής και/ ή επανασυνδιασμού στους εξελικτικούς αλγόριθμους (Sarker et al.,2002).

Βασικός εξελικτικός αλγόριθμος

1. Θέσε $i=0$

2. Παραγωγή αρχικού πληθυσμού $P(i)$ τυχαία

3. ΕΠΑΝΑΛΗΨΗ

(α) Αξιολόγηση κάθε ατόμου στον $P(i)$

(β) Επιλογή γονέων από $P(i)$ βάσει της αξίας τους

(γ) Εφαρμογή διαδικασιών αναζήτησης στους γονείς και παραγωγή της γενεάς $P(i+1)$

4. **MEXPI** ο πληθυσμός να συγκλίνει ή να φτάσουμε στο μέγιστο αριθμό επαναλήψεων

Σχήμα 2.4: Βασικός εξελικτικός αλγόριθμος (Sarker et al.,2002)

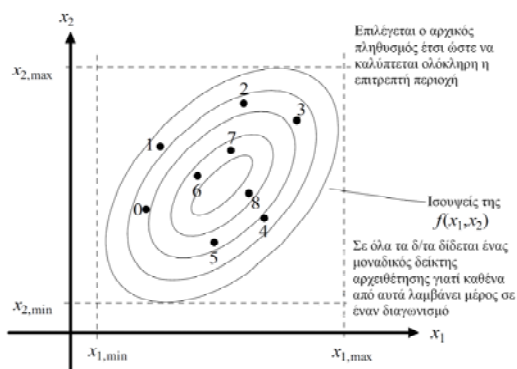
Ένας εξελικτικός αλγόριθμος που έχει χρησιμοποιηθεί σε προβλήματα ομαδοποίησης, πρωτοπαρουσιάστηκε από τους Price και Storn το 1995 και είναι η μέθοδος της Διαφορικής Εξέλιξης (Differential Evolution – ΔΕ). Η βασική ιδέα της ΔΕ είναι η διαδοχική ανανέωση ενός πληθυσμού λύσεων μέσω πρόσθεσης, αφαίρεσης και ανταλλαγής συστατικών, μέχρι ο πληθυσμός να συγκλίνει στο βέλτιστο (σχήμα 2.4).

Στη γενική της μορφή, στη ΔΕ προεπιλέγονται τα όρια των παραμέτρων τα οποία ορίζουν την περιοχή από την οποία επιλέγονται τυχαία Np διανύσματα του αρχικού πληθυσμού (Σχήμα 2.2: 1). Κάθε διάνυσμα σημαίνεται με ένα αριθμό από 0 έως $Np-1$. Όπως και άλλες μεθόδους που βασίζονται στον πληθυσμό, η ΔΕ παράγει νέα σημεία τα οποία είναι διαταραχές των υπάρχοντων σημείων. Συγκεκριμένα η ΔΕ, ταρασσει διανύσματα με την σταθμισμένη διαφορά δύο τυχαία επιλεγμένων διανυσμάτων του πληθυσμού (Σχήμα 2.2: 2). Για την παραγωγή δοκιμαστικού διανύσματος u_0 , η ΔΕ προσθέτει τη σταθμισμένη διαφορά των τυχαίων διανυσμάτων σε ένα τρίτο τυχαίο διάνυσμα του πληθυσμού (Σχήμα 2.2: 3). Στη φάση της επιλογής, το δοκιμαστικό διάνυσμα ανταγωνίζεται έναντι του διανύσματος του πληθυσμού με την ίδια σήμανση, όπου στη συγκεκριμένη περίπτωση είναι το 0. Η εικόνα 4 του σχήματος 2.2, δείχνει το βήμα επιλογής και αποθήκευσης κατά το οποίο το διάνυσμα με τη μικρότερη τιμή

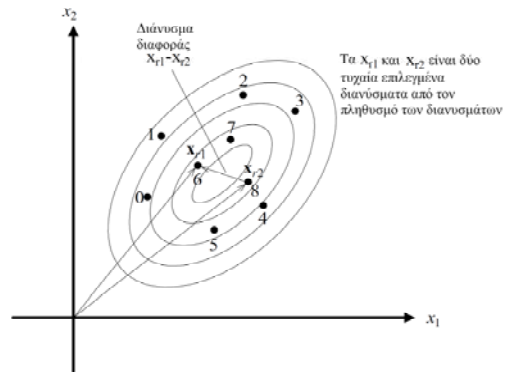
αντικειμενικής συνάρτησης μαρκάρεται σαν μέλος της επόμενης γενεάς. Στις εικόνες 5 και 6, φαίνεται ότι η διαδικασία επαναλαμβάνεται μέχρι Np διανύσματα να έχουν ανταγωνιστεί έναντι ενός τυχαία παραγόμενου δοκιμαστικού διανύσματος. Μόλις δοκιμαστεί και το τελευταίο δοκιμαστικό διάνυσμα, οι Np επιζώντες γίνονται γονείς για την επόμενη γενεά του κύκλου εξέλιξης. (Price et al, 2005)

Παρόλο που το παραπάνω βασικό σχέδιο της ΔΕ δουλεύει αρκετά καλά, η απόδοσή της μπορεί να βελτιωθεί αν η μεθοδολογία της προσαρμοστεί στο εκάστοτε πρόβλημα βελτιστοποίησης.

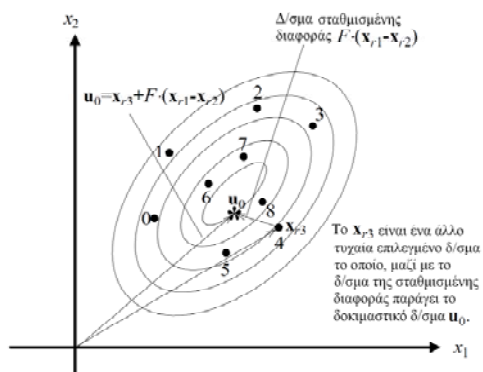
Συγκεκριμένα, σε συγκριτικά πειράματα που έχουν γίνει για προβλήματα ομαδοποίησης μεταξύ της ΔΕ, του αλγορίθμου K-means, γενετικών αλγορίθμων, ενός άλλου ευρετικού αλγόριθμου του particle swarm και τυχαιοποιημένης αναζήτησης, διαπιστώθηκε ότι η ΔΕ είναι ξεκάθαρα ανώτερη από τις υπόλοιπες μεθόδους, όσον αφορά την ακρίβεια αλλά και την ισχύ των λύσεων. Δηλαδή, προκύπτουν παρόμοια αποτελέσματα σε επαναλαμβανόμενες εκτελέσεις του αλγόριθμου. Το τελευταίο είναι πολύ σημαντικό για τον χρήστη, αφού ένας αλγόριθμος αναζήτησης δεν αρκεί να είναι μόνο ακριβής, αλλά πρέπει να παράγει αξιόπιστες και αναπαραξιμες λύσεις (Paterlini et al., 2006).



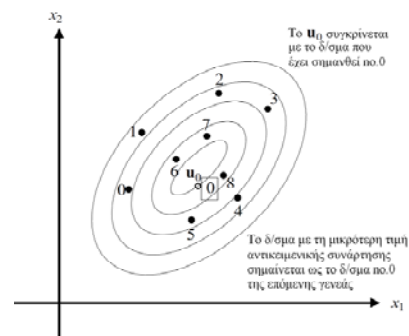
1. Αρχικοποίηση πληθυσμού



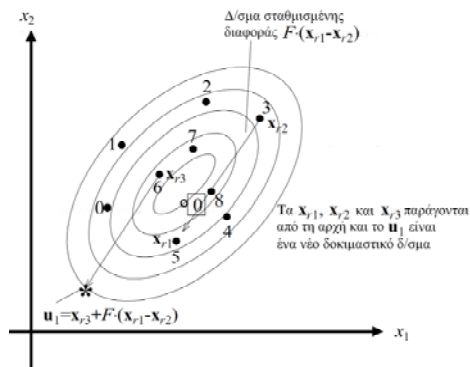
2. Παραγωγή διαταραχής



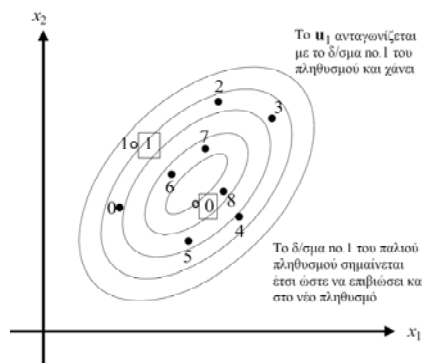
3. Μετάλλαξη



4. Επιλογή



5. Παραγωγή με μετάλλαξη ενός νέου πληθυσμού διανυσμάτων με τυχαία διατάραξη



6. Επιλογή

Σχήμα 2.5: Διαφορική Εξέλιξη (Price et al, 2005)

ΚΕΦΑΛΑΙΟ 3

ΠΟΛΥΚΡΙΤΗΡΙΑ ΟΜΑΔΟΠΟΙΗΣΗ

3.1 ΕΙΣΑΓΩΓΗ

Στο προηγούμενο κεφάλαιο έγινε μια σύντομη παρουσίαση του προβλήματος της ομαδοποίησης, καθώς και ορισμένων τεχνικών επίλυσής του. Ωστόσο, ιδιαίτερο ενδιαφέρον παρουσιάζουν οι τεχνικές ανάλυσης ομάδων με πολυκριτήριο χαρακτήρα.

Σε ένα «παραδοσιακό» πρόβλημα ομαδοποίησης οι πληροφορίες που χρησιμοποιούνται για την περιγραφή των αντικειμένων ή εναλλακτικών είναι εντελώς αντικειμενικές. Ένα αντικείμενο περιγράφεται από p μεταβλητές οι οποίες χρησιμοποιούνται για την εξαγωγή μιας δομής μεταξύ όλων των αντικειμένων. Από την άλλη μεριά, σε ένα πολυκριτήριο σύστημα αποφάσεων, η απόσταση μεταξύ των αντικειμένων βασίζεται σε μεγάλο βαθμό από το σύστημα προτίμησης του εκάστοτε αποφασίζοντα.

Παρόλο που έχουν γίνει πολλές μελέτες στα πλαίσια της πολυκριτήριας ανάλυσης, για την κατάταξη εναλλακτικών σε σύνολα που έχουν προκαταβολικά οριστεί, όπως γίνεται στο πρόβλημα της ταξινόμησης, δεν έχει δοθεί μέχρι στιγμής τόση προσοχή σε προβλήματα μη επιτηρούμενης μάθησης (unsupervised learning) με πολυκριτήριο χαρακτήρα, όπως η ομαδοποίηση.

Ένα πρόβλημα πολυκριτήριας ομαδοποίησης μπορεί να προσεγγιστεί με πολλούς διαφορετικούς τρόπους (Doreian et al., 2005):

- Χαλαρώνοντας το σε πρόβλημα ομαδοποίησης με ένα κριτήριο το οποίο να προέρχεται από το συνδυασμό όλων των κριτηρίων.
- Χρησιμοποιώντας αλγορίθμους ομαδοποίησης με περιορισμούς, όπου ένα επιλεγμένο κριτήριο θεωρείται το κριτήριο ομαδοποίησης και τα άλλα ορίζουν τους περιορισμούς
- Χρησιμοποιώντας τεχνικές αποδοχής ομαδοποίησης που χρησιμοποιούνται στις ομαδοποίησης που προκύπτουν από αλγορίθμους ομαδοποίησης που χρησιμοποιούν κάθε φορά ένα κριτήριο.
- Δημιουργώντας και χρησιμοποιώντας αλγορίθμους ειδικά για αυτό το πρόβλημα.

Στη συνέχεια θα παρουσιαστούν επιλεκτικά μερικές ενδιαφέρουσες τεχνικές επίλυσης του συγκεκριμένου προβλήματος, οι οποίες να χρησιμοποιούν τις πληροφορίες προτίμησης του αποφασίζοντα, αφού πρώτα γίνει η διατύπωσή του.

3.2 ΔΙΑΤΥΠΩΣΗ ΠΡΟΒΛΗΜΑΤΟΣ

Σε ένα πολυκριτήριο πρόβλημα ομαδοποίησης έχουμε ένα σύνολο εναλλακτικών (ενέργειες ή αντικείμενα):

$$A = \{x_1, \dots, x_i, \dots, x_n\}$$

οι οποίες αξιολογούνται από p κριτήρια:

$$F = \{f_1, \dots, f_j, \dots, f_p\}$$

Θεωρείται ότι όλα τα κριτήρια είναι προς μεγιστοποίηση.

Τα παραπάνω δεδομένα σε ένα πολυκριτήριο πρόβλημα παρουσιάζονται σε ένα πίνακα αξιολόγησης, στον οποίο οι γραμμές αναπαριστούν τις ενέργειες και οι στήλες τα κριτήρια:

Ενέργειες x_i	$f_1(x_i)$	$f_2(x_i)$		$f_j(x_i)$		$f_n(x_i)$
x_1	$f_1(x_1)$	$f_2(x_1)$		$f_j(x_1)$		$f_n(x_1)$
x_2	$f_1(x_2)$	$f_2(x_2)$		$f_j(x_2)$		$f_n(x_2)$
x_i	$f_1(x_i)$	$f_2(x_i)$		$f_j(x_i)$		$f_n(x_i)$
x_p	$f_1(x_p)$	$f_2(x_p)$		$f_j(x_p)$		$f_n(x_p)$

Σε αυτόν τον πίνακα το $f_j(x_i)$ υποδηλώνει την αξία της ενέργειας x_i στο κριτήριο f_j .

Ο ρόλος που παίζει κάθε κριτήριο f_j σε μια διαδικασία λήψης αποφάσεων ορίζεται από το συντελεστή σχετικής σημαντικότητας w_j , που είναι προσαρτημένο στο κριτήριο f_j . Οι συντελεστές είναι γνωστοί ως βάρη γιατί όσο μεγαλύτερη η τιμή, τόσο πιο σημαντικό είναι το κριτήριο. Τα βάρη είναι κανονικοποιημένα:

$$w_j \geq 0 \quad \forall j=1,2,\dots,p, \quad \sum_{j=1}^m w_j = 1$$

Πολλές μέθοδοι πολυκριτήριας ομαδοποίησης βασίζονται σε συγκρίσεις των εναλλακτικών ανά δύο. Η προτίμηση της μιας εναλλακτικής έναντι της άλλης ορίζεται μέσω μια συνάρτησης προτίμησης:

$$P_j(x, y) = \begin{cases} 0 & \text{εάν } x_j < y_j \\ g_j(d_j(x, y)) & \text{εάν } x_j \geq y_j \end{cases}, \quad \forall x, y \in A$$

Όπου $d_j(x, y)$ είναι η διαφορά των ενεργειών x και y στο κριτήριο j .

Για τη μορφή της συνάρτησης g_j έχουν προταθεί 6 περιπτώσεις (γενικευμένα κριτήρια). Ανάλογα με το γενικευμένο κριτήριο που χρησιμοποιείται κατά περίπτωση μπορούν να οριστούν οι εξής παράμετροι:

- κατώφλι αδιαφορίας q_j ,
- κατώφλι προτίμησης p_j ,
- μια ενδιάμεση παράμετρος σ_j που καθορίζει το σημείο αλλαγής στην καμπή στη συνάρτηση του γενικευμένου κριτηρίου (το κριτήριο του Gauss).

Με βάση αυτές τις πληροφορίες προτίμησης, επιχειρείται ο διαχωρισμός των εναλλακτικών σε k ομογενή σύνολα, τα οποία δεν είναι εκ των προτέρων γνωστά:

$$C = \{C_1, \dots, C_h, \dots, C_k\}$$

Κάθε ομάδα προσδιορίζεται από ένα σημείο αναφοράς $r_1, \dots, r_h, \dots, r_k$, το οποίο ανάλογα με την τεχνική που εφαρμόζεται υπολογίζεται με διαφορετικό τρόπο. Οι εναλλακτικές μέσα στην ίδια ομάδα πρέπει να μοιάζουν πιο πολύ μεταξύ του, απ' ό,τι με τις εναλλακτικές που βρίσκονται σε διαφορετικά σύνολα.

3.3 ΜΙΑ ΠΡΟΕΚΤΑΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ K-MEANS

Οι De Smet και Guzman (2004), πρότειναν μια επέκταση του αλγορίθμου k-mean σε ένα πολυκριτήριο πλαίσιο εργασίας. Αυτή η επέκταση βασίζεται στον ορισμό μίας πολυκριτηρίας απόστασης, η οποία προκύπτει από τη δομή προτίμησης που ορίζει ο αποφασίζοντας.

Το μοντέλο προτίμησης που χρησιμοποιείται λαμβάνει υπόψην του τις ακόλουθες σχέσεις: Προτίμηση (Preference- P), Αδιαφορία (Indifference- I), και Ασύγκριτικότητα (Incomparability- J), οι οποίες προκύπτουν από τη σύγκριση μεταξύ δύο ενεργειών x_i και x_l . Οι τρεις αυτές σχέσεις συνιστούν ένα μοντέλο προτίμησης για το σύνολο A , αν για κάθε δύο στοιχεία x_i και x_l του A , μια και μόνο μια από τις παρακάτω ιδιότητες είναι αληθής:

- $x_i P x_l$, δηλαδή η x_i προτιμάται από την x_l .
- $a_j P x_l$, δηλαδή η a_j προτιμάται από την x_l .
- $x_i I x_l$, δηλαδή η x_i είναι αδιάφορη από την x_l .
- $x_i J x_l$, δηλαδή η x_i ασύγκριτη με την x_l .

Για το διαχωρισμό των εναλλακτικών, οι De Smet και Guzman θεωρούν ότι όλες οι ενέργειες μέσα στην ίδια ομάδα πρέπει να είναι προτιμητέες, αδιάφορες ή ασύγκριτες περίπου στις ίδιες ενέργειες, βάσει τις προτιμήσεις του αποφασίζοντα.

Για την ποσοτικοποίηση της ομοιότητας των εναλλακτικών ορίζεται μια πολυκριτήρια απόσταση. Αυτή η απόσταση βασίζεται στην έννοια του προφίλ κάθε ενέργειας και σε μια δομή προτίμησης.

Προφίλ Προτίμησης

Το προφίλ $P(x_i)$ της ενέργειας $x_i \in A$ ορίζεται ένας συνδυασμός τεσσάρων συνιστοσών $\langle J(x_i), P(x_i), I(x_i), P^+(x_i) \rangle$, όπου:

- $J(x_i) = \{x_l \in A \mid x_i J x_l\} = P_1(x_i)$,
- $P(x_i) = \{x_l \in A \mid x_i P x_l\} = P_2(x_i)$,
- $I(x_i) = \{x_l \in A \mid x_i I x_l\} = P_3(x_i)$,
- $P^+(x_i) = \{x_l \in A \mid x_i P x_l\} = P_4(x_i)$.

Απόσταση

Ο ορισμός της απόστασης γίνεται λαμβάνοντας υπόψιν τον πολυκριτήριο χαρακτήρα των εναλλακτικών και η βασική σύλληψη είναι ότι δύο ενέργειες είναι τόσο κοντά όσο τα προφίλ τους.

Αν $P(x_i)$ είναι το προφίλ της x_i , η απόσταση μεταξύ δύο ενεργειών x_i και x_l ορίζεται ως εξής:

$$d(x_i, x_l) = 1 - \frac{\sum_{k=1}^4 |P_k(x_i) \cap P_k(x_l)|}{n}$$

Υπολογισμός σημείου αναφοράς

Η κεντρική ιδέα της μεθόδου είναι η αναπαράσταση κάθε πιθανής ομάδας από ένα εικονικό κεντρικό στοιχείο. Οι ενέργειες κατατάσσονται στην ομάδα με το κοντινότερο κεντρικό στοιχείο. Έτσι, νέες ομάδες φτιάχνονται και νέα κεντρικά σημεία προσδιορίζονται, τα προφίλ των οποίων είναι όσο το δυνατόν πιο κοντά στα προφίλ των αντικειμένων που ανήκουν σε κάθε ομάδα. Η διαδικασία επαναλαμβάνεται μέχρι η δομή των ομάδων να παραμείνει αμετάβλητη. Στη συγκεκριμένη περίπτωση ο ορισμός του κεντρικού στοιχείου (πρότυπο) βασίζεται σε μια διαδικασία ψήφισης.

Αν $\{x_{h1}, x_{h2}, \dots, x_{hn}\}$ είναι οι n ενέργειες της h ομάδας, το προφίλ του κεντρικού σημείου της ομάδας προσδιορίζεται από την ακόλουθη διαδικασία ψηφοφορίας:

$$a_j \in P_k(C_h) \Leftrightarrow k = \arg \max \sum_{a_{hl}} 1_{\{a_j \in P_k(a_{hl})\}}$$

όπου, $1_{\{A\}} = 1$ αν η υπόθεση A είναι αληθής, 0 σε διαφορετική περίπτωση. Αν διαφορετικές τιμές του k ικανοποιούν την προηγούμενη συνθήκη, η τελική τιμή θα επιλεγεί τυχαία.

Αλγόριθμος

Παρακάτω περιγράφεται η διαδικασία επίλυσης πολυκριτηρίων προβλημάτων ομαδοποίησης. Τα βήματα αυτής της τεχνικής μοιάζουν πάρα πολύ με τον αλγόριθμο k-means, που παρουσιάστηκε νωρίτερα.

Ο k-means αλγόριθμος για προβλήματα πολυκριτήριας ομαδοποίησης

1. Ορισμός αριθμού ομάδων K
2. Ορισμός των προφίλ $P(x_i), \forall x_i$
3. Τυχαία επιλογή ενός αρχικού συνόλου k σημείων αναφοράς $(y_1, y_2, \dots, y_k) | \forall j, \exists h \ y_j = x_h$
4. **Επανάληψη**
5. Προσδιορισμός της ομαδοποίησης P με τέτοιο τρόπο ώστε να κατατάσσει κάθε εναλλακτική με βάση το πλησιέστερο πρότυπο
6. Για κάθε ομάδα $C_h \in P$ υπολόγισε το κεντρικό σημείο r_h^* , το οποίο γίνεται το νέο σημείο αναφοράς κάθε ομάδας C_h .
7. **Μέχρι** τα σημεία αναφοράς να μην αλλάζουν

Σχήμα 3.1: Ο k-means για προβλήματα πολυκριτήριας ομαδοποίησης (Nemery, 2005)

Η προέκταση αυτή του k-means αλγορίθμου παρουσιάζει τα ίδια μειονεκτήματα όπως και ο αρχικός αλγόριθμος. Δηλαδή, δεν υπάρχει μοναδικότητα στην ομαδοποίηση, δεν μπορεί να προσδιοριστεί ο αριθμός των ομάδων και παρατηρείται ευαισθησία των ενεργειών που βρίσκονται στις παρυφές κάθε ομάδας. (Nemery, 2005)

3.4 PROMETHEE CLUSTER

Ο μέθοδος PROMETHEE Cluster είναι μια προέκταση της PROMETHEE TRI σε περιβάλλον μη επιτηρούμενης μάθησης. Η PROMETHEE TRI επιτρέπει την κατάταξη ενεργειών σε κατηγορίες που έχουν οριστεί εκ των προτέρων με τη βοήθεια ενεργειών αναφοράς. Οι ενέργειες συγκρίνονται αυτές τις ενέργειες αναφοράς και μετά κατατάσσονται σε μια κατηγορία.

Από την άλλη μεριά, η PROMETHEE Cluster, διαχωρίζει τις ενέργειες σε ομοιογενή σύνολα χωρίς να ορίζονται αυτά προκαταβολικά και λαμβάνοντας υπ' όψιν τις τιμές των προτιμήσεων. Είναι λοιπόν μια μη επιτηρούμενη τεχνική μάθησης σε πολυκριτήριο πλαίσιο.

Στις μεθόδους PROMETHEE υπολογίζονται συναρτήσεις προτίμησης $P_j(x,y)$ για όλα τα ζεύγη ενεργειών και όλα τα κριτήρια, οπότε λαμβάνεται ο δείκτης προτίμησης $\pi(x,y)$, ο οποίος δείχνει το βαθμό που προτιμάται η x έναντι της y :

$$\pi(x,y) = \sum_{j \in F} P_j(x,y)w_j, \pi(x,y) \in [0,1]$$

Όταν ο δείκτης προτίμησης $\pi(x,y)$ είναι κοντά στο ένα τότε υπάρχει ισχυρή υπεροχή της x έναντι της y . Αντίθετα, όταν είναι κοντά στο μηδέν, υπάρχει οριακή υπεροχή.

