ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ ΜΕ ΘΕΜΑ

‘Ανάπτυξη συστήματος συστάσεων βασιζόμενου στη μοντελοποίηση του προφίλ των χρηστών με στόχο την προσωποποιημένη αναζήτηση’

Αγγελούσης Γεώργιος Α.Μ 2004030037

Επιβλέπων καθηγητής: Σταυρακάκης Γεώργιος
Εξεταστική επιτροπή: Καλαϊτζάκης Κων/νος
Εξεταστική επιτροπή: Ματσατσίνης Νικόλαος

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1 1
ΕΙΣΑΓΩΓΗ 1
ΚΕΦΑΛΑΙΟ 2 5
ΥΦΙΣΤΑΜΕΝΗ ΚΑΤΑΣΤΑΣΗ 5
2.1 Εισαγωγή 5
2.2 Μεθοδολογίες και Συστήματα παρόμοιου προβλήματος 6
2.3 Ο Γενετικός Αλγόριθμος 9
2.4 Το πρόβλημά μας 10
2.5 Συμπεράσματα και Προηγούμενη Κατάσταση 10
2.6 Πλεονεκτήματα της Πολυκριτηριακής Ανάλυσης σε Recommendation Systems 13
ΚΕΦΑΛΑΙΟ 3 15
ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ: ΠΟΛΥΚΡΙΤΗΡΙΑΚΗ ΑΝΑΛΥΣΗ ΕΥΡΕΤΙΚΟΙ/ΜΕΘΕΥΡΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ 15
3.1 Ο παγκόσμιος ιστός 15
3.2 Ανάκτηση και φιλτράρισμα πληροφορίας 18
3.3 Συστήματα συστάσεων 19
3.3.1 Content- Based Filtering 20
3.3.2 Collaborative Filtering 22
3.3.3 Demographic Filtering 25
3.3.4 Economic Filtering 27
3.3.5 Knowledge-based recommendations 28
3.3.6 Υβριδικές Τεχνικές 28
3.4 Μέθοδοι και τεχνικές μοντελοποίησης του προφίλ των χρηστών 30
3.5 Πολυκριτηριακή ανάλυση αποφάσεων 31
3.5.1 Κύρια χαρακτηριστικά των πολυκριτηριακών μεθόδων λήψης αποφάσεων 34
3.5.2 Μεθοδολογικό πλαίσιο της πολυκριτήριας ανάλυσης αποφάσεων 39
3.6 Μέθοδος Maut (Multi-Attribute Utility Theory) 44
3.7 Μέθοδος Mavt (Multiattribute Value Theory) 46
3.8 Μέθοδος Uta (Utility Theory Additive) 47
3.9 Καθορισμός συντελεστών βαρύτητας 54
3.10 Επιλογή του βέλτιστου σεναρίου 55
3.11 Σύστημα λήψης αποφάσεων με χρήση αθροιστικής συνάρτησης ομάδων κριτηρίων (Πολυκριτηριακή θεωρία αξίας ή χρησιμότητας - Multi – Attribute Value or Utility Theory) 56
3.12 Γραμμικά Πρόσθετα Μοντέλα 57

3.13 Ευρετικοί αλγόριθμοι (Heuristics) 58
3.14 Στατικοί ευρετικοί αλγόριθμοι (Static Heuristics) 59
3.15 Δυναμικοί ευρετικοί αλγόριθμοι (Dynamic Heuristics) 60
3.16 Περιεχόμενο, Απόκτηση, Μάθηση &Ανανέωση Μοντέλου Χρήστη 62
3.16.1 Περιεχόμενο ενός Μοντέλου Χρήστη 62
3.16.2 Τρόποι Απόκτησης του Περιεχομένου ενός Μοντέλου Χρήστη 64
3.16.3 Μάθηση και Ανανέωση των Μοντέλων 67
3.17 Μοντελοποίηση Χρήστη-(User Modeling) 68
3.17.1 Τεχνικές Μοντελοποίησης 70
3.18 Γενετικοί Αλγόριθμοι 78
3.18.1 Εισαγωγή 78
3.18.2 Μεθοδολογία Γενετικών Αλγορίθμων 79
3.18.3 Αξιολόγηση Γενετικών Αλγορίθμων 81
ΚΕΦΑΛΑΙΟ 4 84
ΜΕΘΟΔΟΛΟΓΙΑ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ 84
4.1 Εισαγωγή 84
4.2 Γραφική απεικόνιση Συστήματος 85
4.3 Ο Γενετικός Αλγόριθμος στο πρόβλημά μας 87
4.3.1 Το μοντέλο Hybrid 91
4.3.2 Fittest Τιμές 92
4.4 Πολυκριτηριακή Ανάλυση στο πρόβλημά μας 94
4.4.1 AvgSimCF 94
4.4.2 EuclideanCF 95
4.4.3 Πολυκριτήρια με χρήση Βαρών 96
4.5 Μοντέλο Δημιουργίας Προφίλ (PGM) 97
4.6 Μοντέλο Επιλογής Γειτονιάς (NSM) 97
4.7 Μέτρηση Ομοιότητας (Similarity Measure) 98
4.7.1 Collaborative Filtering Approach 98
4.8 Μοντέλο Συστάσεων(RM) 99
ΚΕΦΑΛΑΙΟ 5 103
ΑΝΑΛΥΣΗ ΚΑΙ ΣΧΕΔΙΑΣΗ ΣΥΣΤΗΜΑΤΟΣ 103
5.1 Dataset 103
5.2 Ανάλυση του Προβλήματος- Collaborative Filtering Προσέγγιση 103
5.3 Επέκταση αλγορίθμου για Πολυκριτηριακή ανάλυση 105
5.4 Aggregation Function 106
5.5 Σύνοψη Συστήματος 107
5.6 Σχήμα Αρχιτεκτονικής Συστήματος 112

ΚΕΦΑΛΑΙΟ 6 114
ΠΑΡΟΥΣΙΑΣΗ ΣΥΣΤΗΜΑΤΟΣ ΜΕΣΩ ΕΦΑΡΜΟΓΗΣ ΣΕ ΠΡΑΓΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ 114
6.1 Collaborative Filtering χωρίς πολυκριτηριακή ανάλυση 116
6.2 Multi-criteria Approach 122
6.3 EuclideanCF 124
6.4 Πολυκριτήρια με χρήση Βαρών 127
6.5 Υπολογισμός και Απεικονίσεις MAE 131
6.6 Γραφικές Απεικονίσεις Προβλήματος 132
6.7 Αξιολόγηση Συστημάτων Συστάσεων 143
ΚΕΦΑΛΑΙΟ 7 149
ΣΥΜΠΕΡΑΣΜΑΤΑ-ΜΕΛΛΟΝΤΙΚΕΣ ΕΡΓΑΣΙΕΣ 149
ΒΙΒΛΙΟΓΡΑΦΙΑ 152

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Τα συστήματα αποφάσεων (recommendation systems) είναι εργαλεία καθώς και τεχνικές οι οποίες προσφέρουν προτάσεις αντικειμένων σε χρήστες. Αυτές οι προτάσεις σχετίζονται με διαδικασίες αποφάσεων, όπως τι αντικείμενα να αγοράσουμε, τι μουσική να ακούσουμε ή τι ταινίες να διαλέξουμε ανάλογα με τις προτιμήσεις του χρήστη. Υπό αυτή την έννοια ένα ακριβές και αξιόπιστο σύστημα αποφάσεων θα πρέπει να έχει την δυνατότητα να ενεργεί εκ μέρους του χρήστη. Για παράδειγμα στο site Amazon.com υπάρχει ένα σύστημα αποφάσεων για να προσωποποιεί μια online αγορά αντικειμένου για κάθε πελάτη. Αν θεωρήσουμε ότι η κάθε σύσταση που δίνει το κάθε σύστημα είναι και για διαφορετικό χρήστη κάθε φορά τότε διαπιστώνουνε ότι οι συστάσεις διαφέρουν ανάλογα τον χρήστη. Υπάρχουν όμως και συστάσεις χωρίς προσωποποίηση του χρήστη (non-personalized). Αυτές οι συστάσεις είναι πολύ πιο απλές να παραχθούν και συναντώνται συνήθως σε περιοδικά ή εφημερίδες. Π.χ το Top Ten των CD’s του μήνα. Αν και είναι χρήσιμες καθώς και αποτελεσματικές σε μερικές περιπτώσεις, αυτού του είδους οι συστάσεις δεν απευθύνονται σε συστήματα αποφάσεων.

Στην **προσωποποιημένη αναζήτηση** με βάση τον χρήστη σε ένα Recommendation System οι συστάσεις δίνονται σαν λίστες αντικειμένων. Από αυτές τις λίστες το σύστημα προσπαθεί να προβλέψει τις προτιμήσεις του χρήστη με διάφορες μεθόδους η οποίες αναλύονται μετέπειτα στα πλαίσια της διπλωματικής εργασίας.

Η μεγάλη ανάπτυξη και ποικιλία των πληροφοριών στο Web,η ταχεία εισαγωγή των υπηρεσιών του διαδικτύου καθώς επίσης και η διαθεσιμότητα των πολλών επιλογών αντί να δημιουργεί οφέλη για τον χρήστη τον μπερδεύουν στο να πάρει κάποια απόφαση. Τα συστήματα αυτά έχει αποδειχθεί ότι είναι ζωτικής σημασίας για την λύση του προβλήματος της υπερβολικής πληροφορίας που διοχετεύεται στο Internet.