Η κεντρική ιδέα των μεθόδων PROMETHEE είναι η ποσοτικοποίηση της υπεροχής μιας ενέργειας x έναντι όλων των υπολοίπων και της υπεροχής όλων των

υπολοίπων ενεργειών έναντι της x . Αυτή η ιδέα οδηγεί στον ορισμό της ροής εξόδου, της ροής εισόδου καθώς επίσης και της καθαρής ροής κάθε ενέργειας:

- Ροή εξόδου (η υπεροχή της x έναντι των υπολοίπων ενεργειών):

$$\phi^+(x) = \sum_j \pi(x, y_j)$$

- Ροή εισόδου (η υπεροχή των υπολοίπων ενεργειών έναντι της x):

$$\phi^-(x) = \sum_j \pi(y_j, x)$$

- Καθαρή ροή (το συνολικό μέγεθος αξιολόγησης της x έναντι των υπολοίπων ενεργειών):

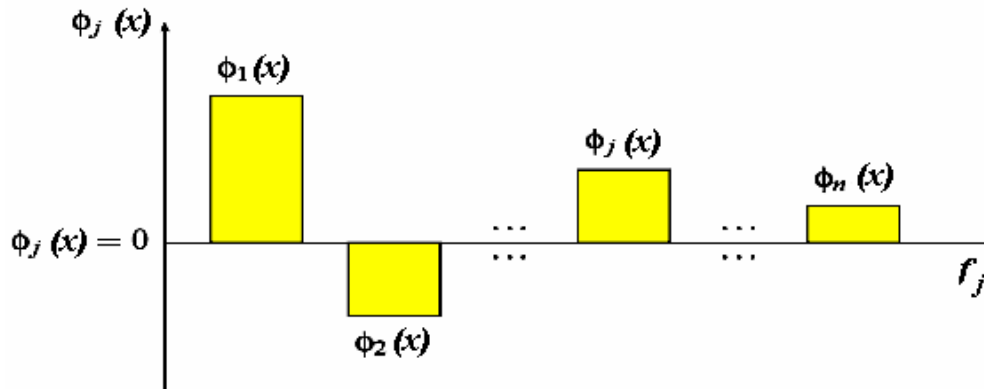
$$\phi(x) = \phi^+(x) + \phi^-(x)$$

Στις μεθόδους PROMETHEE TRI και PROMETHEE Cluster χρησιμοποιείται και η έννοια της *καθαρής ροής ενός κριτηρίου*, η οποία ορίζεται ως εξής:

$$\phi_j(x) = \frac{1}{n-1} \sum_{x \in A} (P_j(x, y) - P_j(y, x))$$

Στην ουσία δηλαδή η καθαρή ροή ενός κριτηρίου ποσοτικοποιεί τη θέση της ενέργειας x ως προς το κριτήριο f_j σε σχέση με όλες τις άλλες ενέργειες του συνόλου A .

Ο προσδιορισμός της καθαρής ροής για όλα τα κριτήρια επιτρέπει τον ορισμό του *προφίλ* μιας ενέργειας (σχήμα 3.2), το οποίο βοηθάει τον αποφασίζοντα να εκτιμήσει την «ποιότητά» της. Το προφίλ εξάγεται από τις καθαρές ροές $\phi_j(x)$ του κάθε κριτηρίου και εκφράζει την υπεροχή της ενέργειας έναντι των υπολοίπων ($\phi_j(x) > 0$) ή την υπεροχή των υπολοίπων ενεργειών έναντι της x ($\phi_j(x) < 0$).



Σχήμα 3.2: Το προφίλ μιας ενέργειας

Ο ορισμός των κατηγοριών ή των ομάδων στις μεθόδους PROMETHEE TRI και PROMETHEE Cluster γίνεται με τη βοήθεια ενεργειών αναφοράς ή κεντρικών ενεργειών $r_1, \dots, r_h, \dots, r_k$, οι οποίες είναι αντιπροσωπευτικές των κατηγοριών. Οι κεντρικές ενέργειες δεν είναι πραγματικές, συνήθως είναι φανταστικές. Η κατάταξη μια ενέργειας a σε μια κατηγορία γίνεται συγκρίνοντας το προφίλ της x_a με τα προφίλ όλων

των ενεργειών αναφοράς. Αυτό γίνεται ορίζοντας τη διαφοροποίηση μεταξύ της ενέργειας x και μιας ενέργειας αναφοράς r_h ως το άθροισμα των διαφορών για όλα τα κριτήρια:

$$e(x, r_h) = \sum_{j \in F} |\phi_j(x) - \phi_j(r_h)| w_j$$

Μια ενέργεια x_i ταξινομείται στην ομάδα C_z που η διαφοροποίηση είναι ελάχιστη:

$$x_i \in C_z \text{ αν } e(x_i, r_{zi}) = \min_{h=1, \dots, k} \{e(x_i, r_h)\}$$

Η κατάταξη λοιπόν των ενεργειών σε κατηγορίες σύμφωνα με την PROMETHEE TRI αποτελείται από δύο φάσεις. Κατά την πρώτη φάση υπολογίζεται η καθαρή ροή κάθε κριτηρίου για κάθε εναλλακτική και κατά τη δεύτερη φάση κατατάσσεται κάθε εναλλακτική στην κατηγορία από την οποία έχει τη μικρότερη διαφοροποίηση.

Η επέκταση της PROMETHEE TRI σε πρόβλημα ομαδοποίησης είναι η PROMETHEE Cluster, η οποία χρησιμοποιεί και στοιχεία της προέκτασης του k-means αλγορίθμου που περιγράφηκε παραπάνω (σχήμα 3.1). Η συγκεκριμένη προσέγγιση διαχωρισμού του συνόλου A σε k ομάδες περιγράφεται ως εξής από τους Figueira et al (2003):

1. Ο αποφασίζοντας καθορίζει τον αριθμό των ομάδων k
2. Θεωρούνται k τυχαία σημεία αναφοράς
3. Εφαρμόζεται η μέθοδος PROMETHEE TRI και το αποτέλεσμα είναι k ομάδες
4. Ορίζεται ξανά ένα πιο κατάλληλο σημείο αναφοράς σε κάθε ομάδα βάσει των ενεργειών που περιλαμβάνει
5. Εφαρμόζεται ξανά η μέθοδος PROMETHEE TRI με τα νέα σημεία αναφοράς

Η διαδικασία ολοκληρώνεται όταν δεν μεταβάλλεται πλέον η σύσταση των ομάδων.

Η κατάταξη μια ενέργειας σε μια συγκεκριμένη ομάδα γίνεται με τον ίδιο τρόπο όπως και στην PROMETHEE TRI. Επιπλέον το νέο κεντρικό σημείο αναφοράς, r_h , κάθε ομάδας προσδιορίζεται από τη διάμεσο των ροών των ενεργειών της κάθε ομάδας:

$$\phi_j(r_h) = \text{median}\{\phi_j(x_{h_q}), q = 1, \dots, l\}$$

όπου $C_h = \{x_{h1}, \dots, x_{hl}\}$. Αυτό διασφαλίζει ότι το r_h είναι το σημείο που ελαχιστοποιεί το άθροισμα των (L^w_I) αποκλίσεων μέσα στην ομάδα h :

$$E = \sum_{q=1}^l \sum_{j \in F} |\phi_j(r_h) - \phi_j(x_{h_q})| w_j$$

PROMETHEE Cluster

Εισερχόμενα: $A=\{x_1, \dots, x_i, \dots, x_n\}$, και $\varphi_j(x_i)$, $\forall x_i \in A, \forall j \in F$

Αρχικοποίηση k τυχαία κεντρικά σημεία αναφοράς $H=\{r_1, \dots, r_h, \dots, r_k\}$

Κάθε ομάδα C_h σχετίζεται με ένα σημείο αναφοράς r_h

Επανάληψη

Για όλα τα εισερχόμενα διανύσματα x_i του A

Κατατάσσεται το x_i στην ομάδα C_h με τη μικρότερη απόκλιση από r_h

Για όλες τις ομάδες C_h , όπου $h \in \{1, \dots, k\}$

Ανανεώνονται τα σημεία αναφοράς r_h

Μέχρι η σύσταση των ομάδων να μην μεταβάλλεται

Σχήμα 3.3: PROMETHEE Cluster (Figueira et al, 2003)

Ως προέκταση του αλγορίθμου k-means, η PROMETHEE Cluster χαρακτηρίζεται από τα ίδια μειονεκτήματα.

Αυτή η μέθοδος είναι η πρώτη που ασχολείται με προβλήματα ομαδοποίησης με πολυκριτήριο χαρακτήρα και λαμβάνει υπ' όψιν της τα κριτήρια ανεξάρτητα. Παρολαυτά, τα προφίλ των ενεργειών αναφοράς κατασκευάζονται καταρχήν με σύγκριση ενός συγκεκριμένου κριτηρίου. Μόνο οι επιλεγμένες ενέργειες αναφοράς λαμβάνονται υπ' όψιν. Όταν υπολογίζεται η καθαρή ροή μια ενέργειας για ένα κριτήριο, χρησιμοποιούνται μόνο οι ενέργειες αναφοράς. Υπολογίζεται τότε η απόκλιση της ενέργειας με ένα σταθμισμένο άθροισμα. Γι' αυτό η επιλογή των αρχικών ενεργειών αναφοράς, όπως επίσης και ο αριθμός των ομάδων επηρεάζουν σημαντικά το αποτέλεσμα. (Nemery, 2006)

3.5 ΔΙΑΤΕΤΑΓΜΕΝΗ ΠΟΛΥΚΡΙΤΗΡΙΑ ΟΜΑΔΟΠΟΙΗΣΗ

Στα πλαίσια τις πολυκριτήριας ομαδοποίησης έχει δοθεί πολύ έμφαση στην κατάταξη αντικειμένων σε διατεταγμένα σύνολα που είναι εκ των προτέρων ορισμένα. Αυτού του είδους τα προβλήματα ονομάζονται ταξινόμηση και τα σύνολα κατηγορίες. Έχουν υπάρξει διάφορες προσεγγίσεις για αυτού του είδους τα προβλήματα, όπως η ELECTRE TRI, UTADIS, PROMETHEE TRI κ.α. Κοινό χαρακτηριστικό όλων αυτών των μεθόδων είναι ο χαρακτηρισμός κάθε συνόλου από τα «όριά» του ή από «κεντρικά» αντικείμενα.

Αντίθετα δεν έχει δοθεί τόση έμφαση μέχρι στιγμής στην ανάπτυξη μεθόδων για το πρόβλημα του ορισμού αυτών των κατηγοριών. Οι τεχνικές λοιπόν που βοηθούν τον αποφασίζοντα να καταλήξει σε διατεταγμένες ομάδες αντικειμένων είναι η Διατεταγμένη Πολυκριτήρια Ομαδοποίηση (multicriteria ordered clustering).

Μια πρώτη προσπάθεια ανεύρεσης διατεταγμένων ομάδων θα ήταν καταρχήν η κατάταξη των αντικειμένων και στη συνέχεια ο διαχωρισμός τους με βάση τη θέση τους στην κατάταξη. Έχουν χρησιμοποιηθεί ευρέως πολλά στατιστικά μοντέλα για την αντιμετώπιση του προβλήματος της κατάταξης, όπως η ανάλυση παλινδρόμησης, ανάλυση ομάδων και άλλα. Όμως, οι στατιστικές μέθοδοι παρουσιάζουν ορισμένα μειονεκτήματα, όπως για παράδειγμα, ο ορισμός εξαρτημένων μεταβλητών και ο αποκλεισμός σημαντικών παραγόντων. Η δημιουργία κατηγοριών από τα καταταγμένα αντικείμενα μπορεί να οδηγήσει σε διάφορα προβλήματα ασυνέχειας και απαιτεί για παράδειγμα να έχει δοθεί ο αριθμός των αντικειμένων σε κάθε κατηγορία. (Nemery, 2006)

Μια μέθοδος που έχει αναπτυχθεί για την αντιμετώπιση των προβλημάτων διατεταγμένης πολυκριτήριας ομαδοποίησης είναι από τους Nemery και De Smet (2005), βασίζεται στη μέθοδο PROMETHEE.

Καταρχήν, υπολογίζεται ένας πίνακας σύγκρισης των εναλλακτικών, όπως και στην PROMETHEE, με βάση τη συνολική προτίμηση μιας ενέργειας από μια άλλη. Η προτίμηση μιας ενέργειας x_i από μια άλλη ενέργεια x_l για το κριτήριο f_j δίνεται από τη σχέση:

$$P_{j(x_i, x_l)} = \min[\max(0, \frac{d_{j(x_i, x_l)} - q_l}{p_l - q_l}), 1]$$

όπου το $d_{j(x_i, x_l)} = f_{j(x_i)} - f_{j(x_l)}$ αναπαριστά τη διαφορά μεταξύ της ενέργειας x_i από την ενέργεια x_l στο κριτήριο f_j . Η συνολική προτίμηση της ενέργειας x_i από την ενέργεια x_l υπολογίζεται ως εξής:

$$\Pi_{(x_i, x_l)} = \Pi_{il} = \sum_{j=1}^p w_j \cdot P_{j(x_i, x_l)}$$

Όταν ολοκληρωθούν αυτές οι συγκρίσεις για κάθε ζεύγος εναλλακτικών του A , μπορεί να κατασκευαστεί ο πίνακας προτιμήσεων $\Pi = (\pi_{il})$. Ένας διατεταγμένος διαχωρισμός του συνόλου των εναλλακτικών A σε K ομάδες, που συμβολίζεται $P = [C_1, C_2, \dots, C_k]$, ορίζεται ως:

- $A = \bigcup_{h=1, \dots, K} C_h$, δηλαδή όλες οι εναλλακτικές του συνόλου A θα καταταγούν σε κάποια ομάδα.
- $C_i \cap C_j = \emptyset$, δηλαδή οι ομάδες που θα δημιουργηθούν δεν θα έχουν κοινές εναλλακτικές

- $C_1 \succ C_2 \succ \dots \succ C_K$, δηλαδή η ομάδα C_1 έχει χαμηλότερη κατάταξη από την ομάδα C_2 . Θεωρείται ότι όσο μικρότερη κατάταξη έχει μια ομάδα, τόσο καλύτερη είναι

Το πρόβλημα της διατεταγμένης ομαδοποίησης περιλαμβάνει τον εντοπισμό του διατεταγμένου διαχωρισμού P του A που είναι πιο συμβατός με τις πληροφορίες που περιέχονται στον πίνακα προτιμήσεων Π . Η συμβατότητα αυτή μπορεί να εκφραστεί με δυο δείκτες που σχετίζονται με την ετερογένεια και τη συνάφεια.

Από τη μία, οι ενέργειες που ανήκουν στην ίδια ομάδα μπορούν να χαρακτηριστούν από σχετικά μικρούς βαθμούς προτίμησης. Στην ιδανική περίπτωση, δυο ενέργειες x_i και x_l , μεταξύ των οποίων ο αποφασίζοντας δεν μπορεί να δει ξεκάθαρη διαφορά (ιδανικά $\pi_{il} = \pi_{li}$), πρέπει να τοποθετηθούν μαζί στην ίδια ομάδα C_k . Η παραβίαση του παραπάνω χαρακτηριστικού που πρέπει να έχει ένα διαχωρισμός μπορεί να εκφραστεί από το *συνολικό δείκτη ετερογένειας* HI_P :

$$HI_P = \max_{\forall i=1, \dots, K} (HI_h)$$

Όπου HI_h είναι ο δείκτης ετερογένειας κάθε ομάδας και δίνεται από τη σχέση:

$$HI(C_h) = HI_h = \max_{\forall i, l | x_i, x_l \in C_h} (\pi_{il})$$

Όπως είναι προφανές, όσο μικρότερο είναι το HI_h , τόσο καλύτερη είναι η ομάδα C_h από άποψη ομοιογένειας.

Από την άλλη, αν μια ενέργεια x_i προτιμάται ισχυρά έναντι μιας άλλης x_l , δεν είναι αποδεκτό αυτές οι δύο εναλλακτικές να βρίσκονται στην ίδια ομάδα. Αντίθετα θα πρέπει να ανήκουν σε διαφορετικές ομάδες $x_i \in C_h$ και $x_l \in C_m$, όπου $C_h \prec C_m$, δηλαδή να καταταγούν σε ομάδες η σειρά προτίμησης των οποίων να είναι συμβατή με τη σειρά προτίμησης των εναλλακτικών. Ένας *συνολικό δείκτης ασυμβατότητας* μπορεί να είναι ο ακόλουθος:

$$CI_P = \max_{\forall m > h} CI_{mh}$$

Όπου CI_{mh} είναι ο δείκτης ασυμβατότητας μεταξύ των ομάδων C_m και C_h :

$$CI(C_m, C_h) = CI_{mh} = \max_{\forall x_i \in C_m, \forall x_l \in C_h | \forall h, m: m > h} (\pi_{il})$$

Φυσικά, μικρότερη τιμή CI_{mh} σημαίνει καλύτερη συμβατότητα μεταξύ των ομάδων C_m και C_h .

Τέλος, ποιοτική βαθμολόγηση ενός διαχωρισμού P μπορεί να γίνει ορίζοντας έναν δείκτη καταλληλότητας FI_P ως η μέγιστη τιμή μεταξύ του δείκτη ετερογένειας και του δείκτη ασυμβατότητας:

$$FI_P = \max(HI_P, CI_P).$$

Όσο μικρότερος είναι ο FI_p τόσο καλύτερος ο διαχωρισμός P .

Η παραπάνω τεχνική, λόγω της πολυπλοκότητας του προβλήματος, μπορεί να συνδυαστεί με έναν εξελικτικό αλγόριθμο βελτιστοποίησης.

ΚΕΦΑΛΑΙΟ 4

ΕΞΕΛΙΚΤΙΚΟΣ ΑΛΓΟΡΙΘΜΟΣ

ΠΟΛΥΚΡΙΤΗΡΙΑΣ

ΟΜΑΔΟΠΟΙΗΣΗΣ

4.1 ΕΙΣΑΓΩΓΗ

Στη συγκεκριμένη εργασία επιχειρείται η ομαδοποίηση πολυκριτήριων δεδομένων με συνδυασμό του αλγορίθμου της Διαφορικής Εξέλιξης με μεθόδους πολυκριτήριας ομαδοποίησης, όπως είναι η PROMETHEE Cluster. Αρχικά παρουσιάζεται ο αλγόριθμος της Διαφορικής Εξέλιξης, ενώ στη συνέχεια περιγράφονται οι τρόποι ομαδοποίησης που χρησιμοποιήθηκαν στην παρούσα εργασία, καθώς και δώδεκα διαφορετικές εκδοχές της αντικειμενικής συνάρτησης βάσει της οποίας γίνεται η επιλογή στον αλγόριθμο Δ.Ε.

4.2 ΔΙΑΦΟΡΙΚΗ ΕΞΕΛΙΞΗ

Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο, η βασική ιδέα της ΔΕ είναι η διαδοχική ανανέωση ενός πληθυσμού λύσεων μέσω πρόσθεσης, αφαίρεσης και ανταλλαγής συστατικών, μέχρι ο πληθυσμός να συγκλίνει στο βέλτιστο. Στη συνέχεια γίνεται μια σύντομη παρουσίαση του συγκεκριμένου αλγόριθμου, με έμφαση στα χαρακτηριστικά του εκείνα που τον κάνουν να διαφέρει και σε πολλές περιπτώσεις να είναι πιο αποτελεσματικός από άλλους εξελικτικούς αλγορίθμους.

4.2.1 Δομή Πληθυσμού

Ένα από τα πιο χρήσιμα εργαλεία του αλγόριθμου της Διαφορικής Εξέλιξης είναι ένα ζεύγος πληθυσμών διανυσμάτων, του τρέχοντος πληθυσμού και ενός δοκιμαστικού πληθυσμού, οι οποίοι και οι δύο αποτελούνται από N_p διανύσματα (λύσεις), D διαστάσεων (κριτηρίων).

Ο τρέχων πληθυσμός P_x αποτελείται από τα διανύσματα εκείνα που έχουν ήδη γίνει αποδεκτά, είτε σαν αρχικά σημεία είτε μετά από συνδυασμό να άλλα διανύσματα:

$$P_{x,g} = (x_{i,g}), i = 0, 1, \dots, N_{p-1}, g = 0, 1, \dots, g_{\max}$$

$$x_{i,g} = (x_{j,i,g}), j = 0, 1, \dots, D-1$$

Οι δείκτες ξεκινούν από το 0 για να είναι πιο απλή η εργασία με τους πίνακες. Ο δείκτης $g=0,1,...,g_{max}$ υποδεικνύει τη γενεά ενός πληθυσμού ή ενός διανύσματος.

Κατά την αρχικοποίηση, ο αλγόριθμος μεταλλάσσει τυχαίες επιλεγμένες εναλλακτικές για την παραγωγή ενός ενδιάμεσου πληθυσμού $P_{v,g}$ από N_p μεταλλαγμένες εναλλακτικές $v_{i,g}$:

$$P_{x,g} = (v_{i,g}), \quad i = 0,1,...,N_{p-1}, \quad g = 0,1,...,g_{max}$$

$$v_{i,g} = (x_{j,i,g}), \quad j = 0,1,...,D-1$$

Τότε, κάθε διάνυσμα του τρέχοντος πληθυσμού επανασυνδιάζεται με ένα μεταλλαγμένο για την παραγωγή ενός δοκιμαστικού πληθυσμού P_u από N_p δοκιμαστικά διανύσματα, $u_{i,g}$:

$$P_{x,g} = (u_{i,g}), \quad i = 0,1,...,N_{p-1}, \quad g = 0,1,...,g_{max}$$

$$u_{i,g} = (x_{j,i,g}), \quad j = 0,1,...,D-1$$

Κατά τη διάρκεια του επανασυνδιασμού, τα δοκιμαστικά διανύσματα αντικαθιστούν τον μεταλλαγμένο πληθυσμό, οπότε ένα διάνυσμα αντιπροσωπεύει και τους δύο πληθυσμούς.

4.2.2 Αναπαράσταση παραμέτρων

Ο αλγόριθμος της διαφορικής εξέλιξης κωδικοποιεί όλες τις παραμέτρους, ανεξάρτητα από τον τύπο τους, σαν αριθμητικά σημεία (floating-point number). Ακόμα και ακέραιες αλλά και διακριτές μεταβλητές αναπαριστώνται σαν πραγματικοί αριθμοί για να προσθέσουν ποικιλία στην διασπορά των διαφορών τους. Η κωδικοποίηση συνεχών παραμέτρων σαν σημεία και η διαχείρισή τους με αριθμητικές πράξεις προσφέρει σημαντικά πλεονεκτήματα, όπως:

- Ευκολία στη χρήση
- Αποδοτικότητα στη διαχείριση της μνήμης
- Χαμηλή υπολογιστική πολυπλοκότητα
- Χαμηλή υπολογιστική προσπάθεια
- Μεγαλύτερη ελευθερία στο σχεδιασμό μια μεταλλαγμένης διασποράς.

4.2.3 Αρχικοποίηση

Για να λειτουργεί σωστά ο αλγόριθμος της διαφορικής εξέλιξης, ο αρχικός πληθυσμός πρέπει να είναι διεσπαρμένος σε ολόκληρο τον χώρο των λύσεων. Γι' αυτό, πριν αρχικοποιηθεί ο πληθυσμός, πρέπει να προσδιοριστούν τα άνω και κάτω όρια για κάθε παράμετρο. Αυτές οι τιμές μπορούν να συγκεντρωθούν σε δύο διανύσματα D διαστάσεων b_L και b_U , όπου οι δείκτες L και U υποδεικνύουν τα κάτω και τα άνω όρια αντίστοιχα.

Στα πραγματικά προβλήματα, τα όρια των παραμέτρων σπάνια είναι ίσα, συνήθως γιατί οι παράμετροι αντιστοιχούν σε διαφορετικές μαθηματικές ή φυσικές έννοιες. Σε πολλές περιπτώσεις, η ύπαρξη φυσικών ορίων ή λογικών περιορισμών κάνει εύκολο τον προσδιορισμό των ορίων.

Αφού λοιπόν, προσδιοριστούν τα άνω και κάτω όρια, μια γεννήτρια τυχαίων αριθμών δίνει για κάθε παράμετρο κάθε διανύσματος μια τιμή μέσα στα προκαθορισμένα όρια. Για παράδειγμα, η αρχική τιμή ($g=0$) για την παράμετρο j του διανύσματος i είναι:

$$x_{j,i,0} = rand_j(0,1) \cdot (b_{j,U} - b_{j,L}) + b_{j,L}$$

Η γεννήτρια τυχαίων αριθμών, $rand_j(0,1)$, επιστρέφει ένα σύνολο ομοιόμορφα διεσπαρμένων αριθμών στην κλίμακα $[0,1)$. Ο δείκτης j δείχνει ότι για κάθε παράμετρο παράγεται ένας νέος τυχαίος αριθμός.