Το Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Last.fm, και IMDb είναι παραδείγματα επιτυχημένων sites τα οποία χρησιμοποιούν συστήματα αποφάσεων. Επιπλέον πολλές εταιρίες των media δημιουργούν και εξελίσσουν συστήματα αποφάσεων σαν μέρος των υπηρεσιών που παρέχουν στους πελάτες τους. Για παράδειγμα το Netflix το οποίο ασχολείται με online ενοικιάσεις ταινιών βράβευσε με ένα εκατομμύριο δολάρια την ομάδα που βελτίωσε την απόδοση του recommendation system του.
Επιπλέον ιδρύματα σε όλο τον κόσμο καθώς και προπτυχιακά αλλά και μεταπτυχιακά μαθήματα ασχολούνται με τα συστήματα συστάσεων. Επίσης tutorials και συνεντεύξεις πάνω στα RS ειναι ευρέως διαδεδομένες.

Στην παρούσα διπλωματική εργασία εξετάζεται η ανάπτυξη ενός συστήματος συστάσεων (recomendation system) βασιζόμενου στην μοντελοποίηση του προφίλ των χρηστών με στόχο την προσωποποιημένη αναζήτηση ανάλογα με τα γενικά και ειδικά χαρακτηριστικά του χρήστη και του προς αναζήτηση αντικειμένου (πχ. μουσική, video, βιβλίο, άρθρο, κλπ). Η σύστασή του αποτελείται από τη δημιουργία ενός γενετικού αλγορίθμου ο οποίος αναλύεται λεπτομερώς στα παρακάτω κεφάλαια , καθώς επίσης και με τη χρήση τεχνικών πολυκριτηριακής ανάλυσης αποφάσεων για τα πολλαπλά κριτήρια των δεδομένων (χρηστών και αντικειμένων για σύσταση).

Οι **γενετικοί αλγόριθμοι** είναι τεχνικές στοχαστικής αναζήτησης που καθοδηγούν έναν πληθυσμό λύσεων χρησιμοποιώντας τις αρχές της εξέλιξης και της φυσικής γενετικής.

Η **πολυκριτηριακή ανάλυση αποφάσεων** έχει ως βασικό αντικείμενο την αντιμετώπιση ενός προβλήματος που παρουσιάζεται κατά την προσπάθεια εξέτασης όλων των παραμέτρων ενός προβλήματος και των κριτηρίων που επηρεάζουν τη λήψη της κατάλληλης απόφασης. Το πρόβλημα αυτό αφορά τον τρόπο με τον οποίο μπορεί να γίνει η σύνθεση όλων των παραμέτρων ώστε να επιτευχθεί η λήψη ορθολογικών αποφάσεων.

Αναλυτικά η δομή της παρούσας εργασίας έχει ως εξής:

Στο 1^{ο} κεφάλαιο ορίζεται τι ακριβώς αντιπροσωπεύει, από τι αποτελείται, καθώς επίσης παραδείγματα επιτυχημένων recommendation systems.

Στο 2^{ο} κεφάλαιο αναλύεται η υφιστάμενη κατάσταση των συστημάτων αποφάσεων, ο τρόπος ανάκτησης και φιλτραρίσματος της πληροφορίας στα συστήματα αυτά καθώς επίσης τα είδη με βάση τον χρήστη και τα ειδικά χαρακτηριστικά του, οι μέθοδοι και οι τεχνικές μοντελοποίησης του προφιλ των χρηστών, οι μέθοδοι και οι τεχνικές προσωποιημένης αναζήτησης και τέλος τα είδη των διάφορων συστημάτων αποφάσεων.

Στο 3^{ο} κεφάλαιο αναλύεται η πολυκριτηριακή ανάλυση αποφάσεων τα κύρια χαρακτηριστικά και η μεθολογία στη λήψη αποφάσεων. Επίσης παρουσιάζονται τα γραμμικά πρόσθετα μοντέλα και τέλος τα είδη και η λειτουργία των ευρετικών αλγορίθμων σε θεωρητικό υπόβαθρο.

Στο 4^{ο} κεφάλαιο ακολουθεί η παρουσίαση ορισμών σχετικά με την μοντελοποίηση χρήστη και αναλύεται ο μηχανισμός που δημιουργεί προφίλ χρηστών. Παρουσιάζονται οι υπάρχουσες τεχνικές και μέθοδοι μοντελοποίησης του προφίλ χρήστη ενώ παρατίθενται και κάποιοι αλγόριθμοι, συμβατικοί αλλά και τεχνητής νοημοσύνης, που χρησιμοποιούνται για τον σκοπό αυτό. Παρουσιάζονται επίσης η ταξινόμηση και οι τεχνικές του user modeling ενώ παρουσιάζονται τεχνικές και μέθοδοι ανανέωσης του προφίλ χρηστών.

Στο 5^{ο} κεφάλαιο ακολουθεί η ανάπυξη και σχεδίαση ενός recommendation system με βάση τις μεθόδους του 4^{ου} κεφαλαίου.