Η διασπορά όμως των αρχικών σημείων μπορεί να προτιμηθεί να γίνει με μη ομοιόμορφο τρόπο. Η απόφαση για το είδος της διασποράς εξαρτάται από το κατά πόσο είναι γνωστή η περιοχή του βέλτιστου. Αν η τοποθεσία που βρίσκεται το βέλτιστο είναι πάνω-κάτω γνωστή, μια κατανομή τύπου Gauss μπορεί να αποδειχτεί καλύτερη, παρόλο που υπάρχει ο κίνδυνος ο πληθυσμός να συγκλίνει πρόωρα. Γενικά, οι ομοιόμορφες κατανομές προτιμούνται γιατί ανταποκρίνονται καλύτερα στην έλλειψη γνώσης της περιοχής του βέλτιστου.

4.2.4 Μετάλλαξη

Μετά την αρχικοποίηση, ο αλγόριθμος της διαφορικής εξέλιξης μεταλλάσσει και επανασυνδιάζει τον πληθυσμό για την παραγωγή ενός πληθυσμού Np δοκιμαστικών λύσεων. Συγκεκριμένα, η *διαφοροποιημένη μετάλλαξη* προσθέτει μια σταθμισμένη διαφορά διανυσμάτων, τα οποία έχουν επιλεγεί με τυχαίο τρόπο, σε ένα τρίτο διάνυσμα, το *διάνυσμα βάση*. Παρακάτω φαίνεται πως μπορεί να συνδυαστούν τρία διαφορετικά, τυχαίως επιλεγμένα διανύσματα, για την παραγωγή ενός μεταλλαγμένου διανύσματος, $v_{i,g}$:

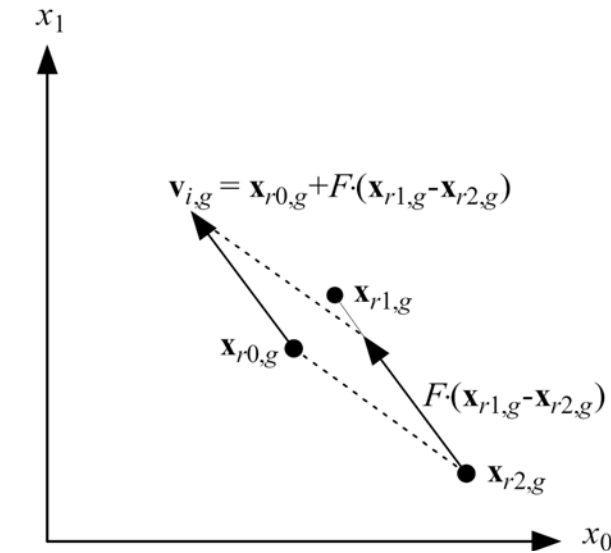
$$v_{i,g} = x_{r0,g} + F \cdot (x_{r1,g} - x_{r2,g})$$

Ο παράγοντας στάθμισης $F \in (0,1+)$, είναι ένας θετικός πραγματικός αριθμός που ελέγχει το ρυθμό εξέλιξης ενός πληθυσμού. Παρόλο που δεν υπάρχει άνω όριο για τον F , σπάνια υπάρχουν αποτελεσματικές τιμές πάνω από του 1.

Το *διάνυσμα βάση* με δείκτη r_0 συνήθως είναι ένα τυχαία επιλεγμένο διάνυσμα, το οποίο πρέπει να είναι διαφορετικό από το *διάνυσμα στόχο* i . Ωστόσο, η τυχαία επιλογή εμπεριέχει τον κίνδυνο κάποια διανύσματα να επιλέγονται συνεχώς ενώ άλλα να παραλειφθούν τελείως, κάτι τι οποίο θα επηρεάσει άμεσα την απόδοση του αλγόριθμου. Για να αποφευχθεί αυτό το διάνυσμα βάση μπορεί να προσδιοριστεί με πολλούς άλλους τρόπους, όπως για παράδειγμα η γενική στοχαστική δειγματοληψία (stochastic universal sampling).

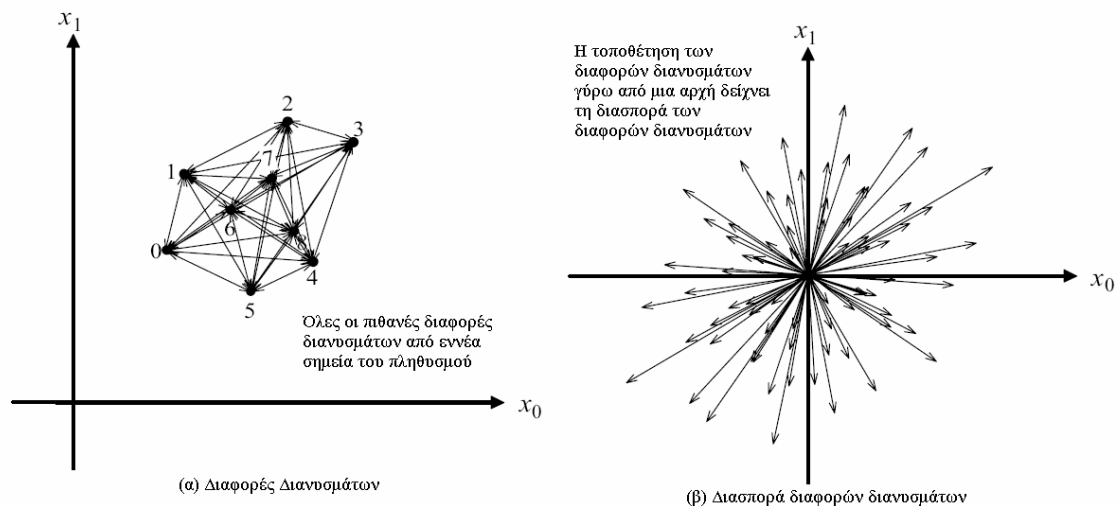
Τα διανύσματα $r1$ και $r2$, από τα οποία προκύπτει το *διάνυσμα διαφορά*, εκτός από ότι πρέπει να είναι διαφορετικά από το διάνυσμα βάση και στόχο, επιλέγονται και

αυτά τυχαία. Το σχήμα 4.1 δείχνει πως κατασκευάζεται το μεταλλαγμένο διάνυσμα $v_{i,g}$, σε ένα χώρο δύο διαστάσεων (δύο παραμέτρων).



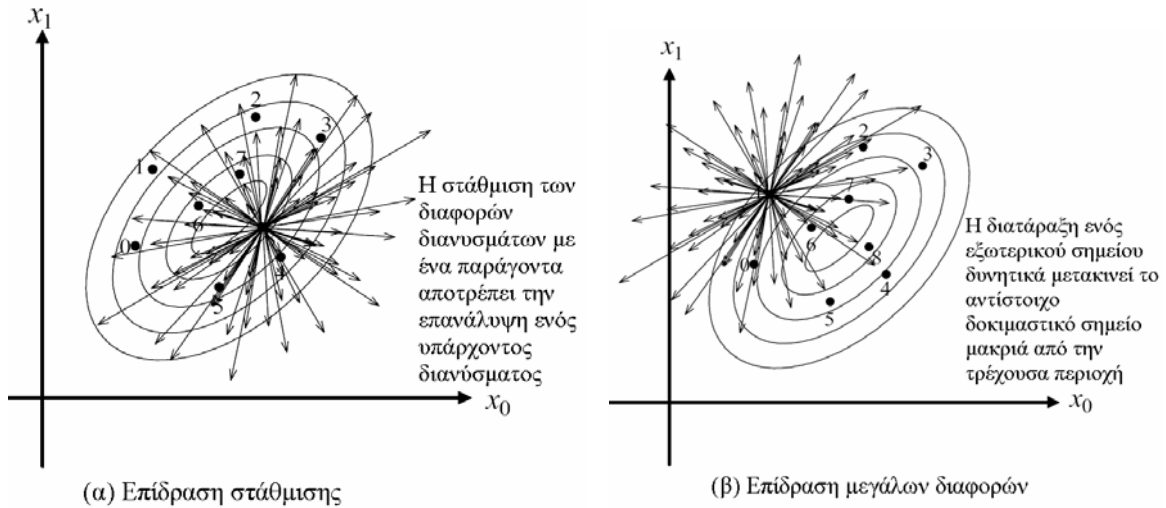
Σχήμα 4.1: Διαφοροποιημένη Μετάλλαξη: η σταθμισμένη διαφορά $F(x_{r1,g} - x_{r2,g})$ προστίθεται στο διάνυσμα βάση $x_{r0,g}$ για την παραγωγή μίας μεταλλαγμένης λύσης $v_{i,g}$. (Price et al, 2005)

Στο σχήμα 4.2 εμφανίζονται τα διανύσματα διαφορών που σχηματίζονται από όλα τα εφικτά ζεύγη εννέα διανυσμάτων. Μεταφέροντας τις διαφορές διανυσμάτων σε μια κοινή αρχή φαίνεται πιο ξεκάθαρα η διασπορά τους. Επειδή όλες οι διαφορές διανυσμάτων έχουν και ένα αντίστροφο και την ίδια πιθανότητα να επιλεγούν, ο μέσος όρος της διασποράς τους είναι μηδέν.



Σχήμα 4.2: (α) Εννέα διανύσματα και (β) η διασπορά των διαφορών διανυσμάτων τους (Price et al, 2005)

Η στάθμιση των διαφορών διανυσμάτων διασφαλίζει ότι τα δοκιμαστικά διανύσματα δεν επαναλαμβάνουν υπαρκτά σημεία (Σχήμα 4.3(α)). Επιπλέον, όπως φαίνεται στο σχήμα 4.3(β), η στάθμιση μπορεί να συμβάλει, λόγω του μεγάλου αριθμού διανυσμάτων διαφορετικού μήκους, στην αποτροπή παγίδευσης σε τοπικό ελάχιστο.



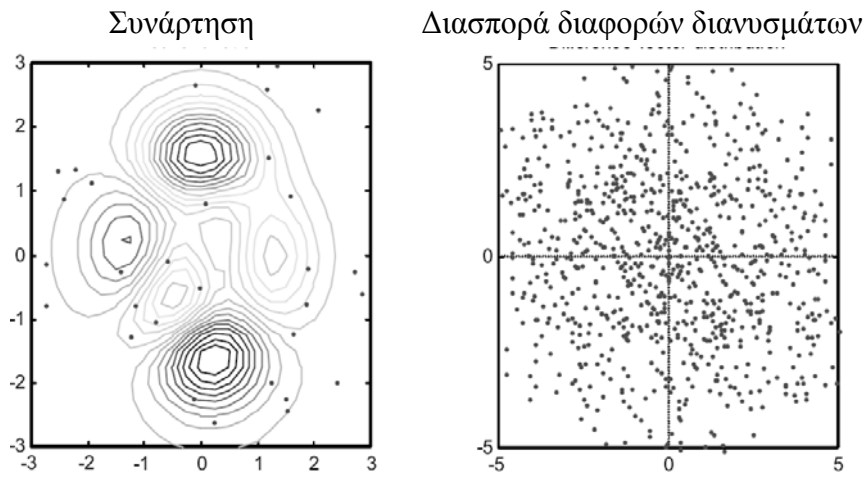
Σχήμα 4.3: (α) Οι επιδράσεις της στάθμισης και (β) των μεγάλων διαφορών διανυσμάτων (Price et al, 2005)

Ένα από τα βασικότερα πλεονεκτήματα των διαφορών διανυσμάτων είναι ότι το μέγεθος του βήματος και ο προσανατολισμός του προσαρμόζονται αυτόματα στο πεδίο της αντικειμενικής συνάρτησης. Το σχήμα 4.4 παρουσιάζει αυτή την ιδιότητα για τη συνάρτηση:

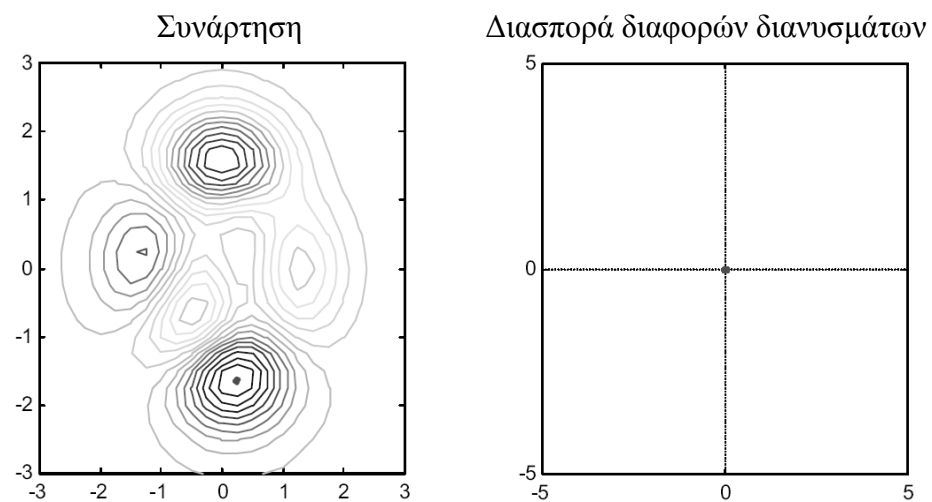
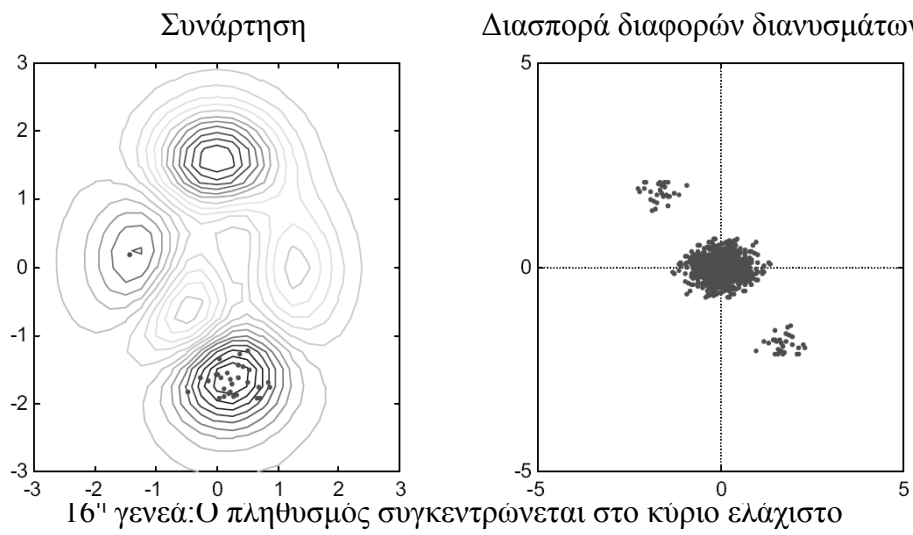
$$f(x_1, x_2) = 3(1 - x_1)^2 \cdot \exp(x_1^2 + (x_2 + 1)^2) - 10\left(\frac{x_1}{5} - x_1^3 - x_2^5\right) \cdot \exp(x_1^2 + x_2^2) - \frac{1}{3} \cdot \exp((x_1 + 1)^2 + x_2^2)$$

Για κάθε γενεά, παρουσιάζονται οι τιμές που λαμβάνει η αντικειμενική συνάρτηση (αριστερό διάγραμμα) και η διασπορά των διαφορών διανυσμάτων (δεξιό διάγραμμα). Για περισσότερη σαφήνεια, στο διάγραμμα των διαφορών διανυσμάτων εμφανίζονται μόνο τα τελικά σημεία των διαφορών.

Καθώς εξελίσσεται, ο πληθυσμός συγκεντρώνεται γύρω από ανταγωνιστικά ελάχιστα. Σε αυτή τη φάση, η διασπορά των διαφορών είναι πολυπαραγοντική, όπως και η συνάρτηση. Περιέχει βήματα προσαρμοσμένα στην αναζήτηση μέσα σε κάθε περιοχή, αλλά και μεγαλύτερα βήματα ικανά να μετακινούν τα διανύσματα από περιοχή σε περιοχή. Μόλις ο πληθυσμός τακτοποιηθεί στο βέλτιστο περιοχή, η διασπορά των διαφορών διανυσμάτων γίνεται μονοπαραγοντική και τα βήματα έχουν το μέγεθος και την κατεύθυνση που απαιτεί η τοπική αναζήτηση.



1^η γενεά: Ο πληθυσμός της Δ.Ε. και η διασπορά των διαφορών διανυσμάτων



34^η γενεά: Η Δ.Ε. βρίσκει το ελάχιστο

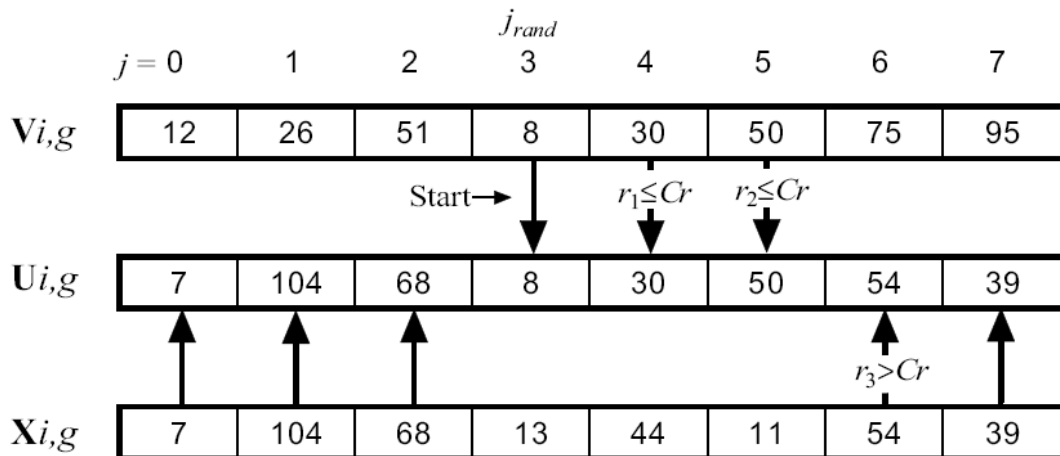
Σχήμα 4.4: Παράδειγμα εξέλιξης αλγόριθμου Δ.Ε. (Price et al, 2005)

4.2.5 Διασταύρωση

Για να ολοκληρωθεί η διαφοροποιημένη μετάλλαξη, ο αλγόριθμος της Διαφορικής Εξέλιξης χρησιμοποιεί *ομοιόμορφη διασταύρωση*. Κατά την διασταύρωση, η οποία αναφέρεται και σαν διακεκριμένος επανασυνδιασμός, φτιάχνονται δοκιμαστικά διάνυσματα από τις τιμές των παραμέτρων που έχουν αντιγραφεί από δύο διαφορετικά διάνυσματα. Αφού αντιγράφονται μόνο πληροφορίες, η διασταύρωση μπορεί να εφαρμοστεί σε δυαδικά, πραγματικά ή ακόμα και συμβολικά δεδομένα.

Η διασταύρωση μπορεί να γίνει με πολλούς διαφορετικούς τρόπους: διασταύρωση ενός σημείου, διασταύρωση N σημείων, εκθετική διασταύρωση και άλλα. Στη συγκεκριμένη εργασία χρησιμοποιήθηκε ο μηχανισμός της εκθετικής διασταύρωσης.

Στη Διαφοροποιημένη Εξέλιξη η εκθετική διασταύρωση ξεκινάει επιλέγοντας τυχαία μια παράμετρο και αντιγράφοντάς την από το μεταλλαγμένο διάνυσμα στη αντίστοιχη δοκιμαστική παράμετρο, έτσι ώστε το δοκιμαστικό διάνυσμα που θα προκύψει να είναι διαφορετικό από το διάνυσμα με το οποίο θα συγκριθεί (δηλαδή το διάνυσμα στόχο, $x_{i,g}$). Η πηγή των δοκιμαστικών παραμέτρων προσδιορίζεται συγκρίνοντας την πιθανότητα διασταύρωσης Cr με έναν αριθμό, οποίος επιλέγεται για κάθε παράμετρο χωριστά από ένα ομοιόμορφα κατανομημένο σύνολο αριθμών μεταξύ 0 και 1. Όσο $rand_j \leq Cr$, οι παράμετροι συνεχίζουν να επιλέγονται από το μεταλλαγμένο διάνυσμα, αλλά από τη πρώτη φορά και έπειτα που $rand_j > Cr$, οι παράμετροι λαμβάνονται από το διάνυσμα στόχο. Το σχήμα 4.5 δείχνει μια περίπτωση στη οποία η εκθετική διασταύρωση παράγει δύο σημεία διασταύρωσης.



Σχήμα 4.5: Εκθετική διασταύρωση. Ξεκινώντας από την τυχαία επιλεγμένη παράμετρο, $j_{rand}(=3)$, οι δοκιμαστικοί παράμετροι κληρονομούνται από το μεταλλαγμένο διάνυσμα $v_{i,j}$, όσο $rand_j \leq Cr$ (π.χ. $j=4, 5$). Από την πρώτη φορά που $rand_j > Cr$, όλες οι εναπομείναντες δοκιμαστικές παράμετροι (π.χ. $j=6, 7, 0, 1, 2$) κληρονομούνται από το διάνυσμα στόχο, $x_{i,g}$. (Price et al, 2005)

Το όνομα εκθετική διασταύρωση ανταποκρίνεται στο γεγονός ότι ο αριθμός των κληρονομημένων παραμέτρων είναι μια εκθετική μεταβλητή τυχαία κατανομημένη. Για παράδειγμα, η πιθανότητα η αρχική, τυχαία επιλεγμένη παράμετρος να είναι η μοναδική παράμετρος του δοκιμαστικού διανύσματος που να προέρχεται από το μεταλλαγμένο διάνυσμα ισούται με την πιθανότητα η πρώτη σύγκριση $rand_j$ και Cr να αποτύχει, δηλαδή $rand_j > Cr$. Έτσι, οι πιθανότητες το αποτέλεσμα μια διασταύρωσης να καταλήξει σε ακριβώς n μεταλλαγμένες παραμέτρους είναι:

$$p(x = n) = (1 - Cr) \cdot Cr^{n-1} = Cr^{n-1} - Cr^n$$

Η πιθανότητα διασταύρωσης $Cr \in [0,1]$, είναι μια τιμή που καθορίζεται από το χρήστη και ελέγχει το ποσοστό των τιμών των παραμέτρων που αντιγράφονται από η μεταλλαγμένη λύση.

4.2.6 Επιλογή

Υπάρχουν δύο φάσεις στη διαδικασία εξέλιξης στις οποίες μπορεί να εφαρμοστεί η διαδικασία της επιλογής. Μερικοί γενετικοί αλγόριθμοι εφαρμόζουν *επιλογή γονέων* για να επιλέξουν ποια διανύσματα θα υποστούν επανασυνδιασμό. Συνήθως τα διανύσματα με τις καλύτερες τιμές αντικειμενικής συνάρτησης έχουν τη μεγαλύτερη πιθανότητα επιλογής, οπότε αυτά είναι και τα πιο πιθανά να επιλεγούν για ζευγάρωμα.

Σε αντίθεση με την επιλογή γονέων, η *επιλογή επιζώντα*, που ονομάζεται και *αντικατάσταση*, επιλέγει την επόμενη γενεά διανυσμάτων από την τρέχουσα γενεά διανυσμάτων και τα δοκιμαστικά διανύσματα.

Στη Διαφορική Εξέλιξη συγκεκριμένα, αν το δοκιμαστικό διάνυσμα $u_{i,g}$, έχει ίση ή μικρότερη τιμή αντικειμενικής συνάρτησης από αυτή του διανύσματος στόχου $x_{i,g}$, τότε αντικαθιστά το διάνυσμα στόχο στην επόμενη γενεά, διαφορετικά διατηρεί τη θέση του στον πληθυσμό για τουλάχιστον άλλη μια γενεά:

$$x_{i,g+1} = \begin{cases} u_{i,g} & \text{αν } f(u_{i,g}) \leq f(x_{i,g}) \\ x_{i,g} & \text{διαφορετικά} \end{cases}$$

Συγκρίνοντας κάθε δοκιμαστικό διάνυσμα με το διάνυσμα στόχο από το οποίο κληρονομεί παραμέτρους, η Διαφορική Εξέλιξη ολοκληρώνει τον επανασυνδιασμό και την επιλογή πιο καλά απ' ότι άλλοι εξελικτικοί αλγόριθμοι.