Στο 6^{ο} κεφάλαιο παρουσιάζονται τα συμπεράσματα της μελέτης αυτής και οι πιθανές προεκτάσεις της προτεινόμενης μεθοδολογίας.

ΚΕΦΑΛΑΙΟ 2

ΥΦΙΣΤΑΜΕΝΗ ΚΑΤΑΣΤΑΣΗ

2.1 Εισαγωγή

Ένα recommendation system μπορεί να προτείνει σε έναν χρήστη ποια ταινία να δει ή ποιο τραγούδι να ακούσει. Έτσι ένα τέτοιο σύστημα θα πρέπει να μπορέσει να δράσει για λογαριασμο του χρήστη. Για να επιτύχει τον στόχο του θα πρέπει να διαθέτει την γνώση για της προτιμήσεις του κάθε χρήστη καθώς και τον τρόπο και τα κριτήρια με τα οποία επιλέγει κάποιο προϊόν. Τα περισσότερα συστήματα συστάσεων χρησιμοποιούν την Collaborative Filtering προσέγγιση, ενώ άλλα βασίζονται στην content-based. Υπάρχουν επίσης συστήματα τα οποίο είναι υβριδικά (hybrid) τα οποία χρησιμοποιούν τον συνδιασμό των δύο προηγούμενων μεθόδων.

Σε αυτό το σημείο αξίζει να σημειώσουμε ότι ένα σημαντικό στοιχείο για τα συστήματα αυτά είναι η προσωοποίηση του χρήστη. Αυτό συμβαίνει για να εκμεταλλευτούν καλύτερα τις πληροφορίες για τον κάθε χρήστη καθώς επίσης και να σχεδιάσουν προϊόντα και υπηρεσίες με στόχο τον κάθε χρήστη συγκεκριμένα. Η επιρροή των προσωποποιημένων υπηρεσιών, η οποία εξαρτάται πλήρως από το προφίλ των χρηστών και την ακρίβειά τους, έχει γίνει η κυρίαρχη κατάσταση στην ανάτυξη των προσωποποιημένων τεχνολογιών.

Η πλειονότητα των Recommendation Systems χρησιμοποιούν μια συνολική αριθμητική αξιολόγηση σαν είσοδο για τον αλγόριθμο σύστασης. Αυτή η αξιολόγιση εξαρτάται από ένα μόνο κριτήριο το οποίο αναπαριστά την συνολική προτίμηση ενός χρήστη πάνω σε ένα αντικείμενο. Παρόλα αυτά η ανάγκη για δημιουργία συστημάτων που μπορούν να προβλέψουν τι προϊόν θέλουν και γιατί το θέλουν οι χρήστες μας οδήγησε στα πολυδιάστατα συστήματα. Τα συστήματα αυτά βασίζονται στην πολυκριτηριακή ανάλυση δεδομένων (Multiple Criteria Decision Analysis). Στην περίπτωση αυτή υπάρχει πολλές ομάδες

αντικειμένων, ενεργειών, προϊόντων οι οποίες αξιολογούνται από πολλές σκοπιές και αναφέρονται ως κριτήρια.

2.2 Μεθοδολογίες και Συστήματα παρόμοιου προβλήματος

Ο τομέας της πολυκριτηριακής ανάλυσης δεδομένων έχει καθιερωθεί στην επιστήμη των αποφάσεων και περιέχει ένα μεγάλο αριθμό θεωριών,τεχνικών και μεθοδολογιών. Μια συνηθισμένη προσέγγιση ισχυρίζεται ότι η Multiple Criteria Decision Analysis ενεργοποιεί ένα αξιόπιστο και πειστικό μοντέλο όταν πολλαπλές εναλλακτικές πρέπει να χρησιμοποιηθούν σε διαφορετικές 'δηλώσεις' του προβλήματος.

Η προσωποποίηση του χρήστη είναι ένας ερευνητικός τομέας ο οποίος στοχεύει στην δημιουργία μοντέλων αξιολόγησης της ανθρώπινης συμπεριφοράς σε υπολογιστικό περιβάλλον.

Ένα συστήμα το οποίο επιλύει παρόμοιο πρόβλημα με το οποίο ασχολείται η παρούσα διπλωματική εργασία είναι αυτό των K.Lakiotaki, N. F. Matsatsinis, and A.Tsoukiàs, "Multi-Criteria User Modeling in Recommender Systems", *IEEE Intelligent Systems.*

Στο σύστημα αυτό ένας αριθμός χρηστών αξιολογεί μια ομάδα ταινιών. Για κάθε ταινία υπάρχουν πολλαπλά κριτήρια και ο χρήστης ζητείται να αξιολογίσει με φθίνουσα σειρά όλες τις εναλλακτικές με σκοπό την συγκρότιση μιας λίστας προτίμησης.