Μόλις προσδιοριστεί ο νέος πληθυσμός, επαναλαμβάνονται οι διαδικασίες της μετάλλαξης, του επανασυνδιασμού και της επιλογής μέχρι να προσδιοριστεί το βέλτιστο, ή να ικανοποιηθεί κάποιο συγκεκριμένο κριτήριο τερματισμού (για παράδειγμα ο αριθμός των γενεών φτάσει το προκαθορισμένο μέγιστο, g_{max}).

4.2.7 Κριτήρια τερματισμού

Κάποιες φορές είναι προφανές πότε μια βελτιστοποίηση πρέπει να σταματήσει. Για παράδειγμα, στις περιπτώσεις που η τιμή της αντικειμενικής συνάρτησης στο ελάχιστο είναι εκ των προτέρων γνωστή, η βελτιστοποίηση σταματά όταν η τιμή της αντικειμενικής συνάρτησης πλησιάσει την ελάχιστη τιμή της μέσα σε κάποια

συγκεκριμένα όρια ανοχής. Συνήθως όμως δεν είναι γνωστή η ελάχιστη τιμή της αντικειμενικής συνάρτησης. Σε αυτή την περίπτωση χρειάζεται να οριστεί κάποιο κριτήριο που να προσδιορίζει πότε να σταματήσει ο αλγόριθμος:

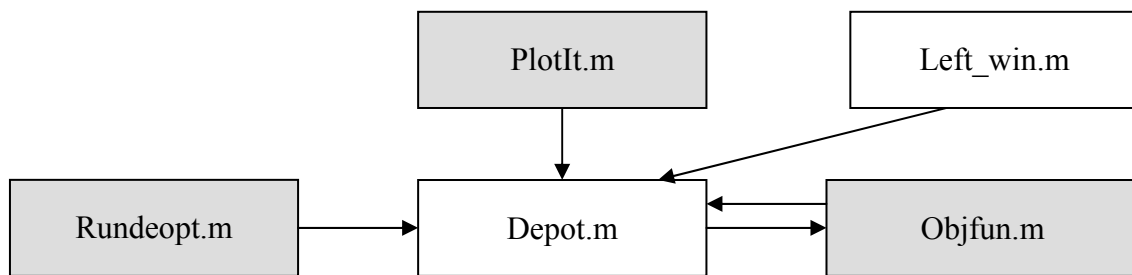
Μερικά κριτήρια τερματισμού είναι:

- Συχνά η βελτιστοποίηση τερματίζει μετά από ένα συγκεκριμένο αριθμό επαναλήψεων $g_{\max}$. Εναλλακτικά η βελτιστοποίηση μπορεί να σταματήσει όταν έχουν περάσει $\Delta g_{\max}$ επαναλήψεις στις οποίες δεν έχει γίνει αποδεκτό κάποιο δοκιμαστικό διάνυσμα. Με πειραματισμούς προσδιορίζεται μια καλή τιμή για το $g_{\max}$ ή το $\Delta g_{\max}$.
- Μια βελτιστοποίηση μπορεί να τερματιστεί όταν ένα μέγεθος που σχετίζεται με τον πληθυσμό φτάσει μια προκαθορισμένη τιμή. Για παράδειγμα, μπορεί να τερματιστεί όταν η διαφορά μεταξύ της χειρότερης και της καλύτερης τιμής της αντικειμενικής συνάρτησης ενός πληθυσμού πέσει κάτω από ένα προκαθορισμένο όριο. Η μέθοδος αυτή απαιτεί ιδιαίτερη προσοχή προκειμένου να μην τερματιστεί πρόωρα η βελτιστοποίηση.
- Μερικές φορές είναι διαθέσιμος μόνο συγκεκριμένος χρόνος για την βελτιστοποίηση. Σε αυτές τις περιπτώσεις πρέπει να τερματιστεί ανεξάρτητα από την κατάσταση του πληθυσμού ή τον αριθμό των επαναλήψεων.
- Λόγω της έμφυτης αβεβαιότητας για το πότε μια βελτιστοποίηση τελειώνει, συνήθως βοηθάει η προσωπική παρακολούθηση συγκεκριμένων δεδομένων. Η παρακολούθηση της τιμής της αντικειμενικής συνάρτησης, ο αριθμός των δοκιμαστικών διανυσμάτων που γίνονται αποδεκτά σε κάθε γενιά κ.α., συνήθως κάνουν εμφανές το πότε πρέπει να τερματιστεί η βελτιστοποίηση.

4.2.8 Υλοποίηση σε περιβάλλον MATLAB

Παρακάτω περιγράφεται ο κώδικας DeMat, όπως περιγράφεται από τους Price et al (2005), ο οποίος αποτελείται από μια συλλογή αρχείων MATLAB που παρέχουν ένα σκελετό για την επίλυση προβλημάτων βελτιστοποίησης με τη Διαφορική Εξέλιξη. Ο συγκεκριμένος κώδικας χρησιμοποιήθηκε, με τις κατάλληλες προσαρμογές, που θα παρουσιαστούν σε παρακάτω κεφάλαιο, και στην παρούσα εργασία για την πραγματοποίηση πολυκριτήριας ομαδοποίησης.

Ο κώδικας DeMat αναπτύχθηκε για να κάνει εύκολη την εφαρμογή της ΔΕ σχεδόν σε οποιοδήποτε πρόβλημα βελτιστοποίησης. Αποτελείται από τα αρχεία που φαίνονται στο Σχήμα 4.12.



Σχήμα 4.12: Τα αρχεία DeMat. Τα αρχεία που είναι σκιαγραφημένα γκρι πρέπει να προσαρμόζονται σε κάθε νέο πρόβλημα. Το αρχείο left_win.m μπορεί να χρειάζεται τροποποίηση όταν να αλλάζει η επιλογή των κριτηρίων

Σε συντομία τα αρχεία είναι:

- Deopt.m: Αυτό το αρχείο περιέχει τη βασική μηχανή της ΔΕ μαζί με διάφορες διαδικασίες μετάλλαξης και επιλογής. Ακόμα διαχειρίζεται τη διασταύρωση και εφαρμόζει τα όρια των περιορισμών. Το αρχείο τέλος, ελέγχει την εξαγωγή των αποτελεσμάτων.
- Objfun.m: Αυτό το αρχείο περιέχει την αντικειμενική συνάρτηση, η οποία υπολογίζει τις τιμές της αντικειμενικής συνάρτησης (κόστος) και αξιολογεί τους περιορισμούς. Τα αποτελέσματα της αξιολόγησης της αντικειμενικής συνάρτησης επιστρέφονται με τη δομή S_MSE.
- Left_win.m: Αυτή είναι μια συνάρτηση που ορίζει κάτω από ποιες συνθήκες επιλέγεται το δοκιμαστικό διάλυση. Η επιλογή αυτή βασίζεται στους περιορισμούς και τις τιμές της αντικειμενικής συνάρτησης.
- Rundeopty.m: Αυτό είναι το κύριο αρχείο για την πραγματοποίηση βελτιστοποίησης και για τον πειραματισμό με στρατηγικές της ΔΕ, διαστάσεις πληθυσμού κ.α.
- PlotIt.m: Αυτό το αρχείο περιλαμβάνει βοηθητικές λειτουργίες για τη γραφική παρουσίαση της πορείας της βελτιστοποίησης μέσω κατάλληλων γραφημάτων.

4.3 ΠΟΛΥΚΡΙΤΗΡΙΑ ΟΜΑΔΟΠΟΙΗΣΗ

Στην παρούσα εργασία εφαρμόστηκαν δύο μέθοδοι για τον διαχωρισμό των εναλλακτικών σε ομάδες, η μια βασίζεται στις αποστάσεις από τα σημεία αναφοράς των ομάδων και η άλλη στις ροές της μεθόδου PROMETHEE.

Για το διαχωρισμό των εναλλακτικών με βάση τις αποστάσεις από τα σημεία αναφοράς κάθε ομάδας χρησιμοποιήθηκε η Ευκλείδεια απόσταση:

$$d(x_i, r_h) = (x_i - r_h)^2$$

Υπολογίζονται δηλαδή οι αποστάσεις όλων των εναλλακτικών από κάθε σημείο αναφοράς μιας ομάδας και στη συνέχεια εντάσσεται κάθε ενέργεια στην πλησιέστερη ομάδα.

Αντίστοιχα στην περίπτωση που η ομαδοποίηση των εναλλακτικών γίνεται με τη χρήση των ροών της μεθόδου PROMETHEE, η μέτρηση της διαφοροποίησης μεταξύ των εναλλακτικών και του σημείου αναφοράς κάθε ομάδας γίνεται όπως προτείνεται από τους Figueira et.al. (2003), και έχει περιγραφεί αναλυτικά στην ενότητα 3.4:

$$e(x, r_h) = \sum_{j \in F} |\phi_j(x) - \phi_j(r_h)| w_j$$

Δηλαδή, η διαφοροποίηση μεταξύ της ενέργειας x και ενός σημείου αναφοράς r_h εκφράζεται από το άθροισμα των σταθμισμένων διαφορών για όλα τα κριτήρια.

Στη συνέχεια κάθε εναλλακτική x_i ταξινομείται στην ομάδα C_z που η διαφορά μεταξύ της εναλλακτικής και του σημείου αναφοράς r_z της ομάδας είναι ελάχιστη:

$$x_i \in C_z \text{ αν } e(x_i, r_{zi}) = \min_{h=1, \dots, k} \{e(x_i, r_h)\}$$

Για τον προσδιορισμό της μεθόδου ομαδοποίησης εισάγεται στο αρχείο depot.m η παράμετρος euclassif. Όταν δίδεται στη συγκεκριμένη παράμετρο η τιμή 0 η ομαδοποίηση γίνεται με βάση τις ροές της PROMETHEE, ενώ όταν δίδεται η τιμή 1 γίνεται βάσει τις αποστάσεις από τα σημεία αναφοράς.

4.4 ΑΝΤΙΚΕΙΜΕΝΙΚΗ ΣΥΝΑΡΤΗΣΗ

Η αντικειμενική συνάρτηση βάσει της οποίας γίνεται η επιλογή των καλύτερων λύσεων βασίζεται στον δείκτη FI_p που παρουσιάζεται από τους Nemery και De Smet (2005), εφαρμόζοντάς τον όμως με διάφορες παραλλαγές.

Όπως έχει αναφερθεί και στην ενότητα 3.5, η ποιοτική βαθμολόγηση ενός διαχωρισμού P μπορεί να γίνει ορίζοντας έναν δείκτη καταλληλότητας FI_p ως τη μέγιστη τιμή μεταξύ του δείκτη ετερογένειας και του δείκτη ασυμβατότητας: $FI_p = \max(HI_p, CI_p)$.

Όσο μικρότερος είναι ο δείκτης FI_p τόσο καλύτερος ο διαχωρισμός P .

Ο συνολικός δείκτης ετερογένειας HI_p δίδεται από τη σχέση:

$$HI_p = \max_{\forall h=1, \dots, K} (HI_h), \text{ όπου}$$

$$HI_h = \max_{\forall i, l | x_i, x_l \in C_h} (\pi_{il}) \text{ ο δείκτης ετερογένειας κάθε ομάδας}$$

Όπως είναι προφανές, όσο μικρότερος είναι ο δείκτης HI_h , τόσο καλύτερη είναι η ομάδα C_h από άποψη ομοιογένειας.

Ο συνολικός δείκτης ασυμβατότητας CI_p δίδεται από τη σχέση:

$CI_p = \max_{\forall m > h} CI_{mh}$, όπου CI_{mh} είναι ο δείκτης ασυμβατότητα μεταξύ των ομάδων

$$C_m \text{ και } C_l: CI_{mh} = \max_{\forall x_i \in C_m, \forall x_j \in C_h | \forall h, m: m > h} (\pi_{il})$$

Φυσικά, μικρή τιμή του δείκτη CI_{mh} σημαίνει ότι οι περισσότερες εναλλακτικές που ανήκουν στην καλύτερη ομάδα είναι καλύτερες μια προς μια, από τις εναλλακτικές που ανήκουν στην χειρότερη ομάδα.

Οι παραλλαγές της αντικειμενικής συνάρτησης προκύπτουν από τον διαφορετικό τρόπο υπολογισμού των δεικτών HI και CI , του HI_p και του CI_p .

Συγκεκριμένα, μια παραλλαγή προκύπτει από τον τρόπο υπολογισμού των δεικτών HI και CI . Εκτός από τη χρήση του μέγιστου, που περιγράφηκε παραπάνω, οι δείκτες HI και CI μπορούν να υπολογιστούν με τη χρήση του αθροίσματος:

$$HI_h = \sum_{\forall i, l | x_i, x_l \in C_h} (\pi_{il})$$

$$CI_{mh} = \sum_{\forall x_i \in C_m, \forall x_j \in C_h | \forall h, m: m > h} (\pi_{ij})$$

Σε αυτή την περίπτωση:

$$FI_p = HI_p + CI_p$$

Μια δεύτερη παραλλαγή της αντικειμενικής συνάρτησης μπορεί να προκύψει από τον τόπο υπολογισμού του δείκτη HI_p . Εκτός από τη χρήση του μέγιστου, που περιγράφηκε παραπάνω, μπορεί να χρησιμοποιηθεί η μέση τιμή του δείκτη HI :

$$HI_p = \frac{1}{K} \sum_{h=1}^K (HI_h)$$

Τέλος, μπορούν να προκύψουν παραλλαγές και από τον τρόπο υπολογισμού του δείκτη CI . Εκτός από τον τρόπο υπολογισμού που υποδεικνύεται παραπάνω, όπου ο δείκτης CI αντανάκλα το βαθμό στον οποίο μια εναλλακτική σε μια χειρότερη κατηγορία προτιμάται από μια εναλλακτική σε κάποια καλύτερη κατηγορία, ο δείκτης CI μπορεί να εμπεριέχει και τον βαθμό στον οποίο μια εναλλακτική κάποιας καλής κατηγορίας προτιμάται μιας εναλλακτικής σε μια κακή κατηγορία:

$$CI_{mh} = \max_{\forall x_i \in C_m, \forall x_j \in C_h | \forall h, m: m > h} (\pi_{ij}) - \max_{\forall x_i \in C_m, \forall x_j \in C_h | \forall h, m: m > h} (\pi_{ji}) \text{ ή}$$

$$CH_{mh} = \sum_{\forall x_i \in C_m, \forall x_j \in C_h} (\pi_{ij}) - \sum_{\forall x_i \in C_m, \forall x_j \in C_h} (\pi_{ji})$$

Τέλος, υπάρχει και η δυνατότητα να γίνεται χρήση δύο διαφορετικών δεικτών:

$$CI_{mh} = \max_{\forall x_i \in C_m, \forall x_j \in C_h | \forall h, m: m > h} (\pi_{ij}) \text{ και } CI_{mh}^+ = \max_{\forall x_i \in C_m, \forall x_j \in C_h | \forall h, m: m > h} (\pi_{ji}) \text{ ή}$$

$$CI_{mh} = \sum_{\forall x_i \in C_m; \forall x_l \in C_h | \forall h, m: m > h} (\pi_{ij}) \quad \text{και} \quad CI_{mh}^+ = \sum_{\forall x_i \in C_m; \forall x_l \in C_h | \forall h, m: m > h} (\pi_{ji}) .$$

Σε αυτή την περίπτωση εκτός από τους δείκτες HI_p , CI_p περιλαμβάνεται και ο δείκτης CI_{pp} :

$$CI_{pp} = \min_{\forall m > h} CI_{mh}^+, \text{ οπότε}$$

$$FI_p = \max(HI_p, CI_p) - CI_{pp} \quad \text{ή}$$

$$FI_p = HI_p + CI_p - CI_{pp} .$$

Επομένως, στην παρούσα εργασία έχουν χρησιμοποιηθεί δώδεκα διαφορετικές αντικειμενικές συναρτήσεις, οι οποίες έχουν προκύψει από τους δώδεκα συνδυασμούς των παραπάνω παραλλαγών. Όλες οι περιπτώσεις αντικειμενικών συναρτήσεων έχουν χρησιμοποιηθεί και με τους δύο τρόπους ομαδοποίησης που αναφέρθηκαν παραπάνω. Συνεπώς εξετάστηκαν συνολικά εικοσιτέσσερις διαφορετικές μεθοδολογίες πολυκριτήριας ομαδοποίησης με τη βοήθεια του αλγόριθμου της Διαφορικής Εξέλιξης.

ΚΕΦΑΛΑΙΟ 5

ΠΕΙΡΑΜΑΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ

5.1 ΔΕΔΟΜΕΝΑ

Για την εφαρμογή του εξελικτικού αλγόριθμου πολυκριτήριας ομαδοποίησης που περιγράφηκε στο κεφάλαιο 4 χρησιμοποιήθηκαν πραγματικά δεδομένα από τον τραπεζικό κλάδο. Τα δεδομένα αυτά χρησιμοποιήθηκαν από τους Mousseau et al. (2000) για την παρουσίαση μιας μεθόδου ταξινόμησης εναλλακτικών σε προκαθορισμένες κατηγορίες με τη βοήθεια της μεθόδου ELECTRE-TRI.

Τα δεδομένα απαρτίζονται από εκατό εναλλακτικές κάθε μια από τις οποίες αξιολογείται από επτά κριτήρια. Για τις ανάγκες της παρούσας εργασίας τα δεδομένα κανονικοποιήθηκαν μεταξύ 0 και 1. Στο άρθρο των Mousseau et al. (2000) οι εναλλακτικές αυτές ταξινομούνται σε 3 κατηγορίες, γι' αυτό και στην παρούσα εργασία διαχωρίζονται αρχικά σε 3 ομάδες. Επιχειρείται όμως και ο διαχωρισμός τους σε 2 και σε 4 ομάδες, προκειμένου να αξιολογηθεί η προτεινόμενη μεθοδολογία σε περιπτώσεις όπου ο αριθμός των ομάδων στην ανάλυση διαφέρει από το πραγματικό.

Εκτός από τα παραπάνω δεδομένα και τον αριθμό των ομάδων, για να εφαρμοστεί ο εξελικτικός αλγόριθμος πολυκριτήριας ομαδοποίησης, απαιτούνται κι άλλα εισερχόμενα όπως τα βάρη των κριτηρίων και οι παράμετροι που σχετίζονται με την εφαρμογή του γενικευμένου κριτηρίου στην μέθοδο PROMETHEE. Διερευνήθηκαν οι συνδυασμοί τριών περιπτώσεων βαρών με τέσσερις περιπτώσεις παραμέτρων. Στον πίνακα 5.1 φαίνονται αναλυτικά όλοι οι συνδυασμοί βαρών και παραμέτρων που χρησιμοποιήθηκαν:

Βάρη	Παράμετροι Συναρτήσεων Προτίμησης			
0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2	0.7	0.01	0.1	0.3
0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3				
1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				

Πίνακας 5.1: Συνδυασμοί βαρών και παραμέτρων

Όπως φαίνεται λοιπόν, αρχικά χρησιμοποιήθηκαν τα βάρη που χρησιμοποιήθηκαν από τους Mousseau et al. (2000), στη συνέχεια έγιναν δοκιμές με τα βάρη ROC που αντιστοιχούν στην κατάταξη των κριτηρίων βάσει της πραγματικής σημαντικότητάς τους, και τέλος με ίσα βάρη για όλα τα κριτήρια. Όσον αφορά τις παραμέτρους των συναρτήσεων προτίμησης, και στις τέσσερις περιπτώσεις είχαν την ίδια τιμή για όλα τα κριτήρια.

	HI	CI	HI_p	CI_p	FI_p
$\Delta E1$	$HI_h = \max_{\forall i,l xi, xl \in C_h} (\pi_{il})$	$CI_{mh} = \max_{\forall x_i \in C_m; \forall x_l \in C_h \forall h, m: m > h} (\pi_{il})$	$HI_p = \max_{\forall h=1, \dots, K} (HI_h)$	$CI_p = \max_{\forall m > h} CI_{mh}$	$FI_p = \max(HI_p, CI_p)$
$\Delta E2$	$HI_h = \sum_{\forall i,l x_i, x_l \in C_h} (\pi_{il})$	$CI_{mh} = \sum_{\forall x_i \in C_m; \forall x_l \in C_h \forall h, m: m > h} (\pi_{il})$	$HI_p = \max_{\forall h=1, \dots, K} (HI_h)$	$CI_p = \max_{\forall m > h} CI_{mh}$	$FI_p = HI_p + CI_p$
$\Delta E3$	$HI_h = \max_{\forall i,l xi, xl \in C_h} (\pi_{il})$	$CI_{mh} = \max_{\forall x_i \in C_m; \forall x_l \in C_h \forall h, m: m > h} (\pi_{il})$	$HI_p = \frac{1}{K} \sum_{h=1}^K (HI_h)$	$CI_p = \max_{\forall m > h} CI_{mh}$	$FI_p = \max(HI_p, CI_p)$
$\Delta E4$	$HI_h = \max_{\forall i,l xi, xl \in C_h} (\pi_{il})$	$CI_{mh} = \max_{\forall x_i \in C_m; \forall x_j \in C_h \forall h, m: m > h} (\pi_{ij}) - \max_{\forall x_i \in C_m; \forall x_j \in C_h \forall h, m: m > h} (\pi_{ji})$	$HI_p = \max_{\forall h=1, \dots, K} (HI_h)$	$CI_p = \max_{\forall m > h} CI_{mh}$	$FI_p = \max(HI_p, CI_p)$
$\Delta E5$	$HI_h = \max_{\forall i,l xi, xl \in C_h} (\pi_{il})$	$CI_{mh} = \max_{\forall x_i \in C_m; \forall x_j \in C_h \forall h, m: m > h} (\pi_{ij})$ $CI_{mh}^+ = \max_{\forall x_i \in C_m; \forall x_j \in C_h \forall h, m: m > h} (\pi_{ji})$	$HI_p = \max_{\forall h=1, \dots, K} (HI_h)$	$CI_p = \max_{\forall m > h} CI_{mh}$ $CI_{pp} = \min_{\forall m > l} CI_{ml}^+$	$FI_p = \max(HI_p, CI_p) - CI_{pp}$
$\Delta E6$	$HI_h = \sum_{\forall i,l x_i, x_l \in C_h} (\pi_{il})$	$CI_{mh} = \sum_{\forall x_i \in C_m; \forall x_l \in C_h \forall h, m: m > h} (\pi_{il})$	$HI_p = \frac{1}{K} \sum_{h=1}^K (HI_h)$	$CI_p = \max_{\forall m > h} CI_{mh}$	$FI_p = HI_p + CI_p$
$\Delta E7$	$HI_h = \sum_{\forall i,l x_i, x_l \in C_h} (\pi_{il})$	$CH_{mh} = \sum_{\forall x_i \in C_m; \forall x_j \in C_h} (\pi_{ij}) - \sum_{\forall x_i \in C_m; \forall x_j \in C_h} (\pi_{ji})$	$HI_p = \max_{\forall h=1, \dots, K} (HI_h)$	$CI_p = \max_{\forall m > h} CI_{mh}$	$FI_p = HI_p + CI_p$
$\Delta E8$	$HI_h = \sum_{\forall i,l x_i, x_l \in C_h} (\pi_{il})$	$CI_{mh} = \sum_{\forall x_i \in C_m; \forall x_l \in C_h \forall h, m: m > h} (\pi_{ij})$	$HI_p = \max_{\forall h=1, \dots, K} (HI_h)$	$CI_p = \max_{\forall m > h} CI_{mh}$	$FI_p = HI_p + CI_p - CI_{pp}$

		$CI_{mh}^+ = \sum_{\forall x_i \in C_m; \forall x_l \in C_h \forall h, m: m > h} (\pi_{ji})$		$CI_{pp} = \min_{\forall m > l} CI_{ml}^+$	
$\Delta E9$	$HI_h = \sum_{\forall i, l x_i, x_l \in C_h} (\pi_{il})$	$CH_{mh} = \sum_{\forall x_i \in C_m; \forall x_j \in C_h} (\pi_{ij}) - \sum_{\forall x_i \in C_m; \forall x_j \in C_h} (\pi_{ji})$	$HI_P = \frac{1}{K} \sum_{h=1}^K (HI_h)$	$CI_P = \max_{\forall m > h} CI_{mh}$	$FI_P = HI_P + CI_P$
$\Delta E10$	$HI_h = \sum_{\forall i, l x_i, x_l \in C_h} (\pi_{il})$	$CI_{mh} = \sum_{\forall x_i \in C_m; \forall x_l \in C_h \forall h, m: m > h} (\pi_{ij})$ $CI_{mh}^+ = \sum_{\forall x_i \in C_m; \forall x_l \in C_h \forall h, m: m > h} (\pi_{ji})$	$HI_P = \frac{1}{K} \sum_{h=1}^K (HI_h)$	$CI_P = \max_{\forall m > h} CI_{mh}$ $CI_{pp} = \min_{\forall m > l} CI_{ml}^+$	$FI_P = HI_P + CI_P - CI_{pp}$
$\Delta E11$	$HI_h = \max_{\forall i, l x_i, x_l \in C_h} (\pi_{il})$	$CI_{mh} = \max_{\forall x_i \in C_m; \forall x_j \in C_h \forall h, m: m > h} (\pi_{ij}) - \max_{\forall x_i \in C_m; \forall x_j \in C_h \forall h, m: m > h} (\pi_{ji})$	$HI_P = \frac{1}{K} \sum_{h=1}^K (HI_h)$	$CI_P = \max_{\forall m > h} CI_{mh}$	$FI_P = \max(HI_P, CI_P)$
$\Delta E12$	$HI_h = \max_{\forall i, l x_i, x_l \in C_h} (\pi_{il})$	$CI_{mh} = \max_{\forall x_i \in C_m; \forall x_j \in C_h \forall h, m: m > h} (\pi_{ij})$ $CI_{mh}^+ = \max_{\forall x_i \in C_m; \forall x_j \in C_h \forall h, m: m > h} (\pi_{ji})$	$HI_P = \frac{1}{K} \sum_{h=1}^K (HI_h)$	$CI_P = \max_{\forall m > h} CI_{mh}$ $CI_{pp} = \min_{\forall m > l} CI_{ml}^+$	$FI_P = \max(HI_P, CI_P) - CI_{pp}$

Πίνακας 5.2: Μέθοδοι υπολογισμού αντικειμενικής συνάρτησης

5.2 ΜΕΘΟΔΟΛΟΓΙΑ

Κάθε μια από τις 24 διαφοροποιήσεις του εξελικτικού αλγόριθμου πολυκριτήριας ομαδοποίησης που αναφέρθηκαν στο κεφάλαιο 4 και φαίνονται συνοπτικά στον πίνακα 5.2, εφαρμόστηκε για τρεις περιπτώσεις ομάδων (2,3 και 4) και 12 συνδυασμούς βαρών και παραμέτρων. Συνολικά δηλαδή κάθε διαφοροποίηση του αλγόριθμου δοκιμάστηκε με 36 διαφορετικές περιπτώσεις, προκειμένου να διαπιστωθεί από ποια προκύπτει η καλύτερη ομαδοποίηση. Για κάθε ένα από τα παραπάνω $24 \times 36 = 864$ σενάρια συγκεντρώθηκαν τα αποτελέσματα από 10 επαναλήψεις του αλγόριθμου.

Η σύγκριση των ομαδοποιήσεων γίνεται με έναν μη επιτηρούμενο δείκτη, την εντροπία, όπως αυτή εφαρμόζεται από τους Winters- Hilt et al. (2007). Η δομή της κάθε ομαδοποίησης συγκρίνεται με την κατάταξη σε 3 κλάσεις στην οποία έχουν καταλήξει οι Mousseau et al. (2000) με τη βοήθεια μιας μεθόδου που βασίζεται στην ELECTRE-TRI, θεωρώντας ότι αυτή είναι η «σωστή» ομαδοποίηση.

Έτσι λοιπόν η εντροπία e_i για την ομάδα C_i που διαμορφώνεται από τον αλγόριθμο ομαδοποίησης μπορεί να γραφτεί ως εξής:

$$e_i = - \sum_{h=1}^3 p_{ih} \log p_{ih}$$

Όπου p_{ih} η πιθανότητα ένα αντικείμενο της ομάδας i να ανήκει στην πραγματική κατηγορία h , και μπορεί να υπολογιστεί αθροίζοντας τον αριθμό των αντικειμένων της κατηγορίας h που έχουν καταταγεί στην ομάδα i , n_{ih} , και διαιρώντας τον με το συνολικό αριθμό των αντικειμένων της ομάδα i , n_i .

Η συνολική εντροπία μπορεί να υπολογιστεί από το άθροισμα των επιμέρους εντροπιών σταθμισμένο από το μέγεθος κάθε ομάδας:

$$e = \sum_{i=1}^K \frac{m_i}{m} e_i ,$$

όπου K είναι ο αριθμός των ομάδων που διαμορφώνονται από τον αλγόριθμο ομαδοποίησης και m ο συνολικός αριθμός των αντικειμένων. Φυσικά όσο μικρότερη η εντροπία, τόσο καλύτερη η ομαδοποίηση. Αφού υπολογίζονται οι εντροπίες σε κάθε μία από τις 10 επαναλήψεις του αλγόριθμου υπολογίζεται ο μέσος όρος, η καλύτερη και η χειρότερη εντροπία, καθώς και η τυπική απόκλιση τους.

Τέλος, εφαρμόζεται ο αλγόριθμος της Διαφορικής Εξέλιξης για κλασσική, όχι πολυκριτήρια, ομαδοποίηση με βάση τις αποστάσεις των εναλλακτικών από τα σημεία αναφοράς κάθε ομάδας.

Εκτός από την εντροπία, η αξιολόγηση των αποτελεσμάτων γίνεται και μη επιτηρούμενους δείκτες οι οποίοι βασίζονται στις ροές μεταξύ των εναλλακτικών των ομάδων που διαμορφώθηκαν από κάθε εφαρμογή του αλγόριθμου. Συγκεκριμένα υπολογίζονται οι δείκτες:

- $HI_p_max = \max_{\forall h=1,...,K} (HI_h)$, όπου

$$HI_h = \max_{\forall i,l | x_i, x_l \in C_h} (\pi_{il}) \text{ ο δείκτης ετερογένειας κάθε ομάδας}$$

- $CI_p_max = \max_{\forall m>h} CI_{mh}$, όπου CI_{mh} είναι ο δείκτης ασυμβατότητας μεταξύ των

$$\text{ομάδων } C_m \text{ και } C_l: CI_{mh} = \max_{\forall x_i \in C_m; \forall x_l \in C_h | \forall h,m:m>h} (\pi_{il})$$

- $HI_p_mean = \max_{\forall h=1,...,K} (HI_h_mean)$, όπου ο δείκτης

$$HI_h_mean = \frac{1}{n_h * n_h} \sum_{\forall i,l | x_i, x_l \in C_h}^{n_h} (\pi_{il}), \text{ είναι ο μέσος βαθμός προτίμησης μεταξύ των } n_h$$

εναλλακτικών που ανήκουν στην ομάδα C_h .

- $CI_p_mean = \frac{1}{K * K} \sum_{\forall m>h} CI_{mh_mean}$, όπου CI_{mh} είναι ο δείκτης ασυμβατότητας μεταξύ των ομάδων C_m και C_l :

$$CI_{mh_mean} = \frac{1}{n_m * n_h} \sum_{\forall x_i \in C_m; \forall x_l \in C_h | \forall h,m:m>h} (\pi_{il}), \text{ αντιστοιχεί στο μέσο βαθμό προτίμησης}$$

των εναλλακτικών μιας χειρότερης ομάδας σε σχέση με τις εναλλακτικές μιας καλύτερης ομάδας

Σε όλες τις παραπάνω περιπτώσεις δεικτών, όσο μικρότερη η τιμή του δείκτη, τόσο καλύτερη η ομαδοποίηση.

5.3 ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

5.3.1 Αποτελέσματα Διαχωρισμού σε 2 Ομάδες

Η παρουσίαση των αποτελεσμάτων ξεκινάει από το διαχωρισμό των εναλλακτικών σε δύο ομάδες με τη βοήθεια των ροών της PROMETHEE. Από τις τιμές της εντροπίας, που παρουσιάζονται στον πίνακα 5.3, γίνεται εμφανές ότι οι μέθοδοι με τις κατά μέσο όρο καλύτερες επιδόσεις είναι οι ΔΕ1, ΔΕ4 και ΔΕ5. Σε ανάλογο συμπέρασμα καταλήγει κανείς κοιτώντας τις καλύτερες και τις χειρότερες τιμές της εντροπίας. Ακόμα παρατηρείται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τα πραγματικά βάρη με $\sigma=0,7$ και με τα βάρη ROC και $\sigma=0,7$. Και στις τρεις περιπτώσεις βαρών οι μικρότερες τιμές εντροπίας εμφανίζονται όταν η παράμετρος των συναρτήσεων προτίμησης είναι 0,7 και 0,3.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.86	0.81	0.74	0.83	0.78	0.76	0.70	0.71	0.81	0.79	0.70	0.67
ΔΕ2	0.86	0.83	0.74	0.71	0.87	0.87	0.84	0.71	0.88	0.87	0.86	0.79
ΔΕ3	0.99	0.97	0.77	0.80	0.96	0.96	0.77	0.74	0.96	0.93	0.87	0.81
ΔΕ4	0.84	0.81	0.75	0.82	0.82	0.77	0.70	0.71	0.80	0.79	0.72	0.66
ΔΕ5	0.87	0.81	0.73	0.83	0.79	0.76	0.70	0.72	0.79	0.79	0.71	0.66
ΔΕ6	0.84	0.82	0.78	0.76	0.83	0.82	0.79	0.78	0.85	0.83	0.79	0.79
ΔΕ7	0.87	0.87	0.85	0.88	0.82	0.87	0.83	0.83	0.70	0.86	0.82	0.82
ΔΕ8	0.86	0.83	0.75	0.73	0.87	0.87	0.84	0.71	0.88	0.87	0.86	0.80
ΔΕ9	0.88	0.85	0.88	0.87	0.65	0.86	0.82	0.86	0.36	0.86	0.82	0.81
ΔΕ10	0.84	0.82	0.78	0.75	0.83	0.82	0.79	0.79	0.85	0.83	0.79	0.79
ΔΕ11	0.99	0.91	0.77	0.80	0.96	0.95	0.75	0.74	0.95	0.94	0.86	0.81
ΔΕ12	0.99	0.91	0.79	0.77	0.96	0.95	0.74	0.74	0.94	0.94	0.88	0.81

Πίνακας 5.3: ΤΙΜΕΣ ΕΝΤΡΟΠΙΑΣ ΔΙΑΧΩΡΙΣΜΟΥ ΒΑΣΕΙ ΤΩΝ ΡΟΩΝ ΤΗΣ PROMETHEE ΣΕ 2 ΟΜΑΔΕΣ

Στους πίνακες 5.4, 5.5, 5.6 και 5.7 παρουσιάζονται αντίστοιχα οι δείκτες H_{ip_max} , H_{ip_mean} , Cl_{ip_max} και Cl_{ip_mean} . Με βάση και τους τέσσερις δείκτες επιβεβαιώνεται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τη χρήση των πραγματικών βαρών και βαρών ROC και $\sigma=0,7$, και ακολουθούν τα ίσα βάρη πάντα με $\sigma=0,7$. Κατά μέσο όρο οι καλύτεροι δείκτες H_{ip_max} επιτυγχάνονται από τις μεθόδους ΔΕ2, ΔΕ3 και ΔΕ1.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	1.00	0.84	0.52	0.22	1.00	0.85	0.50	0.21	1.00	0.95	0.60	0.22
ΔΕ2	1.00	0.84	0.52	0.22	1.00	0.85	0.50	0.21	1.00	0.94	0.60	0.22
ΔΕ3	1.00	0.84	0.52	0.22	1.00	0.85	0.50	0.21	1.00	0.94	0.60	0.22
ΔΕ4	1.00	0.99	0.58	0.25	1.00	0.99	0.64	0.23	1.00	1.00	0.85	0.29
ΔΕ5	1.00	0.98	0.55	0.25	1.00	0.98	0.63	0.23	1.00	1.00	0.83	0.29
ΔΕ6	1.00	0.94	0.56	0.26	1.00	0.98	0.59	0.23	1.00	1.00	0.87	0.29
ΔΕ7	1.00	0.85	0.52	0.25	1.00	0.94	0.68	0.27	1.00	0.98	0.80	0.35
ΔΕ8	1.00	0.95	0.84	0.40	1.00	0.95	0.80	0.39	1.00	0.98	0.88	0.42
ΔΕ9	1.00	0.84	0.52	0.25	1.00	0.94	0.68	0.27	1.00	0.98	0.80	0.35
ΔΕ10	1.00	0.85	0.52	0.25	1.00	0.94	0.59	0.22	1.00	0.97	0.76	0.29
ΔΕ11	1.00	0.97	0.87	0.40	1.00	0.95	0.80	0.40	1.00	0.98	0.88	0.42
ΔΕ12	1.00	0.85	0.52	0.25	1.00	0.94	0.58	0.22	1.00	0.97	0.76	0.29

Πίνακας 5.4: ΔΕΙΚΤΗΣ H_{ip_max} ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ PROMETHEE

Αντίστοιχα, οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη H_{ip_mean} είναι οι ΔΕ5, ΔΕ6 και ΔΕ4.

Βάρη σ	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.38	0.27	0.13	0.04	0.39	0.28	0.12	0.04	0.39	0.29	0.12	0.04
ΔΕ2	0.38	0.27	0.13	0.04	0.39	0.28	0.12	0.04	0.40	0.30	0.12	0.04
ΔΕ3	0.38	0.27	0.13	0.04	0.39	0.28	0.12	0.04	0.40	0.30	0.12	0.04
ΔΕ4	0.35	0.27	0.11	0.04	0.32	0.24	0.11	0.03	0.30	0.19	0.09	0.03
ΔΕ5	0.32	0.25	0.12	0.04	0.30	0.25	0.10	0.03	0.29	0.18	0.09	0.03
ΔΕ6	0.34	0.26	0.12	0.04	0.31	0.26	0.11	0.03	0.29	0.19	0.09	0.03
ΔΕ7	0.38	0.28	0.13	0.04	0.39	0.28	0.14	0.04	0.41	0.28	0.14	0.05
ΔΕ8	0.37	0.27	0.13	0.04	0.39	0.26	0.13	0.04	0.40	0.26	0.13	0.04
ΔΕ9	0.38	0.28	0.13	0.04	0.39	0.28	0.14	0.04	0.41	0.28	0.14	0.05
ΔΕ10	0.37	0.28	0.13	0.04	0.39	0.28	0.13	0.04	0.41	0.29	0.14	0.04
ΔΕ11	0.37	0.27	0.13	0.04	0.38	0.26	0.13	0.04	0.39	0.26	0.13	0.04
ΔΕ12	0.37	0.28	0.13	0.04	0.39	0.28	0.13	0.04	0.41	0.29	0.14	0.04

Πίνακας 5.5: ΔΕΙΚΤΗΣ HIp_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ PROMETHEE

Οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη CIp_max είναι οι ΔΕ6, ΔΕ4 και ΔΕ5.

Βάρη σ	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.64	0.45	0.26	0.12	0.61	0.45	0.21	0.09	0.73	0.61	0.18	0.05
ΔΕ2	0.65	0.46	0.26	0.12	0.71	0.47	0.21	0.09	0.72	0.64	0.17	0.05
ΔΕ3	0.65	0.47	0.27	0.12	0.62	0.47	0.21	0.09	0.67	0.66	0.18	0.05
ΔΕ4	0.69	0.62	0.21	0.09	0.61	0.53	0.17	0.06	0.55	0.37	0.06	0.03
ΔΕ5	0.75	0.52	0.23	0.09	0.61	0.52	0.15	0.06	0.54	0.46	0.07	0.03
ΔΕ6	0.72	0.45	0.23	0.08	0.61	0.51	0.18	0.06	0.50	0.42	0.06	0.03
ΔΕ7	0.57	0.43	0.27	0.11	0.65	0.50	0.32	0.13	0.67	0.60	0.36	0.14
ΔΕ8	0.70	0.67	0.62	0.27	0.68	0.54	0.41	0.20	0.86	0.45	0.38	0.22
ΔΕ9	0.57	0.42	0.27	0.11	0.66	0.51	0.32	0.12	0.67	0.60	0.37	0.14
ΔΕ10	0.57	0.42	0.24	0.08	0.60	0.52	0.28	0.09	0.67	0.61	0.32	0.14
ΔΕ11	0.86	0.76	0.67	0.27	0.91	0.54	0.39	0.20	0.82	0.45	0.38	0.22
ΔΕ12	0.57	0.42	0.24	0.08	0.60	0.52	0.28	0.09	0.67	0.59	0.32	0.14

Πίνακας 5.6: ΔΕΙΚΤΗΣ CIp_max ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ PROMETHEE

Τέλος, κατά μέσο όρο οι καλύτεροι δείκτες CIp_mean επιτυγχάνονται από τις μεθόδους ΔΕ10 και ΔΕ12.

Βάρη	1/7,1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.16	0.10	0.04	0.01	0.13	0.07	0.02	0.01	0.16	0.07	0.01	0.00
ΔΕ2	0.16	0.09	0.04	0.01	0.16	0.08	0.02	0.01	0.23	0.09	0.01	0.00
ΔΕ3	0.17	0.10	0.04	0.01	0.13	0.07	0.02	0.01	0.14	0.08	0.01	0.00
ΔΕ4	0.23	0.15	0.03	0.01	0.21	0.15	0.02	0.00	0.21	0.14	0.01	0.00
ΔΕ5	0.27	0.10	0.03	0.01	0.23	0.14	0.01	0.00	0.20	0.18	0.01	0.00
ΔΕ6	0.23	0.09	0.03	0.01	0.22	0.13	0.02	0.00	0.20	0.16	0.01	0.00
ΔΕ7	0.15	0.11	0.04	0.01	0.15	0.11	0.05	0.01	0.13	0.10	0.04	0.01
ΔΕ8	0.19	0.16	0.14	0.06	0.18	0.14	0.10	0.04	0.23	0.13	0.10	0.03
ΔΕ9	0.15	0.11	0.04	0.01	0.16	0.11	0.04	0.01	0.13	0.10	0.04	0.01
ΔΕ10	0.14	0.09	0.04	0.01	0.13	0.09	0.03	0.01	0.12	0.08	0.03	0.01
ΔΕ11	0.21	0.18	0.16	0.06	0.27	0.14	0.09	0.04	0.29	0.14	0.10	0.03
ΔΕ12	0.14	0.09	0.04	0.01	0.13	0.09	0.03	0.01	0.12	0.08	0.03	0.01

Πίνακας 5.7: ΔΕΙΚΤΗΣ Clp_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ PROMETHEE

Στη συνέχεια παρουσιάζονται τα αποτελέσματα που αφορούν το διαχωρισμό των εναλλακτικών σε δύο ομάδες με τη βοήθεια των αποστάσεων από τα σημεία αναφοράς κάθε ομάδας. Στον πίνακα 5.8 παρουσιάζονται οι τιμές της εντροπίας. Παρατηρείται ότι κατά μέσο όρο οι καλύτερες τιμές τις εντροπίας επιτυγχάνονται με την εφαρμογή των μεθόδων ΔΕ5, ΔΕ1 και ΔΕ4. Ακόμα παρατηρείται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τα πραγματικά βάρη και τα βάρη ROC με $\sigma=0,7$.

Βάρη	1/7,1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.79	0.78	0.82	0.85	0.80	0.76	0.71	0.75	0.83	0.77	0.68	0.67
ΔΕ2	0.85	0.85	0.79	0.78	0.86	0.87	0.84	0.70	0.88	0.87	0.86	0.83
ΔΕ3	0.96	0.96	0.78	0.71	0.99	0.95	0.77	0.74	0.97	0.96	0.91	0.83
ΔΕ4	0.82	0.79	0.81	0.85	0.81	0.77	0.71	0.74	0.77	0.76	0.67	0.67
ΔΕ5	0.81	0.80	0.83	0.85	0.77	0.76	0.71	0.74	0.83	0.76	0.69	0.67
ΔΕ6	0.83	0.82	0.78	0.76	0.83	0.81	0.79	0.79	0.85	0.83	0.79	0.79
ΔΕ7	0.82	0.88	0.90	0.83	0.82	0.86	0.86	0.89	0.62	0.86	0.82	0.86
ΔΕ8	0.85	0.85	0.80	0.78	0.86	0.87	0.85	0.71	0.88	0.87	0.86	0.83
ΔΕ9	0.81	0.85	0.88	0.85	0.62	0.86	0.84	0.88	0.61	0.86	0.82	0.82
ΔΕ10	0.83	0.82	0.78	0.76	0.83	0.81	0.79	0.79	0.85	0.83	0.79	0.79
ΔΕ11	0.97	0.96	0.78	0.74	0.98	0.95	0.77	0.74	0.98	0.96	0.91	0.81
ΔΕ12	0.98	0.95	0.79	0.71	0.99	0.94	0.77	0.74	0.97	0.96	0.90	0.81

Πίνακας 5.8: ΤΙΜΕΣ ΕΝΤΡΟΠΙΑΣ ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΤΗΣ ΑΠΟΣΤΑΣΗΣ

Στους πίνακες 5.9, 5.10, 5.11 και 5.12 παρουσιάζονται αντίστοιχα οι δείκτες HIp_max, HIp_mean, Clp_max και Clp_mean. Με βάση και τους τέσσερις δείκτες επιβεβαιώνεται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τη χρήση των πραγματικών βαρών και βαρών ROC και $\sigma=0,7$, και ακολουθούν τα ίσα βάρη πάντα με $\sigma=0,7$. Κατά μέσο όρο οι καλύτεροι δείκτες HIp_max επιτυγχάνονται από τις μεθόδους ΔΕ1 ΔΕ2 και ΔΕ3.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	1.00	0.84	0.52	0.22	1.00	0.85	0.50	0.21	1.00	0.93	0.60	0.22
ΔΕ2	1.00	0.84	0.52	0.22	1.00	0.85	0.50	0.21	1.00	0.93	0.60	0.22
ΔΕ3	1.00	0.84	0.52	0.22	1.00	0.85	0.50	0.21	1.00	0.93	0.60	0.22
ΔΕ4	1.00	0.95	0.68	0.25	1.00	0.96	0.70	0.23	1.00	0.99	0.89	0.31
ΔΕ5	1.00	0.95	0.65	0.25	1.00	0.96	0.68	0.23	1.00	0.99	0.87	0.29
ΔΕ6	1.00	0.96	0.71	0.25	1.00	0.95	0.70	0.23	1.00	0.99	0.87	0.29
ΔΕ7	1.00	0.84	0.52	0.25	1.00	0.94	0.67	0.27	1.00	0.98	0.81	0.35
ΔΕ8	1.00	0.97	0.86	0.41	1.00	0.95	0.86	0.40	1.00	0.98	0.89	0.51
ΔΕ9	1.00	0.84	0.52	0.25	1.00	0.94	0.68	0.27	1.00	0.98	0.80	0.35
ΔΕ10	1.00	0.85	0.52	0.25	1.00	0.89	0.59	0.22	1.00	0.97	0.76	0.29
ΔΕ11	1.00	0.94	0.87	0.44	1.00	0.95	0.82	0.41	1.00	0.98	0.88	0.42
ΔΕ12	1.00	0.85	0.52	0.25	1.00	0.89	0.59	0.22	1.00	0.97	0.76	0.29

Πίνακας 5.9: ΔΕΙΚΤΗΣ H_{pr_max} ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Αντίστοιχα, οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη H_{pr_mean} είναι οι ΔΕ6 ΔΕ4 και ΔΕ5.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.37	0.27	0.13	0.04	0.39	0.28	0.12	0.04	0.41	0.29	0.12	0.04
ΔΕ2	0.38	0.27	0.13	0.04	0.39	0.28	0.12	0.04	0.40	0.29	0.12	0.04
ΔΕ3	0.38	0.27	0.13	0.04	0.38	0.28	0.12	0.04	0.40	0.29	0.12	0.04
ΔΕ4	0.35	0.28	0.10	0.03	0.36	0.27	0.09	0.03	0.35	0.24	0.10	0.03
ΔΕ5	0.35	0.28	0.10	0.04	0.34	0.27	0.09	0.03	0.35	0.25	0.10	0.03
ΔΕ6	0.36	0.26	0.09	0.03	0.36	0.28	0.09	0.03	0.36	0.24	0.10	0.03
ΔΕ7	0.38	0.28	0.13	0.04	0.39	0.28	0.14	0.04	0.41	0.28	0.15	0.05
ΔΕ8	0.38	0.27	0.13	0.05	0.39	0.27	0.13	0.04	0.40	0.27	0.13	0.05
ΔΕ9	0.38	0.28	0.14	0.04	0.39	0.28	0.14	0.04	0.41	0.28	0.15	0.05
ΔΕ10	0.37	0.28	0.13	0.04	0.39	0.28	0.13	0.04	0.41	0.29	0.14	0.04
ΔΕ11	0.37	0.27	0.13	0.05	0.38	0.26	0.13	0.04	0.39	0.26	0.13	0.05
ΔΕ12	0.37	0.28	0.13	0.04	0.39	0.28	0.13	0.04	0.41	0.29	0.14	0.04