Στη δεύτερη φάση του προβλήματος δίνεται σαν είσοδος ένας πίνακας K+1 διάστασης. Στο σημείο αυτό χρησιμοποιείται η μέθοδος πολυκριτηριακής ανάλυσης δεδομένων. Ο πολυκριτηριακός πίνακας εισόδου αναλύεται με σκοπό την δημιουργία ενός μονού k διαστάσεων διανύσματος για κάθε χρήστη (weight vector). Κατά την διάρκεια της πολυκριτηριακής ανάλυσης εφαρμόζεται ο αλγόριθμος UTA, ένας από τους πιο διαδεδομένους αλγορίθμους στον τομέα αυτόν με σκοπό την αξιολόγηση των μοντέλων προτιμήσεων από τις ήδη υπάρχουσες δομές.

Στην συνέχεια του παραπάνω συστήματος ένας clustering αλγόριθμος χωρίζει τα αρχικά δεδομένα σε σκόρπιες ομάδες. Ο σκοπός του αλγορίθμου αυτού είναι η ομαδοποίηση παρόμοιων αντικειμένων με

αρκετή ομοιότητα μεταξύ τους και μεγάλη διαφορά από τα αντικείμενα άλλων ομάδων.

Υπάρχουν διάφοροι τρόποι για την δημιουργία μιας σύστασης στον χρήστη όπως το να διαλέξει το καλύτερο item για τον χρήστη, το να παρουσιάσει τους top-N γείτονες ή τέλος με το να κατηγοριοποιήσει τα items σε κατηγορίες όπως 'highly recommended', 'fairly recommended', 'not recommended' κτλ. Στο συγκεκριμένο παράδειγμα χρησιμοποιείται η Multi-criteria Collaborative Filtering προσέγγιση η οποία βασίζεται στην πολυδιάστατη απόσταση μεταξύ των χρηστών.

Ένα άλλο σύστημα βασισμένο στην μοτελοποίηση των χρηστών είναι αυτό των K. Lakiotaki, P. Delias, V. Sakkalis and N. F. Matsatsinis, "User Profiling based on Multi-criteria Analysis: The role of Utility Functions", *Operational Research: An International Journal*, 9(1),3-16, (2009).

Εδώ προτείνεται η μεθοδολογία για την δημιουργία προφίλ χρηστών με ακρίβεια. Όπως και στο προηγούμενο σύστημα χρησιμοποιείται το πεδίο MCDA. Η ομαδοποίηση των χρηστών με βάση τις προτιμήσεις τους σύμφωνα με την προσέγγιση των K. Lakiotaki, P. Delias, V. Sakkalis and N. F. Matsatsinis παρέχει καλύτερα αποτελέσματα όταν οι utility συναρτήσεις χρησιμοποιούνται για την αναπαράσταση των προτιμήσεων.

Η είσοδος δεδομένων στο συγκεκριμένο σύστημα αφορά τα προϊόντα λαδιού στην Γαλλία. Κάθε χρήστης, στο σύνολο 204, αξιολόγησε 6 προϊόντα λαδιού με βάση μια ομάδα 5 κριτηρίων. Τα κριτήρια ήταν η εικόνα, το χρώμα, η οσμή και η συσκευασία.

Δύο διαφορετικές αναπαραστάσεις του data set χρησιμοποιήθηκαν στην ανάλυση. Η πρώτη αναπαράσταση περιέχει τις τιμές που δόθηκαν από τους χρήστες για τα προϊόντα και η δεύτερη τη συνολική χρησιμότητα των προϊόντων η οποία υπολογίστηκε από τον αλγόριθμο UTA. Η ιδιαιτερότητα σε αυτή την προσέγγιση είναι ότι ο πίνακας αναπαρίστανται σε δυαδική μορφή και όχι με πραγματικές τιμές.

Πιο συγκεκριμένα όλες οι αξιολογήσεις των χρηστών μετατρέπονται σε δυαδική μορφή. Επειδή κάθε χρήστης αξιολόγησε 6 προϊόντα δημιουργείται ένας πίνακας 30 θέσεων για κάθε χρήστη. (λόγο των 5 κριτηρίων για κάθε προϊόν). Στην συνέχεια εφαρμόζεται μια βελτιωμένη έκδοση του UTA αλγορίθμου βασισμένου στην θεωρία MAUT. Στην

συνέχεια 2 clustering προσεγγίσεις χρησιμοποιούνται για την ομαδοποίηση παρόμοιων χρηστών, μια εναλλακτική του k-means αλγορίθμου και η προσέγγιση (AHC) των Wedel and Kamakura,2000.