Πίνακας 5.10: ΔΕΙΚΤΗΣ H_{pr_mean} ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη C_{Ip_max} είναι οι ΔΕ4 και ΔΕ6.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.71	0.43	0.26	0.12	0.69	0.44	0.21	0.09	0.91	0.65	0.20	0.06
ΔΕ2	0.64	0.44	0.26	0.12	0.72	0.45	0.21	0.09	0.87	0.61	0.20	0.06
ΔΕ3	0.66	0.44	0.25	0.12	0.62	0.44	0.21	0.09	0.91	0.59	0.20	0.06
ΔΕ4	0.62	0.50	0.15	0.08	0.66	0.48	0.12	0.06	0.74	0.41	0.08	0.03
ΔΕ5	0.73	0.52	0.15	0.08	0.73	0.47	0.12	0.06	0.83	0.41	0.09	0.03
ΔΕ6	0.73	0.46	0.12	0.08	0.68	0.52	0.12	0.06	0.78	0.41	0.08	0.03
ΔΕ7	0.57	0.44	0.26	0.11	0.60	0.50	0.32	0.13	0.73	0.60	0.37	0.14
ΔΕ8	0.75	0.71	0.69	0.28	0.77	0.54	0.52	0.21	0.89	0.45	0.40	0.24
ΔΕ9	0.57	0.44	0.26	0.11	0.60	0.51	0.32	0.13	0.69	0.60	0.37	0.14
ΔΕ10	0.57	0.42	0.24	0.08	0.60	0.48	0.28	0.09	0.69	0.62	0.32	0.14
ΔΕ11	0.80	0.69	0.69	0.30	0.88	0.54	0.46	0.21	0.95	0.45	0.38	0.22
ΔΕ12	0.57	0.42	0.24	0.08	0.60	0.48	0.28	0.09	0.69	0.62	0.32	0.14

Πίνακας 5.11: ΔΕΙΚΤΗΣ CIp_max ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Τέλος, κατά μέσο όρο οι καλύτεροι δείκτες CIp_mean επιτυγχάνονται από τις μεθόδους ΔΕ3, ΔΕ1 και ΔΕ2.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.14	0.09	0.04	0.01	0.14	0.07	0.02	0.01	0.21	0.07	0.01	0.00
ΔΕ2	0.16	0.09	0.05	0.01	0.17	0.07	0.02	0.01	0.10	0.07	0.01	0.00
ΔΕ3	0.15	0.09	0.04	0.01	0.12	0.07	0.02	0.01	0.21	0.07	0.01	0.00
ΔΕ4	0.15	0.10	0.02	0.01	0.19	0.11	0.01	0.00	0.35	0.08	0.01	0.00
ΔΕ5	0.23	0.11	0.02	0.01	0.23	0.10	0.01	0.00	0.21	0.08	0.01	0.00
ΔΕ6	0.21	0.10	0.01	0.01	0.19	0.12	0.01	0.00	0.30	0.08	0.01	0.00
ΔΕ7	0.15	0.10	0.04	0.01	0.15	0.11	0.05	0.01	0.13	0.10	0.04	0.01
ΔΕ8	0.24	0.16	0.17	0.06	0.20	0.14	0.11	0.04	0.25	0.13	0.10	0.04
ΔΕ9	0.15	0.10	0.04	0.01	0.15	0.11	0.05	0.01	0.13	0.10	0.04	0.01
ΔΕ10	0.14	0.09	0.04	0.01	0.13	0.09	0.03	0.01	0.12	0.08	0.03	0.01
ΔΕ11	0.22	0.17	0.18	0.07	0.27	0.14	0.10	0.05	0.25	0.14	0.10	0.03
ΔΕ12	0.14	0.09	0.04	0.01	0.13	0.09	0.03	0.01	0.12	0.08	0.03	0.01

Πίνακας 5.12: ΔΕΙΚΤΗΣ CIp_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 2 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Συνολικά δηλαδή προκύπτει ότι στον πολυκριτήριο διαχωρισμό σε 2 ομάδες με βάση τις αποστάσεις των εναλλακτικών από τα σημεία αναφοράς κάθε ομάδας ότι οι μέθοδοι ΔΕ1, ΔΕ2, ΔΕ4 και ΔΕ6 έχουν την καλύτερη απόδοση.

Ενδιαφέρον παρουσιάζει η σύγκριση της εντροπίας αλλά και των δεικτών HIp_max, HIp_mean, CIp_max και CIp_mean που προκύπτουν από τον διαχωρισμό σε δύο ομάδες, όχι με πολυκριτήρια μέθοδο, αλλά με μια μέθοδο κλασσικής ομαδοποίησης που βασίζεται στις αποστάσεις. Η μέση τιμή της εντροπίας σε αυτή την περίπτωση είναι 0,83, είναι δηλαδή στα ίδια επίπεδα με τις τιμές της εντροπίας της πολυκριτήριας ομαδοποίησης. Στον πίνακα 5.13 παρουσιάζονται οι τιμές των δεικτών HIp_max, HIp_mean, CIp_max και CIp_mean, υπολογισμένοι με διαφορετικά βάρη και σ κάθε φορά. Είναι εμφανές ότι και σε αυτή την περίπτωση σημαντικά καλύτερες τιμές στους δείκτες προκύπτουν με ίσα και πραγματικά βάρη και $\sigma=0.7$.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
Hip_max	1.00	0.99	0.65	0.31	1.00	0.99	0.71	0.36	1.00	1.00	0.89	0.52
Hip_mean	0.38	0.29	0.14	0.04	0.39	0.30	0.15	0.05	0.41	0.32	0.17	0.06
CI_max	0.71	0.66	0.43	0.18	0.80	0.76	0.54	0.25	0.93	0.92	0.79	0.37
CI_mean	0.25	0.19	0.09	0.03	0.31	0.23	0.11	0.03	0.38	0.29	0.14	0.04

Πίνακας 5.13: ΔΕΙΚΤΕΣ HI ΚΑΙ CI ΚΛΑΣΣΙΚΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ ΣΕ 2 ΟΜΑΔΕΣ

5.3.2 Αποτελέσματα Διαχωρισμού σε 3 Ομάδες

Στη συνέχεια παρουσιάζονται τα αποτελέσματα που αφορούν το διαχωρισμό των εναλλακτικών σε τρεις ομάδες με τη βοήθεια των ροών της PROMETHEE. Στον πίνακα 5.14 παρουσιάζονται οι τιμές της εντροπίας. Παρατηρείται ότι κατά μέσο όρο οι καλύτερες τιμές της εντροπίας επιτυγχάνονται με την εφαρμογή των μεθόδων ΔΕ11, και ΔΕ8. Ακόμα παρατηρείται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με ίσα βάρη και βάρη ROC με 0.7 και τα πραγματικά βάρη με $\sigma=0,3$ και 0,7.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.68	0.68	0.65	0.63	0.64	0.64	0.60	0.60	0.54	0.66	0.75	0.58
ΔΕ2	0.75	0.67	0.64	0.13	0.62	0.63	0.61	0.62	0.52	0.67	0.75	0.59
ΔΕ3	0.70	0.67	0.63	0.64	0.62	0.64	0.62	0.62	0.49	0.60	0.75	0.48
ΔΕ4	0.81	0.69	0.71	0.09	0.91	0.79	0.67	0.69	0.91	0.85	0.75	0.73
ΔΕ5	0.83	0.67	0.72	0.79	0.90	0.82	0.60	0.69	0.90	0.85	0.74	0.75
ΔΕ6	0.81	0.79	0.73	0.79	0.83	0.87	0.68	0.70	0.92	0.87	0.73	0.78
ΔΕ7	0.75	0.72	0.69	0.65	0.81	0.80	0.61	0.62	0.56	0.78	0.57	0.55
ΔΕ8	0.71	0.65	0.67	0.62	0.60	0.62	0.59	0.58	0.53	0.59	0.54	0.54
ΔΕ9	0.76	0.74	0.71	0.69	0.81	0.76	0.60	0.60	0.83	0.77	0.54	0.56
ΔΕ10	0.76	0.71	0.63	0.65	0.80	0.76	0.61	0.55	0.79	0.76	0.53	0.50
ΔΕ11	0.59	0.66	0.60	0.59	0.48	0.50	0.54	0.54	0.39	0.42	0.50	0.48
ΔΕ12	0.74	0.68	0.63	0.64	0.80	0.76	0.60	0.54	0.79	0.76	0.60	0.51

Πίνακας 5.14: ΤΙΜΕΣ ΕΝΤΡΟΠΙΑΣ ΔΙΑΧΩΡΙΣΜΟΥ ΒΑΣΕΙ ΤΩΝ ΡΟΩΝ ΤΗΣ PROMETHEE ΣΕ 3 ΟΜΑΔΕΣ

Στους πίνακες 5.15, 5.16, 5.17 και 5.18 παρουσιάζονται αντίστοιχα οι δείκτες Hip_max, Hip_mean, Cip_max και Cip_mean. Με βάση και τους τέσσερις δείκτες επιβεβαιώνεται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τη χρήση ίσων βαρών, των πραγματικών και βαρών ROC και $\sigma=0,7$. Κατά μέσο όρο οι καλύτεροι δείκτες Hip_max επιτυγχάνονται από τις μεθόδους ΔΕ2, ΔΕ3 και ΔΕ1.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.86	0.69	0.41	0.15	0.90	0.75	0.39	0.17	1.00	0.87	0.40	0.16
ΔΕ2	0.86	0.69	0.41	0.15	0.90	0.75	0.39	0.17	0.97	0.87	0.40	0.15
ΔΕ3	0.86	0.69	0.41	0.15	0.90	0.75	0.39	0.17	0.97	0.87	0.40	0.17
ΔΕ4	0.94	0.83	0.46	0.23	0.99	0.94	0.44	0.22	1.00	0.99	0.42	0.25
ΔΕ5	0.97	0.82	0.48	0.23	1.00	0.93	0.44	0.22	1.00	1.00	0.48	0.25
ΔΕ6	0.95	0.85	0.48	0.23	1.00	0.94	0.42	0.22	1.00	0.99	0.43	0.26
ΔΕ7	0.99	0.89	0.58	0.22	1.00	0.94	0.67	0.25	1.00	0.98	0.79	0.31
ΔΕ8	1.00	0.91	0.62	0.24	1.00	0.92	0.66	0.24	1.00	0.98	0.69	0.26
ΔΕ9	1.00	0.89	0.56	0.23	1.00	0.94	0.68	0.25	1.00	0.98	0.78	0.32
ΔΕ10	1.00	0.91	0.56	0.22	1.00	0.94	0.68	0.25	1.00	0.98	0.78	0.31
ΔΕ11	1.00	0.99	0.56	0.25	1.00	0.93	0.57	0.26	1.00	0.98	0.66	0.21
ΔΕ12	1.00	0.90	0.56	0.23	1.00	0.94	0.67	0.25	1.00	0.98	0.78	0.31

Πίνακας 5.15: ΔΕΙΚΤΗΣ HIp_max ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 3 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ
PROMETHEE

Αντίστοιχα, οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη HIp_mean είναι οι ΔΕ5, ΔΕ4 και ΔΕ6.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.36	0.27	0.12	0.04	0.36	0.24	0.11	0.03	0.36	0.24	0.09	0.02
ΔΕ2	0.36	0.27	0.12	0.04	0.36	0.24	0.11	0.03	0.36	0.24	0.09	0.02
ΔΕ3	0.36	0.27	0.12	0.04	0.37	0.24	0.11	0.03	0.36	0.24	0.09	0.02
ΔΕ4	0.32	0.23	0.10	0.03	0.30	0.22	0.10	0.03	0.26	0.17	0.09	0.02
ΔΕ5	0.33	0.23	0.10	0.03	0.29	0.22	0.10	0.03	0.27	0.15	0.09	0.02
ΔΕ6	0.33	0.23	0.10	0.03	0.32	0.23	0.09	0.03	0.26	0.17	0.09	0.02
ΔΕ7	0.36	0.26	0.12	0.04	0.38	0.26	0.12	0.04	0.38	0.25	0.11	0.03
ΔΕ8	0.36	0.25	0.11	0.03	0.37	0.25	0.11	0.03	0.37	0.25	0.10	0.03
ΔΕ9	0.36	0.26	0.12	0.04	0.38	0.26	0.12	0.04	0.39	0.25	0.11	0.03
ΔΕ10	0.35	0.25	0.11	0.03	0.37	0.25	0.11	0.03	0.39	0.25	0.11	0.03
ΔΕ11	0.35	0.25	0.11	0.03	0.36	0.25	0.11	0.03	0.36	0.23	0.10	0.03
ΔΕ12	0.35	0.25	0.11	0.04	0.37	0.25	0.11	0.03	0.39	0.25	0.11	0.03

Πίνακας 5.16: ΔΕΙΚΤΗΣ HIp_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 3 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ
PROMETHEE

Οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη CIp_max είναι οι ΔΕ4, ΔΕ5 και ΔΕ6.

Βάρη	1/7,1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.77	0.72	0.28	0.10	0.88	0.61	0.62	0.14	1.00	0.67	0.72	0.28
ΔΕ2	0.90	0.79	0.28	0.10	0.90	0.56	0.62	0.16	0.97	0.74	0.72	0.26
ΔΕ3	0.88	0.68	0.28	0.10	0.91	0.61	0.63	0.15	0.99	0.80	0.72	0.29
ΔΕ4	0.66	0.53	0.29	0.10	0.64	0.62	0.34	0.22	0.57	0.48	0.68	0.15
ΔΕ5	0.77	0.60	0.31	0.10	0.67	0.68	0.33	0.14	0.64	0.52	0.43	0.17
ΔΕ6	0.72	0.69	0.32	0.11	0.69	0.64	0.32	0.27	0.56	0.49	0.68	0.20
ΔΕ7	0.77	0.87	0.52	0.21	0.76	0.84	0.50	0.21	1.00	0.74	0.51	0.24
ΔΕ8	0.92	0.82	0.56	0.22	0.94	0.80	0.52	0.19	0.93	0.94	0.42	0.22
ΔΕ9	0.73	0.88	0.49	0.23	0.83	0.80	0.49	0.20	1.00	0.73	0.50	0.24
ΔΕ10	0.70	0.84	0.58	0.23	0.98	0.72	0.48	0.20	1.00	0.43	0.51	0.24
ΔΕ11	1.00	0.84	0.57	0.24	1.00	0.84	0.46	0.20	1.00	0.94	0.42	0.22
ΔΕ12	0.74	0.84	0.53	0.19	0.98	0.62	0.45	0.21	1.00	0.43	0.51	0.24

Πίνακας 5.17: ΔΕΙΚΤΗΣ C_{ip}_max ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 3 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ
PROMETHEE

Τέλος, κατά μέσο όρο οι καλύτεροι δείκτες C_{ip}_mean επιτυγχάνονται από τις μεθόδους ΔΕ5, ΔΕ4και ΔΕ6.

Βάρη	1/7,1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.32	0.31	0.07	0.02	0.44	0.24	0.25	0.03	0.54	0.31	0.30	0.11
ΔΕ2	0.39	0.35	0.07	0.02	0.47	0.21	0.25	0.04	0.50	0.35	0.30	0.11
ΔΕ3	0.43	0.26	0.07	0.02	0.47	0.24	0.25	0.04	0.56	0.41	0.29	0.10
ΔΕ4	0.22	0.16	0.07	0.02	0.32	0.32	0.08	0.06	0.25	0.34	0.27	0.05
ΔΕ5	0.32	0.21	0.06	0.02	0.28	0.34	0.06	0.03	0.23	0.35	0.13	0.07
ΔΕ6	0.31	0.28	0.08	0.02	0.31	0.40	0.06	0.08	0.27	0.37	0.26	0.08
ΔΕ7	0.23	0.43	0.18	0.08	0.34	0.40	0.21	0.09	0.57	0.33	0.24	0.09
ΔΕ8	0.41	0.42	0.21	0.09	0.40	0.38	0.23	0.09	0.53	0.45	0.24	0.09
ΔΕ9	0.23	0.46	0.16	0.08	0.42	0.37	0.20	0.08	0.67	0.33	0.25	0.09
ΔΕ10	0.23	0.43	0.24	0.09	0.55	0.31	0.19	0.08	0.68	0.12	0.23	0.09
ΔΕ11	0.53	0.45	0.28	0.10	0.48	0.37	0.24	0.09	0.53	0.42	0.24	0.09
ΔΕ12	0.28	0.43	0.23	0.06	0.56	0.24	0.16	0.08	0.68	0.12	0.23	0.09

Πίνακας 5.18: ΔΕΙΚΤΗΣ C_{ip}_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 3 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ
PROMETHEE

Από την παρουσίαση των τιμών της εντροπίας του διαχωρισμού των εναλλακτικών σε τρεις ομάδες με τη βοήθεια των αποστάσεων από τα χαρακτηριστικά σημεία κάθε ομάδας, στον πίνακα 5.19, προκύπτει ότι κατά μέσο όρο οι καλύτερες τιμές τις εντροπίας επιτυγχάνονται με την εφαρμογή των μεθόδων ΔΕ11 και ΔΕ8. Ακόμα παρατηρείται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με ίσα βάρη, βάρη ROC και πραγματικά βάρη με $\sigma=0,7$.

Βάρη	1/7,1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0,3, 0,1, 0,1, 0,1, 0,1, 0,1, 0,2				0,5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.57	0.62	0.63	0.66	0.59	0.65	0.66	0.63	0.53	0.68	0.74	0.57
ΔΕ2	0.59	0.63	0.63	0.08	0.57	0.67	0.65	0.62	0.53	0.68	0.65	0.59
ΔΕ3	0.57	0.63	0.63	0.66	0.59	0.66	0.66	0.63	0.50	0.67	-0.73	0.58
ΔΕ4	0.85	0.80	0.74	0.03	0.91	0.84	0.73	0.72	0.82	0.86	0.75	0.72
ΔΕ5	0.87	0.79	0.73	0.79	0.91	0.85	0.72	0.72	0.86	0.86	0.75	0.69
ΔΕ6	0.87	0.85	0.73	0.78	0.89	0.84	0.73	0.73	0.82	0.83	0.75	0.72
ΔΕ7	0.73	0.74	0.70	0.71	0.75	0.79	0.70	0.70	0.59	0.83	0.67	0.60
ΔΕ8	0.63	0.67	0.68	0.66	0.61	0.63	0.62	0.61	0.56	0.66	0.55	0.55
ΔΕ9	0.75	0.74	0.69	0.70	0.77	0.77	0.71	0.66	0.84	0.82	0.66	0.59
ΔΕ10	0.71	0.72	0.66	0.69	0.82	0.80	0.65	0.61	0.81	0.83	0.70	0.53
ΔΕ11	0.54	0.65	0.60	0.63	0.52	0.67	0.54	0.54	0.47	0.57	0.51	0.50
ΔΕ12	0.71	0.70	0.67	0.65	0.81	0.79	0.65	0.64	0.82	0.82	0.64	0.54

Πίνακας 5.19: ΤΙΜΕΣ ΕΝΤΡΟΠΙΑΣ ΔΙΑΧΩΡΙΣΜΟΥ ΒΑΣΕΙ ΤΩΝ ΑΠΟΣΤΑΣΕΩΝ ΣΕ 3 ΟΜΑΔΕΣ

Στους πίνακες 5.20, 5.21, 5.22 και 5.23 παρουσιάζονται αντίστοιχα οι δείκτες H_{ip_max} , H_{ip_mean} , C_{ip_max} και C_{ip_mean} . Με βάση και τους τέσσερις δείκτες επιβεβαιώνεται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τη χρήση ίσων βαρών, των πραγματικών και βαρών ROC και $\sigma=0,7$. Κατά μέσο όρο οι καλύτεροι δείκτες H_{ip_max} επιτυγχάνονται από τις μεθόδους ΔΕ4, ΔΕ3 και ΔΕ2.

Βάρη	1/7,1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0,3, 0,1, 0,1, 0,1, 0,1, 0,1, 0,2				0,5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.86	0.71	0.41	0.15	0.90	0.75	0.40	0.17	1.00	0.87	0.41	0.18
ΔΕ2	0.86	0.71	0.41	0.15	0.90	0.75	0.39	0.17	0.97	0.87	0.50	0.19
ΔΕ3	0.86	0.71	0.41	0.15	0.90	0.75	0.39	0.17	0.97	0.87	0.43	0.17
ΔΕ4	1.00	0.91	0.49	0.24	1.00	0.93	0.57	0.24	1.00	0.99	0.59	0.24
ΔΕ5	1.00	0.88	0.52	0.24	1.00	0.96	0.56	0.24	1.00	0.99	0.60	0.22
ΔΕ6	1.00	0.92	0.50	0.24	1.00	0.95	0.56	0.24	1.00	0.99	0.60	0.24
ΔΕ7	1.00	0.86	0.57	0.23	1.00	0.93	0.66	0.26	1.00	0.97	0.79	0.31
ΔΕ8	1.00	0.89	0.56	0.24	1.00	0.88	0.62	0.24	1.00	0.95	0.74	0.28
ΔΕ9	1.00	0.85	0.59	0.22	1.00	0.93	0.66	0.26	1.00	0.98	0.79	0.32
ΔΕ10	1.00	0.86	0.59	0.22	1.00	0.93	0.67	0.25	1.00	0.97	0.78	0.31
ΔΕ11	1.00	0.99	0.53	0.25	1.00	0.96	0.57	0.21	1.00	0.97	0.68	0.22
ΔΕ12	1.00	0.86	0.59	0.22	1.00	0.93	0.67	0.25	1.00	0.97	0.79	0.31

Πίνακας 5.20: ΔΕΙΚΤΗΣ H_{ip_max} ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 3 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Αντίστοιχα, οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη H_{ip_mean} είναι οι ΔΕ6, ΔΕ4 και ΔΕ5.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.34	0.26	0.12	0.04	0.36	0.24	0.11	0.03	0.37	0.24	0.09	0.03
ΔΕ2	0.35	0.26	0.12	0.04	0.36	0.24	0.11	0.03	0.37	0.24	0.12	0.03
ΔΕ3	0.35	0.26	0.12	0.04	0.36	0.24	0.11	0.03	0.37	0.25	0.10	0.03
ΔΕ4	0.33	0.22	0.10	0.03	0.33	0.22	0.10	0.03	0.30	0.19	0.08	0.02
ΔΕ5	0.30	0.22	0.10	0.03	0.33	0.22	0.09	0.03	0.32	0.20	0.09	0.02
ΔΕ6	0.33	0.22	0.10	0.03	0.34	0.22	0.10	0.03	0.30	0.20	0.08	0.02
ΔΕ7	0.37	0.26	0.12	0.04	0.38	0.26	0.12	0.04	0.38	0.27	0.12	0.03
ΔΕ8	0.36	0.26	0.11	0.03	0.37	0.26	0.12	0.03	0.38	0.26	0.11	0.03
ΔΕ9	0.37	0.26	0.12	0.04	0.38	0.26	0.12	0.04	0.40	0.27	0.12	0.03
ΔΕ10	0.35	0.25	0.11	0.04	0.38	0.26	0.11	0.03	0.39	0.26	0.12	0.03
ΔΕ11	0.35	0.25	0.11	0.03	0.36	0.25	0.11	0.03	0.37	0.25	0.10	0.03
ΔΕ12	0.35	0.25	0.11	0.04	0.38	0.26	0.11	0.04	0.39	0.26	0.11	0.03

Πίνακας 5.21: ΔΕΙΚΤΗΣ HIp_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 3 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη CIp_max είναι οι ΔΕ6, ΔΕ4 και ΔΕ5.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.93	0.87	0.28	0.10	0.90	0.61	0.53	0.15	0.99	0.94	0.70	0.26
ΔΕ2	0.90	0.65	0.29	0.10	1.00	0.60	0.56	0.14	0.97	0.89	0.64	0.23
ΔΕ3	0.93	0.75	0.28	0.10	0.92	0.54	0.56	0.13	0.97	0.69	0.69	0.27
ΔΕ4	0.68	0.54	0.32	0.11	0.70	0.54	0.28	0.08	0.67	0.44	0.47	0.23
ΔΕ5	0.74	0.59	0.33	0.11	0.76	0.70	0.22	0.11	0.84	0.55	0.48	0.17
ΔΕ6	0.71	0.63	0.30	0.10	0.70	0.56	0.24	0.14	0.72	0.50	0.35	0.19
ΔΕ7	0.77	0.82	0.46	0.20	0.86	0.74	0.54	0.22	1.00	0.91	0.54	0.25
ΔΕ8	0.92	0.81	0.57	0.22	0.97	0.81	0.47	0.20	1.00	0.76	0.43	0.22
ΔΕ9	0.81	0.78	0.47	0.22	0.94	0.68	0.43	0.22	0.98	0.91	0.59	0.25
ΔΕ10	0.86	0.80	0.58	0.21	0.90	0.91	0.53	0.22	0.98	0.90	0.54	0.24
ΔΕ11	0.98	0.84	0.57	0.25	1.00	0.73	0.48	0.20	1.00	0.68	0.42	0.22
ΔΕ12	0.85	0.84	0.53	0.23	0.87	0.80	0.54	0.21	0.99	0.95	0.52	0.24

Πίνακας 5.22: ΔΕΙΚΤΗΣ CIp_max ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 3 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Τέλος, κατά μέσο όρο οι καλύτεροι δείκτες CIp_mean επιτυγχάνονται και σε αυτή την περίπτωση από τις μεθόδους ΔΕ6, ΔΕ4 και ΔΕ5.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.51	0.44	0.07	0.02	0.43	0.20	0.20	0.04	0.50	0.53	0.28	0.08
ΔΕ2	0.47	0.22	0.07	0.02	0.59	0.20	0.23	0.03	0.47	0.51	0.23	0.07
ΔΕ3	0.51	0.36	0.07	0.03	0.46	0.17	0.22	0.03	0.47	0.29	0.28	0.09
ΔΕ4	0.22	0.17	0.06	0.02	0.30	0.23	0.07	0.01	0.45	0.21	0.19	0.09
ΔΕ5	0.27	0.22	0.07	0.02	0.33	0.35	0.04	0.02	0.26	0.29	0.17	0.05
ΔΕ6	0.28	0.23	0.06	0.02	0.22	0.24	0.04	0.03	0.47	0.17	0.12	0.07
ΔΕ7	0.23	0.39	0.14	0.06	0.36	0.28	0.18	0.07	0.55	0.30	0.23	0.09
ΔΕ8	0.42	0.38	0.25	0.09	0.45	0.35	0.22	0.09	0.47	0.31	0.24	0.09
ΔΕ9	0.24	0.35	0.16	0.08	0.44	0.21	0.14	0.08	0.40	0.36	0.23	0.09
ΔΕ10	0.43	0.40	0.24	0.07	0.44	0.45	0.19	0.08	0.61	0.45	0.22	0.09
ΔΕ11	0.52	0.45	0.28	0.10	0.49	0.35	0.24	0.09	0.47	0.34	0.24	0.09
ΔΕ12	0.38	0.43	0.22	0.08	0.35	0.36	0.19	0.07	0.56	0.50	0.23	0.09

Πίνακας 5.23: ΔΕΙΚΤΗΣ CIp_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 3 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Συνολικά δηλαδή προκύπτει ότι στον πολυκριτήριο διαχωρισμό σε τρεις ομάδες με βάση τις αποστάσεις των εναλλακτικών από τα σημεία αναφοράς κάθε ομάδας ότι οι μέθοδοι ΔΕ4, ΔΕ5 και ΔΕ6 έχουν την καλύτερη απόδοση.

Ενδιαφέρον παρουσιάζει η σύγκριση της εντροπίας αλλά και των δεικτών Hip_max, Hip_mean, CIp_max και CIp_mean που προκύπτουν από τον διαχωρισμό σε τρεις ομάδες, όχι με πολυκριτήρια μέθοδο, αλλά με μια μέθοδο κλασσικής ομαδοποίησης που βασίζεται στις αποστάσεις. Η μέση τιμή της εντροπίας σε αυτή την περίπτωση είναι 0,63. Στον πίνακα 5.24 παρουσιάζονται οι τιμές των δεικτών Hip_max, Hip_mean, CIp_max και CIp_mean, υπολογισμένοι με διαφορετικά βάρη και σ κάθε φορά. Είναι εμφανές ότι και σε αυτή την περίπτωση σημαντικά καλύτερες τιμές στους δείκτες προκύπτουν με πραγματικά, ίσα και βάρη ROC με $\sigma=0.7$. Συγκρίνοντας τις τιμές των δεικτών στην περίπτωση της κλασσικής ομαδοποίησης με αυτές της πολυκριτηρίας για τις ίδιες περιπτώσεις βαρών και σ προκύπτει ότι η πολυκριτήρια ομαδοποίηση έχει σημαντικά καλύτερα αποτελέσματα και στην περίπτωση των ροών της PROMETHEE και στην περίπτωση των αποστάσεων.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
Hip_max	1.00	0.86	0.52	0.25	1.00	0.90	0.50	0.21	1.00	0.94	0.70	0.25
Hip_mean	0.35	0.26	0.11	0.03	0.37	0.27	0.11	0.03	0.38	0.28	0.12	0.04
CIp_max	1.00	0.90	0.64	0.26	1.00	0.93	0.57	0.23	1.00	0.97	0.52	0.24
CIp_mean	0.58	0.48	0.30	0.11	0.52	0.42	0.26	0.10	0.46	0.36	0.23	0.09

Πίνακας 5.24: ΔΕΙΚΤΕΣ HI ΚΑΙ CI ΚΛΑΣΣΙΚΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ ΣΕ 3 ΟΜΑΔΕΣ

Τέλος, σημειώνεται ότι οι τιμές της εντροπίας και των δεικτών Hip_max, Hip_mean δείχνουν η ομαδοποίηση σε τρεις ομάδες είναι σωστότερη από αυτή σε δύο. Αντίθετα, οι δείκτες CIp_max και CIp_mean οδηγούν προς το αντίθετο συμπέρασμα, αφού στην περίπτωση της πολυκριτηρίας ομαδοποίησης σε δύο ομάδες εμφανίζουν μικρότερες τιμές απ' ότι σε τρεις ομάδες.

5.3.3 Αποτελέσματα Διαχωρισμού σε 4 Ομάδες

Τελευταία παρουσιάζονται τα αποτελέσματα που αφορούν το διαχωρισμό των εναλλακτικών σε τέσσερις ομάδες. Στον πίνακα 5.25 παρουσιάζονται οι τιμές τις εντροπίας που προέκυψαν με χρήση ροών της PROMETHEE. Παρατηρείται ότι κατά μέσο όρο οι καλύτερες τιμές τις εντροπίας επιτυγχάνονται με την εφαρμογή των μεθόδων ΔΕ9, και ΔΕ7. Ακόμα παρατηρείται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τα πραγματικά βάρη και βάρη ROC με 0.7 και αμέσως μετά τα πραγματικά βάρη με $\sigma=0,3$ και ίσα βάρη με $\sigma=0,7$.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0,3, 0,1, 0,1, 0,1, 0,1, 0,1, 0,2				0,5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.60	0.55	0.51	0.53	0.57	0.59	0.53	0.47	0.43	0.46	0.58	0.47
ΔΕ2	0.71	0.61	0.57	0.44	0.76	0.69	0.48	0.33	0.83	0.69	0.64	0.49
ΔΕ3	0.66	0.55	0.67	0.68	0.81	0.60	0.65	0.70	0.77	0.77	0.67	0.57
ΔΕ4	0.59	0.55	0.56	0.56	0.57	0.59	0.58	0.52	0.45	0.43	0.60	0.49
ΔΕ5	0.62	0.56	0.51	0.53	0.56	0.60	0.53	0.51	0.43	0.44	0.58	0.46
ΔΕ6	0.71	0.64	0.43	0.46	0.83	0.66	0.48	0.32	0.83	0.77	0.55	0.47
ΔΕ7	0.69	0.50	0.47	0.52	0.60	0.39	0.39	0.40	0.57	0.44	0.27	0.33
ΔΕ8	0.71	0.64	0.52	0.44	0.78	0.70	0.57	0.37	0.82	0.72	0.63	0.49
ΔΕ9	0.53	0.59	0.47	0.46	0.41	0.28	0.35	0.40	0.23	0.44	0.29	0.39
ΔΕ10	0.72	0.63	0.47	0.49	0.81	0.67	0.47	0.39	0.84	0.74	0.53	0.52
ΔΕ11	0.61	0.56	0.70	0.64	0.81	0.66	0.57	0.60	0.67	0.79	0.64	0.68
ΔΕ12	0.68	0.54	0.65	0.64	0.69	0.58	0.67	0.67	0.64	0.74	0.67	0.73

Πίνακας 5.25: ΤΑΞΙΝΟΜΗΣΗ ΒΑΣΕΙ ΤΩΝ ΡΟΩΝ ΤΗΣ PROMETHEE ΣΕ 4 ΟΜΑΔΕΣ

Στους πίνακες 5.26, 5.27, 5.28 και 5.29 παρουσιάζονται αντίστοιχα οι δείκτες H_{ip_max} , H_{ip_mean} , C_{ip_max} και C_{ip_mean} . Με βάση και τους τέσσερις δείκτες επιβεβαιώνεται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τη χρήση ίσων βαρών, των πραγματικών και βαρών ROC και $\sigma=0,7$. Κατά μέσο όρο οι καλύτεροι δείκτες H_{ip_max} επιτυγχάνονται από τις μεθόδους ΔΕ2, ΔΕ1 και ΔΕ3.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0,3, 0,1, 0,1, 0,1, 0,1, 0,1, 0,2				0,5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.79	0.59	0.37	0.14	0.80	0.60	0.37	0.15	0.93	0.80	0.37	0.16
ΔΕ2	0.76	0.59	0.37	0.14	0.81	0.60	0.37	0.15	0.93	0.80	0.37	0.13
ΔΕ3	0.80	0.59	0.36	0.15	0.81	0.61	0.38	0.15	0.94	0.80	0.37	0.15
ΔΕ4	0.88	0.65	0.45	0.19	0.98	0.75	0.44	0.20	1.00	0.99	0.51	0.22
ΔΕ5	0.88	0.66	0.46	0.18	0.98	0.82	0.44	0.19	1.00	0.99	0.43	0.18
ΔΕ6	0.89	0.66	0.40	0.18	0.95	0.71	0.48	0.21	1.00	0.98	0.50	0.24
ΔΕ7	0.96	0.82	0.55	0.23	1.00	0.93	0.63	0.25	1.00	0.98	0.77	0.31
ΔΕ8	1.00	0.82	0.53	0.24	1.00	0.86	0.56	0.21	1.00	0.97	0.57	0.21
ΔΕ9	0.96	0.83	0.53	0.22	0.99	0.92	0.65	0.24	1.00	0.98	0.76	0.31
ΔΕ10	1.00	0.84	0.56	0.22	1.00	0.94	0.67	0.25	1.00	0.98	0.77	0.31
ΔΕ11	0.99	0.86	0.57	0.26	1.00	0.80	0.51	0.18	1.00	0.97	0.53	0.19
ΔΕ12	1.00	0.84	0.55	0.22	1.00	0.93	0.64	0.25	1.00	0.98	0.77	0.31

Πίνακας 5.26: ΔΕΙΚΤΗΣ H_{ip_max} ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 4 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ PROMETHEE

Αντίστοιχα, οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη HIp_mean είναι οι ΔΕ5, ΔΕ6 και ΔΕ4.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.34	0.24	0.11	0.03	0.35	0.23	0.09	0.03	0.35	0.21	0.07	0.02
ΔΕ2	0.34	0.24	0.11	0.03	0.35	0.23	0.10	0.03	0.35	0.21	0.07	0.02
ΔΕ3	0.34	0.24	0.11	0.03	0.35	0.23	0.09	0.03	0.35	0.21	0.07	0.02
ΔΕ4	0.33	0.23	0.09	0.03	0.30	0.22	0.09	0.03	0.27	0.16	0.07	0.02
ΔΕ5	0.32	0.23	0.09	0.03	0.31	0.21	0.08	0.02	0.27	0.14	0.07	0.02
ΔΕ6	0.31	0.23	0.09	0.03	0.32	0.22	0.09	0.03	0.29	0.16	0.07	0.02
ΔΕ7	0.35	0.25	0.10	0.03	0.36	0.23	0.11	0.03	0.37	0.22	0.10	0.03
ΔΕ8	0.34	0.24	0.10	0.03	0.36	0.23	0.10	0.03	0.36	0.21	0.08	0.02
ΔΕ9	0.35	0.26	0.10	0.03	0.36	0.24	0.11	0.03	0.37	0.22	0.10	0.03
ΔΕ10	0.35	0.24	0.10	0.03	0.36	0.23	0.10	0.03	0.37	0.22	0.09	0.03
ΔΕ11	0.33	0.24	0.10	0.03	0.34	0.22	0.10	0.03	0.33	0.20	0.08	0.02
ΔΕ12	0.35	0.24	0.10	0.03	0.36	0.23	0.11	0.03	0.37	0.22	0.09	0.03

Πίνακας 5.27: ΔΕΙΚΤΗΣ HIp_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 4 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ
PROMETHEE

Οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη CIp_max είναι οι ΔΕ4, ΔΕ5 και ΔΕ6.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.80	0.54	0.30	0.14	0.98	0.65	0.57	0.23	0.96	0.68	0.54	0.29
ΔΕ2	0.82	0.53	0.37	0.17	0.99	0.76	0.58	0.22	0.99	0.64	0.64	0.29
ΔΕ3	0.81	0.54	0.30	0.15	0.93	0.65	0.60	0.23	1.00	0.59	0.53	0.28
ΔΕ4	0.77	0.66	0.35	0.14	0.85	0.79	0.49	0.10	0.64	0.47	0.63	0.22
ΔΕ5	0.75	0.66	0.35	0.14	0.74	0.77	0.46	0.13	0.71	0.49	0.59	0.26
ΔΕ6	0.84	0.66	0.29	0.13	0.90	0.72	0.65	0.11	0.76	0.55	0.65	0.29
ΔΕ7	0.86	0.79	0.45	0.18	0.92	0.78	0.56	0.21	1.00	0.81	0.50	0.26
ΔΕ8	0.92	0.80	0.44	0.16	0.93	0.96	0.48	0.21	0.92	0.89	0.48	0.22
ΔΕ9	0.92	0.88	0.46	0.19	0.92	0.76	0.54	0.22	1.00	0.71	0.46	0.26
ΔΕ10	0.93	0.78	0.40	0.15	0.99	0.71	0.48	0.17	0.94	0.74	0.48	0.25
ΔΕ11	0.98	0.89	0.54	0.27	0.99	0.99	0.52	0.17	1.00	0.83	0.60	0.24
ΔΕ12	0.90	0.82	0.39	0.19	0.99	0.71	0.52	0.19	1.00	0.68	0.46	0.25

Πίνακας 5.28: ΔΕΙΚΤΗΣ CIp_max ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 4 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ
PROMETHEE

Τέλος, κατά μέσο όρο οι καλύτεροι δείκτες CIp_mean επιτυγχάνονται από τις μεθόδους ΔΕ4, ΔΕ5 και ΔΕ1.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0,3, 0,1, 0,1, 0,1, 0,1, 0,1, 0,2				0,5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.37	0.19	0.09	0.05	0.57	0.31	0.23	0.08	0.59	0.35	0.29	0.11
ΔΕ2	0.43	0.19	0.16	0.07	0.60	0.39	0.21	0.08	0.58	0.32	0.29	0.13
ΔΕ3	0.40	0.19	0.09	0.05	0.50	0.30	0.25	0.08	0.62	0.31	0.29	0.14
ΔΕ4	0.38	0.28	0.14	0.04	0.48	0.42	0.20	0.02	0.37	0.38	0.36	0.11
ΔΕ5	0.37	0.29	0.14	0.04	0.38	0.43	0.23	0.06	0.38	0.38	0.31	0.09
ΔΕ6	0.49	0.31	0.08	0.04	0.49	0.36	0.28	0.03	0.38	0.38	0.42	0.13
ΔΕ7	0.42	0.38	0.13	0.05	0.42	0.37	0.24	0.07	0.64	0.36	0.27	0.13
ΔΕ8	0.49	0.38	0.16	0.04	0.49	0.53	0.22	0.08	0.56	0.54	0.25	0.12
ΔΕ9	0.44	0.43	0.16	0.05	0.46	0.37	0.24	0.08	0.64	0.33	0.26	0.13
ΔΕ10	0.51	0.40	0.14	0.05	0.58	0.35	0.22	0.07	0.59	0.37	0.26	0.13
ΔΕ11	0.53	0.49	0.24	0.10	0.55	0.56	0.25	0.08	0.58	0.44	0.30	0.13
ΔΕ12	0.51	0.45	0.16	0.05	0.60	0.33	0.25	0.08	0.66	0.32	0.27	0.14

Πίνακας 5.29: ΔΕΙΚΤΗΣ C_{IP}_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 4 ΟΜΑΔΕΣ ΒΑΣΕΙ ΡΟΩΝ PROMETHEE

Από την παρουσίαση των τιμών της εντροπίας του διαχωρισμού των εναλλακτικών σε τέσσερις ομάδες με τη βοήθεια των αποστάσεων από τα χαρακτηριστικά σημεία κάθε ομάδας, στον πίνακα 5.30, προκύπτει ότι κατά μέσο όρο οι καλύτερες τιμές της εντροπίας επιτυγχάνονται με την εφαρμογή των μεθόδων ΔΕ9 και ΔΕ7. Ακόμα παρατηρείται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με ίσα βάρη, πραγματικά βάρη και βάρη ROC με $\sigma=0,7$.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0,3, 0,1, 0,1, 0,1, 0,1, 0,1, 0,2				0,5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.58	0.58	0.59	0.58	0.51	0.61	0.67	0.53	0.49	0.61	0.70	0.52
ΔΕ2	0.65	0.60	0.60	0.58	0.67	0.72	0.58	0.53	0.66	0.76	0.64	0.55
ΔΕ3	0.74	0.64	0.67	0.63	0.74	0.61	0.68	0.63	0.66	0.76	0.69	0.63
ΔΕ4	0.59	0.58	0.59	0.57	0.51	0.63	0.61	0.54	0.50	0.60	0.67	0.53
ΔΕ5	0.58	0.58	0.63	0.59	0.53	0.61	0.66	0.53	0.52	0.63	0.70	0.55
ΔΕ6	0.64	0.61	0.62	0.58	0.68	0.67	0.60	0.52	0.66	0.74	0.64	0.56
ΔΕ7	0.57	0.58	0.54	0.51	0.53	0.57	0.56	0.49	0.52	0.46	0.43	0.40
ΔΕ8	0.65	0.59	0.57	0.56	0.64	0.64	0.54	0.56	0.72	0.71	0.61	0.58
ΔΕ9	0.35	0.52	0.46	0.48	0.31	0.45	0.36	0.39	0.27	0.28	0.35	0.47
ΔΕ10	0.64	0.68	0.62	0.61	0.67	0.69	0.63	0.55	0.67	0.71	0.59	0.60
ΔΕ11	0.71	0.64	0.67	0.63	0.78	0.65	0.68	0.62	0.73	0.77	0.70	0.63
ΔΕ12	0.71	0.65	0.67	0.63	0.79	0.67	0.68	0.62	0.64	0.71	0.69	0.65

Πίνακας 5.30: ΤΑΞΙΝΟΜΗΣΗ ΒΑΣΕΙ ΤΗΣ ΑΠΟΣΤΑΣΗΣ ΑΠΟ ΤΑ ΣΗΜΕΙΑ ΑΝΑΦΟΡΑΣ ΣΕ 4 ΟΜΑΔΕΣ

Στους πίνακες 5.31, 5.32, 5.33 και 5.34 παρουσιάζονται αντίστοιχα οι δείκτες H_{IP}_max, H_{IP}_mean, C_{IP}_max και C_{IP}_mean. Με βάση και τους τέσσερις δείκτες επιβεβαιώνεται ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τη χρήση ίσων βαρών, των πραγματικών και βαρών ROC και $\sigma=0,7$. Κατά μέσο όρο οι καλύτεροι δείκτες H_{IP}_max επιτυγχάνονται από τις μεθόδους ΔΕ3, ΔΕ1 και ΔΕ6.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.84	0.62	0.37	0.14	0.89	0.62	0.37	0.15	0.96	0.82	0.40	0.17
ΔΕ2	0.84	0.62	0.37	0.14	0.87	0.62	0.37	0.14	0.96	0.82	0.39	0.14
ΔΕ3	0.84	0.61	0.38	0.14	0.88	0.63	0.37	0.15	0.96	0.84	0.40	0.17
ΔΕ4	0.97	0.70	0.45	0.16	1.00	0.77	0.46	0.18	0.99	0.98	0.58	0.22
ΔΕ5	0.93	0.76	0.46	0.17	1.00	0.86	0.46	0.17	1.00	0.98	0.48	0.18
ΔΕ6	0.95	0.72	0.45	0.17	1.00	0.85	0.45	0.17	1.00	0.97	0.56	0.24
ΔΕ7	1.00	0.84	0.55	0.23	1.00	0.91	0.62	0.22	1.00	0.96	0.77	0.29
ΔΕ8	0.99	0.89	0.53	0.23	1.00	0.85	0.58	0.20	1.00	0.95	0.65	0.22
ΔΕ9	1.00	0.85	0.54	0.22	0.99	0.88	0.60	0.24	1.00	0.97	0.76	0.29
ΔΕ10	1.00	0.85	0.56	0.22	1.00	0.89	0.62	0.25	1.00	0.96	0.77	0.31
ΔΕ11	0.99	0.85	0.63	0.25	1.00	0.85	0.49	0.19	1.00	0.93	0.55	0.19
ΔΕ12	1.00	0.85	0.53	0.22	1.00	0.90	0.64	0.25	1.00	0.96	0.77	0.30

Πίνακας 5.31: ΔΕΙΚΤΗΣ H_{ip_max} ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 4 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Αντίστοιχα, οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη H_{ip_mean} είναι οι ΔΕ5, ΔΕ6 και ΔΕ4.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.34	0.23	0.11	0.03	0.35	0.23	0.10	0.03	0.36	0.23	0.10	0.03
ΔΕ2	0.34	0.23	0.11	0.03	0.35	0.23	0.10	0.03	0.36	0.23	0.10	0.02
ΔΕ3	0.34	0.24	0.11	0.03	0.35	0.23	0.10	0.03	0.36	0.22	0.09	0.03
ΔΕ4	0.31	0.22	0.09	0.03	0.31	0.21	0.09	0.03	0.31	0.19	0.07	0.02
ΔΕ5	0.32	0.22	0.09	0.03	0.31	0.20	0.09	0.03	0.32	0.18	0.08	0.02
ΔΕ6	0.32	0.22	0.09	0.03	0.30	0.20	0.09	0.03	0.31	0.19	0.08	0.02
ΔΕ7	0.35	0.25	0.11	0.03	0.37	0.25	0.11	0.03	0.38	0.26	0.11	0.03
ΔΕ8	0.34	0.25	0.10	0.03	0.36	0.25	0.10	0.03	0.37	0.24	0.10	0.03
ΔΕ9	0.35	0.25	0.10	0.03	0.37	0.26	0.11	0.03	0.38	0.26	0.11	0.03
ΔΕ10	0.35	0.25	0.10	0.03	0.36	0.25	0.11	0.03	0.38	0.24	0.11	0.03
ΔΕ11	0.33	0.24	0.10	0.03	0.34	0.23	0.10	0.03	0.35	0.21	0.08	0.02
ΔΕ12	0.34	0.25	0.11	0.03	0.36	0.25	0.11	0.03	0.38	0.24	0.11	0.03

Πίνακας 5.32: ΔΕΙΚΤΗΣ H_{ip_mean} ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 4 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Οι μέθοδοι που κατά μέσο όρο εμφανίζουν καλύτερες τιμές στον δείκτη C_{Ip_max} είναι οι ΔΕ6, ΔΕ5 και ΔΕ4.

Βάρη	1/7,1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.81	0.47	0.28	0.14	0.88	0.66	0.37	0.21	0.99	0.79	0.71	0.27
ΔΕ2	0.87	0.47	0.33	0.13	0.99	0.73	0.44	0.21	0.98	0.91	0.71	0.27
ΔΕ3	0.77	0.48	0.30	0.13	0.82	0.74	0.38	0.21	0.98	0.86	0.72	0.27
ΔΕ4	0.74	0.48	0.24	0.13	0.72	0.72	0.28	0.10	0.86	0.65	0.66	0.26
ΔΕ5	0.77	0.57	0.30	0.13	0.81	0.65	0.35	0.10	0.90	0.64	0.54	0.21
ΔΕ6	0.74	0.53	0.25	0.10	0.72	0.62	0.24	0.09	0.77	0.69	0.39	0.25
ΔΕ7	0.89	0.84	0.48	0.16	0.92	0.84	0.48	0.21	0.97	0.92	0.59	0.25
ΔΕ8	0.93	0.84	0.53	0.15	0.94	0.78	0.52	0.17	0.98	0.91	0.53	0.23
ΔΕ9	0.93	0.76	0.45	0.17	0.94	0.81	0.57	0.20	0.98	0.92	0.58	0.29
ΔΕ10	0.95	0.79	0.43	0.16	0.97	0.81	0.49	0.18	0.99	0.86	0.57	0.25
ΔΕ11	0.98	0.78	0.58	0.26	0.98	0.87	0.48	0.17	1.00	0.93	0.58	0.27
ΔΕ12	0.92	0.73	0.41	0.16	0.97	0.86	0.45	0.18	0.99	0.90	0.56	0.24

Πίνακας 5.33: ΔΕΙΚΤΗΣ CIp_max ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 4 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Τέλος, κατά μέσο όρο οι καλύτεροι δείκτες CIp_mean επιτυγχάνονται από τις μεθόδους ΔΕ6, ΔΕ5 και ΔΕ4.