Άλλα συστήματα με επίλυση παρόμοιου προβλήματος είναι αυτό των Sandra Garcia Esparza, Michael P. O'Mahony, Barry Smyth το οποίο εστιάζει στο UGC για την παροχή reviews χωρίς την χρήση ratings και metadata.Λαμβάνει επίσης υπόψην τα γενικά χαρακτηριστικά των χρηστών όπως π.χ 'μου αρέσουν τα καρτούν' για την σύσταση προβλέψεων. Ενσωματώνει επίσης την πολυκριτηριακή ανάλυση δεδομένων σε μια index-based προσέγγιση στην οποία οι χρήστες και τα αντικείμενα συσχετίζονται μεταξύ τους με όρους που έχουν επιλεγεί εκ των προτέρων.
Επίσης το σύστημα συστάσεων των Nikos G. Manouselis, and Constantina I. Costopoulou χρησιμοποιεί επίσης πολλαπλά κριτήρια για την σύσταση με μια διαφορετική αυτή την φορά προσέγγιση, την web-based. Αυτό γίνεται με την μοντελοποίηση ενός collaborative Filtering συστήματος το οποίο δημιουργεί ένα εργαλείο ελέγχου (testing tool) το οποίο βασίζεται στην Multi-Attribute Utility Theory(MAUT).

Μια μελέτη σχετικά με την εφαρμογή γενετικών αλγορίθμων πάνω σε συστήματα συστάσεων με πολυκριτηριακή ανάλυση καθώς επίσης και με χρήση βαρών στα κριτήρια εξετάζεται από τον Chein-Shung Hwang με την επέκταση ενός παραδοσιακού collaborative filtering συστήματος.

Την ίδια προσέγγιση με βάση το collaborative filtering εφαρμόζουν και οι Gulden Uchyigit και Keith Clark οι οποίοι δημιουργούν ένα agent-based σύστημα το οποίο ομαδοποιεί και ανανεώνει τα ενδιαφέροντα των χρηστών με την πάροδο του χρόνου για να κάνει κάποια σύσταση.

Στην συνέχεια οι John Pagonis και Adrian.F Clark παρουσιάζουν το σύστημα Engene το οποίο βασίζεται σε έναν γενετικό αλγόριθμο ο οποίος χρησιμοποιείται για content-based recommendation systems. Στην μελέτη αυτή συγκρίνεται ένας one-class classifier με έναν one-class k-NN classifier για κείμενα με γραμματικό περιεχόμενο.

Επιπλέον οι Nitai B. Silva, Ing-Ren Tsang, George D.C. Cavalcanti, and Ing-Jyh Tsang ανέπτυξαν ένα σύστημα το οποίο προτείνει φίλους σε μέσα κοινωνικής δικτύωσης με βάση γράφους δικτύων το οποίο ονομάζεται Oro-Aro.

Οι Liwei Liu, Freddy Lecue, Nikolay Mehandjiev, Ling Xu στην δημοσίευσή τους πρότειναν την δημιουργία ενός συστήματος συστάσεων για την σύσταση υπηρεσιών με βάση πολλαπλά κριτήρια εξετάζοντας την ποιότητα της κάθε υπηρεσίας με χρήση Context Similarities.

Τέλος στο σύστημα των K.Palanivel και R.Sivakumar προτείνεται ένα σύστημα συστάσεων ταινιών με βάση user-based και item-based αλγορίθμων με σκοπό την κανονικοποίηση διαφορετικών δεδομένων μεταξύ τους και την αξιοποίηση συστάσεων με βάση την επιλογή γειτνίασης μεταξύ χρηστών και αντικειμένων. Πολλές τεχνικές όμοιες με την παραπάνω υφίστανται στην ανάλυση των recommendation systems με την χρήση clustering, μέθοδος η οποία ξεφεύγει από τα πλαίσια της παρούσας εργασίας και είναι άσκοπο να αναφερθεί.

2.3 Ο Γενετικός Αλγόριθμος

Οι γενετικοί αλγόριθμοι είναι τεχνικές στοχαστικής αναζήτησης που καθοδηγούν έναν πληθυσμό λύσεων χρησιμοποιώντας τις αρχές της εξέλιξης και της φυσικής γενετικής. Εκτενής έρευνα έχει πραγματοποιηθεί εκμεταλλευόμενη τις αρχές του γενετικού αλγορίθμου και απεικονίζοντας τις ικανότητές τούς σε ένα ευρύ φάσμα προβλημάτων βελτιστοποίησης, συμπεριλαμβανομένης της ανάθεσης βαρών σε χαρακτηριστικά. Οι γενετικοί αλγόριθμοι μοντελοποιούνται βάσει των αρχών της εξέλιξης μέσω φυσικής επιλογής. Η συνάρτηση καταλληλότητας (fitness function) χρησιμοποιείται για την αξιολόγηση του πληθυσμού μεμονωμένα.
Ένας γενικός αλγόριθμος ξεκινά με ένα σύνολο λύσεων (πού αναπαρίστάται από τα χρωμοσώματα) ο οποίος ονομάζεται πληθυσμός. Ο αρχικός πληθυσμός δημιουργείται από μία τυχαία επιλογή λύσεων. Σε κάθε βήμα της εξέλιξης (γενιά), οι λύσεις στον τρέχοντα πληθυσμό αξιολογούνται αναφορικά με κάποια προκαθορισμένα ποιοτικά κριτήρια τα οποία αναφέρονται ως συνάρτηση καταλληλότητας. Οι λύσεις από έναν πληθυσμό χρησιμοποιούνται για τη διαμόρφωση του νέου πληθυσμού (νέα γενιά). Οι λύσεις (γονείς) επιλέγονται σύμφωνα με την συνάρτηση καταλληλότητας: όσο πιο κατάλληλοι είναι τόσο πιο πιθανό είναι να επιλεχθούν για αναπαραγωγή. Αυτές οι λύσεις « αναπαράγονται» για τη δημιουργία μίας ή περισσότερων λύσεων (απόγονοι) μέσω της διασταύρωσης και της μετάλλαξης τυχαία. Η