Βάρη	1/7,1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
ΔΕ1	0.42	0.18	0.09	0.05	0.46	0.30	0.12	0.08	0.57	0.35	0.30	0.10
ΔΕ2	0.47	0.17	0.12	0.05	0.57	0.35	0.15	0.08	0.52	0.45	0.31	0.12
ΔΕ3	0.38	0.18	0.09	0.04	0.42	0.36	0.11	0.08	0.55	0.44	0.30	0.09
ΔΕ4	0.37	0.15	0.07	0.04	0.40	0.37	0.08	0.02	0.49	0.36	0.37	0.16
ΔΕ5	0.44	0.18	0.11	0.04	0.47	0.35	0.10	0.02	0.52	0.30	0.25	0.06
ΔΕ6	0.36	0.19	0.06	0.03	0.35	0.34	0.06	0.02	0.45	0.34	0.13	0.13
ΔΕ7	0.42	0.43	0.18	0.04	0.45	0.34	0.18	0.06	0.49	0.32	0.24	0.09
ΔΕ8	0.50	0.41	0.22	0.05	0.47	0.38	0.23	0.05	0.49	0.37	0.25	0.10
ΔΕ9	0.46	0.34	0.15	0.05	0.49	0.35	0.23	0.06	0.47	0.37	0.21	0.11
ΔΕ10	0.50	0.38	0.15	0.04	0.52	0.43	0.22	0.06	0.60	0.38	0.27	0.11
ΔΕ11	0.54	0.38	0.31	0.10	0.50	0.45	0.19	0.07	0.55	0.52	0.29	0.13
ΔΕ12	0.49	0.37	0.15	0.05	0.53	0.44	0.20	0.06	0.61	0.40	0.26	0.11

Πίνακας 5.34: ΔΕΙΚΤΗΣ CIp_mean ΔΙΑΧΩΡΙΣΜΟΥ ΣΕ 4 ΟΜΑΔΕΣ ΒΑΣΕΙ ΑΠΟΣΤΑΣΕΩΝ

Συνολικά δηλαδή προκύπτει ότι στον πολυκριτήριο διαχωρισμό σε τέσσερις ομάδες με βάση τις αποστάσεις των εναλλακτικών από τα σημεία αναφοράς κάθε ομάδας ότι οι μέθοδοι ΔΕ4, ΔΕ5 και ΔΕ6 δίνουν τα καλύτερα αποτελέσματα.

Ενδιαφέρον παρουσιάζει η σύγκριση της εντροπίας αλλά και των δεικτών HIp_max , HIp_mean , CIp_max και CIp_mean που προκύπτουν από τον διαχωρισμό σε τέσσερις ομάδες, όχι με πολυκριτήρια μέθοδο, αλλά με μια μέθοδο κλασσικής ομαδοποίησης που βασίζεται στις αποστάσεις. Η μέση τιμή της εντροπίας σε αυτή την περίπτωση είναι 0,53. Στον πίνακα 5.35 παρουσιάζονται οι τιμές των δεικτών HIp_max , HIp_mean , CIp_max και CIp_mean , υπολογισμένοι με διαφορετικά βάρη και σ κάθε φορά. Είναι εμφανές ότι και σε αυτή την περίπτωση σημαντικά καλύτερες τιμές στους δείκτες προκύπτουν με βάρη ROC, πραγματικά και ίσα βάρη με $\sigma=0.7$. Συγκρίνοντας τις τιμές των δεικτών στην περίπτωση της κλασσικής ομαδοποίησης με αυτές της

πολυκριτήριας για τις ίδιες περιπτώσεις βαρών και παραμέτρων σ προκύπτει ότι η πολυκριτήρια ομαδοποίηση έχει σημαντικά καλύτερα αποτελέσματα και στην περίπτωση των ροών της PROMETHEE και στην περίπτωση των αποστάσεων.

Βάρη	1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7				0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2				0.5, 1/30, 1/30, 1/30, 1/30, 1/30, 1/3			
σ	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7	0,01	0,1	0,3	0,7
Hip_max	1.00	0.91	0.49	0.22	1.00	0.93	0.50	0.19	1.00	0.97	0.58	0.18
Hip_mean	0.34	0.25	0.10	0.03	0.35	0.25	0.10	0.03	0.36	0.26	0.10	0.03
CIp_max	1.00	0.90	0.46	0.20	1.00	0.93	0.46	0.21	1.00	0.97	0.62	0.26
CIp_mean	0.46	0.37	0.21	0.08	0.45	0.36	0.23	0.08	0.49	0.43	0.30	0.12

Πίνακας 5.35: ΔΕΙΚΤΕΣ ΗΙ ΚΑΙ CΙ ΚΛΑΣΣΙΚΗΣ ΟΜΑΔΟΠΟΙΗΣΗΣ ΣΕ 4 ΟΜΑΔΕΣ

Αξιοσημείωτο είναι ότι, από τις τιμές της εντροπίας και των δεικτών Hip_max, Hip_mean διαφαίνεται ότι η ομαδοποίηση σε τέσσερις ομάδες οδηγεί σε καλύτερα αποτελέσματα απ' ότι σε τρεις και δυο. Αντίθετα, από τους οι δείκτες CIp_max και CIp_mean προκύπτει το συμπέρασμα ότι η πιο κατάλληλη είναι η ομαδοποίηση σε δύο ομάδες.

5.4 ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Στη συνέχεια γίνεται μια συνολική σύγκριση και αποτίμηση του συνόλου των αποτελεσμάτων. Αρχικά, στον πίνακα 5.36, παρουσιάζονται οι μέσοι όροι της εντροπίας όλων των μεθόδων και όλων των τιμών της παραμέτρου σ για κάθε αριθμό ομάδων, μεθόδου διαχωρισμού και περίπτωσης βαρών και συγκρίνονται με τις αντίστοιχες τιμές της εντροπίας για κλασσική, και όχι πολυκριτήρια, ομαδοποίηση.

Από την σύγκριση των συγκεκριμένων δεδομένων προκύπτει ότι με όλους τους τρόπους ομαδοποίησης ο διαχωρισμός σε τέσσερις ομάδες δίνει τις καλύτερες τιμές της εντροπίας, κάτι το οποίο προκαλεί εντύπωση, αφού η σύγκριση γίνεται με ομαδοποίηση σε τρεις ομάδες.

Όσο αφορά τη σύγκριση μεταξύ πολυκριτήριας και κλασσικής ομαδοποίησης, οι τιμές της εντροπίας για διαχωρισμό σε δύο ομάδες δείχνουν μικρή διαφορά στην ποιότητα της ομαδοποίησης, με την κλασσική ομαδοποίηση να επιτυγχάνει λίγο υψηλότερη τιμή εντροπίας. Αντίθετα, στον διαχωρισμό σε τρεις και τέσσερις ομάδες, η εφαρμογή κλασσικής ομαδοποίησης είχε σαν αποτέλεσμα σημαντικά χαμηλότερες τιμές εντροπίας, άρα καλύτερη ομαδοποίηση.

Συγκρίνοντας τον διαχωρισμό με τη βοήθεια της μεθόδου PROMETHEE με το διαχωρισμό με βάσει τις αποστάσεις, παρατηρείται ότι στις περισσότερες περιπτώσεις, η μέθοδος PROMETHEE επιτυγχάνει καλύτερο διαχωρισμό.

Αριθμός Ομάδων	Μέθοδος Διαχωρισμού	Ίσα Βάρη	Πραγματικά Βάρη	Βάρη Roc	Κλασσική Ομαδοποίηση
2	PROMETHEE	0,829	0,807	0,813	0,830
	Αποστάσεις	0,830	0,813	0,821	
3	PROMETHEE	0,670	0,668	0,657	0,630
	Αποστάσεις	0,667	0,699	0,677	
4	PROMETHEE	0,576	0,565	0,577	0,526
	Αποστάσεις	0,600	0,598	0,602	

Πίνακας 5.36: Σύγκριση τιμών εντροπίας

Στη συνέχεια, στους πίνακες 5.37, 5.38, 5.39 και 5.40 γίνεται σύγκριση στους μέσου όρους των τιμών των HIp_max, HIp_mean, CIp_max και CIp_mean όλων των μεθόδων και όλων των τιμών της παραμέτρου σ για κάθε αριθμό ομάδων, μεθόδου διαχωρισμού και περίπτωσης βαρών και συγκρίνονται με τις αντίστοιχες τιμές της κλασσικής ομαδοποίησης.

Στον πίνακα 5.37, ο οποίος αναφέρεται στον δείκτη HIp_max, φαίνεται ότι, για τις αντίστοιχες τιμές βαρών, η πολυκριτήρια ομαδοποίηση επιτυγχάνει καλύτερα αποτελέσματα από την κλασσική. Όσο αφορά τον αριθμό των ομάδων, σε όλες τις περιπτώσεις, καλύτερες τιμές του δείκτη επιτυγχάνονται με ομαδοποίηση σε τέσσερις ομάδες. Ενώ η PROMETHEE δείχνει να είναι καλύτερη μέθοδος διαχωρισμού από τις αποστάσεις.

Αριθμός Ομάδων	Μέθοδος Διαχωρισμού	Ίσα Βάρη	Πραγματικά Βάρη	Βάρη Roc	Κλασσική Ομαδοποίηση		
					Ίσα	Πραγμ.	ROC
2	PROMETHEE	0,69	0,70	0,76	0,74	0,76	0,85
	Αποστάσεις	0,69	0,71	0,76			
3	PROMETHEE	0,63	0,65	0,69	0,68	0,65	0,72
	Αποστάσεις	0,63	0,66	0,71			
4	PROMETHEE	0,58	0,61	0,68	0,65	0,65	0,68
	Αποστάσεις	0,60	0,62	0,68			

Πίνακας 5.37: Σύγκριση τιμών HIp_max

Σε αντίστοιχα συμπεράσματα καταλήγει κανείς παρατηρώντας και τον πίνακα 5.38, οποίος αναφέρεται στις τιμές του δείκτη HIp_mean. Η διαφορά εδώ είναι ότι οι διαφορές των τιμών είναι μικρότερες, και πάλι όμως σε όλες τις περιπτώσεις η πολυκριτήρια ομαδοποίηση είναι καλύτερη.

Αριθμός Ομάδων	Μέθοδος Διαχωρισμού	Ίσα Βάρη	Πραγματικά Βάρη	Βάρη Roc	Κλασσική Ομαδοποίηση		
					Ίσα	Πραγμ.	ROC
2	PROMETHEE	0,20	0,20	0,20	0,21	0,22	0,24
	Αποστάσεις	0,20	0,20	0,21			
3	PROMETHEE	0,19	0,18	0,17	0,19	0,19	0,20
	Αποστάσεις	0,18	0,19	0,18			
4	PROMETHEE	0,18	0,17	0,16	0,18	0,18	0,19
	Αποστάσεις	0,18	0,18	0,18			

Πίνακας 5.38: Σύγκριση τιμών HIp_mean

Παρατηρώντας τον πίνακα 5.39, ο οποίος αναφέρεται στον δείκτη CIp_mean, προκύπτουν διαφορετικά συμπεράσματα από τους προηγούμενους δείκτες. Συγκεκριμένα, οι τιμές του δείκτη για διαχωρισμό σε δύο ομάδες είναι μικρότερες, ενώ η ομαδοποίηση σε τέσσερις ομάδες παρουσιάζεται ως η χειρότερη επιλογή. Και εδώ η πολυκριτήρια ομαδοποίησης δείχνει να οδηγεί σε καλύτερα αποτελέσματα σε σχέση με την κλασσική, ενώ η ομαδοποίηση με τη μέθοδο PROMETHEE δεν βοηθάει σε όλες τις περιπτώσεις σε αντιπροσωπευτικότερο διαχωρισμό των εναλλακτικών. Για παράδειγμα, στον διαχωρισμό σε δύο ομάδες με χρήση ίσων βαρών η χρήση των αποστάσεων δείχνει πιο αποτελεσματική, αντίθετα όταν χρησιμοποιούνται τα βάρη ROC η μέθοδος PROMETHEE είναι καλύτερη.

Αριθμός Ομάδων	Μέθοδος Διαχωρισμού	Ίσα Βάρη	Πραγματικά Βάρη	Βάρη Roc	Κλασσική Ομαδοποίηση		
					Ίσα	Πραγμ.	ROC
2	PROMETHEE	0,40	0,38	0,39	0,50	0,59	0,75
	Αποστάσεις	0,39	0,38	0,43			
3	PROMETHEE	0,53	0,55	0,59	0,70	0,68	0,68
	Αποστάσεις	0,54	0,54	0,61			
4	PROMETHEE	0,53	0,60	0,60	0,64	0,65	0,71
	Αποστάσεις	0,51	0,56	0,66			

Πίνακας 39: Σύγκριση τιμών CIp_mean

Τέλος, από τον πίνακα 40, οποίος αφορά τον δείκτη CIp_max, προκύπτουν παρόμοια συμπεράσματα. Εντύπωση προκαλεί η σημαντικά μικρότερη τιμή του δείκτη για ομαδοποίηση σε δύο ομάδες σε όλες τις περιπτώσεις πολυκριτήριας και κλασσικής ομαδοποίησης. Σύμφωνα και με αυτόν τον δείκτη η πολυκριτήρια ομαδοποίηση σχηματίζει πιο ομοιόμορφες ομάδες.

Αριθμός Ομάδων	Μέθοδος Διαχωρισμού	Ίσα Βάρη	Πραγματικά Βάρη	Βάρη Roc	Κλασσική Ομαδοποίηση		
					Ίσα	Πραγμ.	ROC
2	PROMETHEE	0,09	0,08	0,08	0,14	0,17	0,21
	Αποστάσεις	0,09	0,08	0,08			
3	PROMETHEE	0,22	0,25	0,29	0,37	0,33	0,28
	Αποστάσεις	0,22	0,23	0,28			
4	PROMETHEE	0,24	0,30	0,34	0,28	0,28	0,33
	Αποστάσεις	0,23	0,26	0,32			

Πίνακας 40: Σύγκριση τιμών CIp_max

Στον πίνακα 41 παρουσιάζονται οι μέσοι όροι των τιμών των δεικτών HIp_max, HIp_mean, CIp_max και CIp_mean όλων των παραμέτρων σ , των βαρών και των ομάδων για κάθε μέθοδο υπολογισμού της αντικειμενικής συνάρτησης. Παρατηρείται ότι σύμφωνα με τον δείκτη HIp_max οι πιο αποτελεσματικές μέθοδοι είναι οι ΔΕ1, ΔΕ2 και ΔΕ3, αντίθετα σύμφωνα με τους άλλους τρεις δείκτες πιο αποτελεσματικές εμφανίζονται οι μέθοδοι ΔΕ4, ΔΕ5 και ΔΕ6.

Οι μέσοι όροι των παραπάνω δεικτών στη σωστή ομαδοποίηση που κατέληξαν οι Mousseau et al. (2000) είναι: HIp_max=0,69, HIp_mean=0,17, CIp_max=0,55 και CIp_mean=0,19. Παρατηρείται δηλαδή ότι σύμφωνα με τους δείκτες αυτούς η πολυκριτήρια ομαδοποίηση μπορεί να οδηγήσει σε πιο ομοιόμορφες ομάδες.

	PROMETHEE				ΑΠΟΣΤΑΣΕΙΣ			
	HIp_max	HIp_mean	CIp_max	CIp_mean	HIp_max	HIp_mean	CIp_max	CIp_mean
ΔΕ1	0.58	0.19	0.50	0.19	0.58	0.19	0.52	0.20
ΔΕ2	0.57	0.19	0.52	0.21	0.58	0.19	0.52	0.20
ΔΕ3	0.58	0.19	0.50	0.20	0.58	0.19	0.50	0.19
ΔΕ4	0.66	0.16	0.43	0.18	0.68	0.17	0.41	0.16
ΔΕ5	0.66	0.16	0.43	0.18	0.68	0.17	0.44	0.16
ΔΕ6	0.66	0.16	0.45	0.19	0.69	0.17	0.41	0.15
ΔΕ7	0.71	0.20	0.54	0.21	0.71	0.20	0.55	0.20
ΔΕ8	0.72	0.19	0.58	0.25	0.73	0.19	0.59	0.24
ΔΕ9	0.71	0.20	0.53	0.22	0.71	0.20	0.54	0.19
ΔΕ10	0.71	0.19	0.51	0.21	0.70	0.20	0.55	0.22
ΔΕ11	0.72	0.19	0.62	0.27	0.72	0.19	0.61	0.26
ΔΕ12	0.71	0.19	0.51	0.21	0.70	0.20	0.54	0.22

Πίνακας 41: Σύγκριση τιμών, HIp_max, HIp_mean, CIp_max, CIp_mean ανάλογα με τη μέθοδο υπολογισμού της αντικειμενικής συνάρτησης

ΚΕΦΑΛΑΙΟ 6

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η εργασία αυτή έχει ως αντικείμενο την εφαρμογή παραλλαγών του Διαφορικού Εξελικτικού Αλγόριθμου στην επίλυση προβλημάτων ομαδοποίησης με πολυκριτήριο χαρακτήρα. Στα πλαίσια λοιπόν της εργασίας διερευνήθηκε η χρήση δώδεκα διαφορετικών παραλλαγών υπολογισμού της αντικειμενικής συνάρτησης, δύο μεθόδων διαχωρισμού των εναλλακτικών, διαφορετικές τιμές βαρών και παραμέτρων των κριτηρίων. Για τα δεδομένα που αναλύθηκαν υπάρχει ήδη ένας γνωστός διαχωρισμός σε τρεις ομάδες με άλλη μέθοδο, ωστόσο επιχειρήθηκε εδώ η ομαδοποίησή τους σε δύο τρεις και τέσσερις ομάδες.

Τέλος, τα αποτελέσματα αξιολογήθηκαν με δύο διαφορετικές προσεγγίσεις. Αρχικά, συγκρίνοντας την εκάστοτε ομαδοποίηση με τον διαχωρισμό αναφοράς με την βοήθεια της εντροπίας, και στη συνέχεια με τη βοήθεια των τεσσάρων δεικτών ομοιογένειας που βασίζονται στις ροές της μεθόδου PROMETHEE.

Τα βασικά συμπεράσματα που προκύπτουν από την εφαρμογή των παραπάνω είναι:

- Δεν μπορεί να προκύψει σαφές συμπέρασμα ως προς το ποια μέθοδος υπολογισμού της αντικειμενικής συνάρτησης είναι καλύτερη, αφού ανάλογα με τις υπόλοιπες επιλογές του αλγόριθμου, άλλη μέθοδος εμφανίζει τα καλύτερα αποτελέσματα.
- Συγκρίνοντας την πολυκριτήρια ομαδοποίηση με την κλασσική ομαδοποίηση, πάντα με τη χρήση του Διαφορικού Εξελικτικού Αλγόριθμου, προκύπτει σαφής υπεροχή της πρώτης.
- Στις περισσότερες περιπτώσεις ο διαχωρισμός με τη βοήθεια της μεθόδου PROMETHEE οδηγεί σε καλύτερα αποτελέσματα, απ' ότι ο διαχωρισμός με βάση τις αποστάσεις από το σημείο αναφοράς κάθε ομάδας.
- Η χρήση των συγκεκριμένων μεθόδων με διαφορετικούς αριθμούς ομάδων μπορεί να χρησιμοποιηθεί και σε περιπτώσεις που δεν είναι γνωστός ο αριθμός των ομάδων που πρέπει να διαχωριστεί ένα σύνολο δεδομένων.
- Παρατηρείται ότι τα αποτελέσματα μπορεί να είναι πολύ διαφορετικά ανάλογα με τις τιμές των παραμέτρων, όπως τα βάρη και η παράμετρος s , που χρησιμοποιούνται σε κάθε εφαρμογή του αλγόριθμου. Γι' αυτό το

λόγο χρειάζεται να δοθεί έμφαση στον προσδιορισμό αυτών των στοιχείων. Σε αυτή την κατεύθυνση θα βοηθούσε μια μέθοδος βελτιστοποίησης, με την οποία θα μπορούσαν να προσδιοριστούν οι βέλτιστες τιμές των παραμέτρων.

Τέλος, αξίζει να αναφερθεί ότι παρόλο που το πρόβλημα της κλασσικής ομαδοποίησης έχει μελετηθεί αρκετά, δεν συμβαίνει το ίδιο με το πρόβλημα της πολυκριτήριας ομαδοποίησης, ειδικά με χρήση εξελικτικών μεθόδων. Συνεπώς, αποτελεί ένα ανοικτό πεδίο για μελλοντική έρευνα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Calinski, T., Harabasz, J., (1974), “A dendrite method for cluster analysis”, *Comm. Statist.* Vol. 3 (1), pag. 1-27
- De Smet Y., Montano Guzman L. (2004), “Towards multicriteria clustering: An extension of the k-mean algorithm”, *European Journal of Operational Research* Vol.158 pag.390-398
- Doreian P., Batagelj V., and Ferligoj A. (2005). *Generalized Blockmodeling, Structural analysis in the social sciences*. Cambridge University Press
- Δούμπος Μ., (2006), *Πολυκριτήρια Συστήματα Αποφάσεων*, Σημειώσεις Μαθήματος, Πολυτεχνείο Κρήτης Τμήμα Μηχανικών Παραγωγής και Διοίκησης
- Duda R., Hart P., Stork D., (2001), *Pattern Classification*, John Wiley & Sons, INC.
- Figueira J., De Smet Y., Brans J. (2003), “MCDA Methods For Sorting and Clustering Problems: PROMETHEE TRI and PROMETHEE Cluster”, Technical report 12, INESC Coimbra, Coimbra
- Giudici P., 2003, “Applied Data Mining, Statistical methods for business and industry”, John Wiley & Sons Ltd
- Hastie T., Tibshirani R., Friedman J. (2001), *The elements of statistical learning: data mining, inference and prediction*, Springer
- Holland, J.H., 1975, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Harbor
- Marriott, F.H.C., (1971), “Practical Problems in a method of cluster analysis”, *Biometrics* Vol. 27, pag. 501-514
- Marriott, F.H.C., (1982), “Optimisation methods of cluster analysis”, *Biometrics* Vol. 69 (2), pag. 417-422
- Maulik U., Bandyopadhyay S., (2000), “Genetic algorithm-based clustering technique.”, *The journal of Pattern Recognition Society*, Vol. 33 pag.1455-1465
- Mousseau V., Slowinski R., Zielniewicz P., (2000), “A user-oriented implementation of the ELECTRE_TRI method integrating preference elicitation support”, *Computers & Operations Research*, Vol. 27, pag. 757-777
- Nemery P. (2006), “Multicriteria Clustering”, Mémoire de Diplôme d’Etudes Approfondies en Sciences Appliquées, Université Libre de Bruxelles, Brussels, Belgium
- Nemery P., De Smet Y. (2005), “Multicriteria Ordered Clustering”, Technical Report, Université Libre de Bruxelles

- Paterlini S., Krink T. (2006), "Differential Evolution and particle swarm optimization in partitioned clustering", *Computational Statistics & Data Analysis* Vol.50 pag.1220-1247
- Price K., Storn, R., Lampinen, J., 2005, *Differential Evolution, A practical approach to global optimization*, Springer-Verlag Berlin Heidelberg
- Sarker, R., Mohammadian, M., Xin Y., 2002, *Evolutionary Optimization*, Kluwer Academic Publishers
- Tan Pang-Ning, Steinbach M., Kumar V. (2006), *Introduction to data mining*, Addison-Wesley
- Von Luxburg U., Ben-David S., (2005), Towards a statistical theory of clustering, PASCAL Workshop on Statistics and Optimization of Clustering
- Webb A. (2002), *Statistical pattern recognition*, Second Edition, Wiley.
- http://en.wikipedia.org/wiki/Data_clustering
- .