διαδικασία επιλογής λύσεων και εφαρμογής της διασταύρωσης και μετάλλαξης προς παραγωγή μίας περισσότερο επιτυχημένης γενιάς επαναλαμβάνεται μέχρι να επιτευχθεί μία ικανοποιητική λύση.

2.4 Το πρόβλημά μας

Στην διπλωματική αυτή εργασία δημιουργείται ένας γενετικός αλγόριθμος με βάση το collaborative filtering στην γλώσσα προγραμματισμού matlab, με ένα σετ δεδομένων από ταινίες τις οποίες αξιολόγησαν 6078 χρήστες με 4 κριτήρια για κάθε ταινία.Το σετ αυτό μας δόθηκε από την κ.Λακιωτάκη και τον κ.Ματσατσίνη σε μορφή txt . Καθώς δημιουργούμε τον αλγόριθμο αυτό στην matlab ενσωματώνουμε επίσης την επιλογή γειτνίασης για τους χρήστες με σκοπό να να υπολογίσουμε χρήστες με όμοιες προτιμήσεις. Στηριζόμαστε στην πολυκριτηριακή ανάλυση με σκοπό να χειριστούμε και τα 4 κριτήρια για κάθε αντικείμενο.
Χρησιμοποιούμε μεθόδους όπως την roulette wheel selection, την Eucledean καθώς επίσης εντάσουμε στα κριτήρια μας διάφορα βάρη με σκοπό στο τέλος να αξιολογήσουμε της διάφορες μεθόδους μεταξύ τους και να αποφασίσουμε ποια από όλες αυτές μας δίνει τα καλύτερα αποτελέσματα και προτίνει κατάλληλα μια ταινία για έναν συγκεκριμένο χρήστη. Τέλος μετράμε σε κάθε μέθοδο το MAE (Mean Absolute Error) για να συγκρίνουμε τα διάφορα αποτέλεσματα.

2.5 Συμπεράσματα και Προηγούμενη Κατάσταση

Για να κάνουμε καλές προβλέψεις σε κάθε περίπτωση είναι αναγκαίο να έχουμε επαρκή πληροφορία, στην περίπτωσή μας το αρχείο txt που μας δόθηκε από τον κ.Ματσατσίνη και την κ.Λακιωτάκη.

Οι νέες τεχνολογίες μας δίνουν την ευκαιρία να ανακτούμε με ευκολία περισσότερη πληροφορία από το Internet. Για παράδειγμα αν κάποιος θέλει να νοικιάσει μια ταινία έχει αμέτρητες επιλογές. Παρόλα αυτά η τόση πληροφορία μπορεί να οδηγήσει στην υπερφόρτωση με πληροφορία του χρήστη και να κάνει τη λήψη αποφάσεων δύσκολη.

Τα συστήματα συστάσεων και οι τεχνολογίες προσωποποίησης χρησιμεύουν ώστε να ξεπεραστεί το πρόβλημα αυτό δίνοντας συστάσεις με βάση το προφίλ του εκάστοτε χρήστη σχετίζοντας τις σχετικές πληροφορίες με αυτούς. Τα περισσότερα online shopping sites καθώς και άλλες εφαρμογές χρησιμοποιούν recommendation systems. Τα πιο γνωστά είναι το Netflix το οποίο προτείνει ταινίες καθώς και το Amazon.com. Αν οι χρήστες 'προσφέρουν' πληροφορία για τα αντικείμενα που έχουν αγοράσει ή καταναλώσει τότε το σύστημα συστάσεων μπορεί να προβλέψει τις προτιμήσεις των χρηστών για αντικείμενα που δεν έχουν δει ακόμα. Αυτό μπορεί να γίνει με βάση προηγούμενες ενέργειες και τροφοδότηση του χρήστη με σκοπό την σύσταση ενός αντικειμένου το οποίο σχετίζεται άμεσα με αυτόν.

Σε αυτή τη φάση είναι αναγκαίο να επισημάνουμε ότι τα περισσότερα recommendation systems στην ουσία χρησιμοποιούν ένα κριτήριο με σκοπό τη σύσταση. Τα συστήματα αυτά λέγονται single-rating συστήματα. Παρόλο που τα παραπάνω συστήματα είναι πολύ επιτυχημένα σε μεγάλο αριθμό εφαρμογών τα συστήματα με πολλαπλά κριτήρια εφαρμόζονται σε όλο και περισσότερες εφαρμογές στις μέρες μας. Με αυτά τα συστήματα θα ασχοληθούμε και στην παρούσα διπλωματική εργασία.

Το σύστημα Zagat's Guide, το οποίο κάνει συστάσεις εστιατορίων παρέχει τρία κριτήρια για την αξιολόγηση του κάθε εστιατορίου (φαγητό, διακόσμηση και εξυπηρέτηση). Τα online πολυκαταστήματα όπως το Circuitcity.com και το Buy.com χρησιμοποιούν επίσης πολυκριτηριακή ανάλυση για τη σύσταση ηλεκτρονικών συσκευών στων καταναλωτή. (εμφάνιση, απόδοση, διάρκεια ζωής μπαταρίας και τιμή). Αξίζει να σημειώσουμε ότι τα παραπάνω παραδείγματα δεν λαμβάνουν υπόψην την προσωποποιημένη αναζήτηση με βάση τον εκάστοτε χρήστη διότι η αξιολόγηση για κάθε κριτήριο είναι η ίδια για όλους τους χρήστες. Με σκοπό την πλήρης εκμετάλλευση των αξιολογήσεων στην πολυκριτηριακή ανάλυση δεδομένων νέες τεχνικές συστάσεων έχουν αναπτυχθεί, όπως στην Yahoo μια εφαρμογή της οποίας προτείνει ταινίες. Με τις τεχνικές αυτές οι οποίες εκμεταλλεύονται την πολυκριτηριακή ανάλυση δεδομένων θα ασχοληθούμε στην παρούσα διπλωματική εργασία.

Τα προβλήματα με πολυκριτηριακή ανάλυση δεδομένων έχουν μελετηθεί εκτενώς στα πεδία της επιστήμης. Η πλειονότητα των προβλημάτων των

μηχανικών είναι στην βάση τους πολυκριτηριακά. Για παράδειγμα στην σχεδίαση ενός αεροπλάνου, η αξιοπιστία,η αποτελεσματικότητα, το κόστος και ένας επιπλέον συνδιασμός κριτηρίων πρέπει να ληφθούν υπόψην. Τυπικές μέθοδοι για την επίλυση της πολυκριτηριακής ανάλυσης δεδομένων έχουν αναπτυχθεί όπως η λύση Pareto. Σε αυτήν γίνεται η βελτιστοποίηση του σημαντικότερου κριτηρίου και η μετατροπή άλλων κριτηρίων με περιορισμούς.

Σκοπός της πολυκριτηριακής ανάλυσης δεδομένων είναι η σύσταση ενός αντικειμένου όταν πολλαπλά κριτήρια 'αντικρούονται' και ανταγωνίζονται μεταξύ τους. Οι περισσότερες μέθοδοι αποφάσεων βασίζονται σε πολλαπλά κριτήρια. Οι μέθοδοι αυτές καθορίζουν ποιες εναλλακτικές περιπτώσεις προτιμώνται από άλλους με βάση συγκρίσεων σε κάθε κριτήριο.

Στο χώρο του marketing η αγορά ενός προϊόντος μπορεί να θεωρηθεί σαν πρόβλημα πολυκριτηριακής ανάλυσης. Για παράδειγμα στην αγορά ενός αυτοκινήτου λαμβάνουμε υπόψην πολλαπλά κριτήρια όπως την τιμή την μάρκα και το χρώμα. Το μοντέλο ενώσεων είναι το πιο διαδεμένο στον χώρο αυτό. (Green et al.2001). Το μοντέλο αυτό καθορίζει την σπουδαιότητα των βαρών για τα γνωρίσματα των προϊόντων και τις τιμές των γνωρισμάτων αυτών. Έτσι οι προτιμήσεις των χρηστών είναι ένας γραμμικός συνδιασμός βαρών και τιμών.

Η πολυκριτηριακή πληροφορία βρίσκεται επίσης σε διάφορες ηλεκτρονικές αγορές όπως τις δημοπρασίες. Εκεί λαμβάνεται υπόψην όχι μόνο η τιμή αλλά και η ποιότητα το στυλ καθώς και η ημερομηνία παράδοσης του προϊόντος.
Η επόμενη γενιά συστημάτων συστάσεων θεωρεί ως το πιο χρήσιμο τομέα την πολυκριτηριακή ανάλυση. Ο Ricci et al.(2002) ανέπτυξε ένα recommendation system για ταξιδιωτικές πληροφορίες το οποίο χρησιμοποιεί τεχνικές βασισμένες στο προφίλ του χρήστη. Οι συστάσεις δημιουργούνται όταν οι χρήστες αξιολογούν την τοποθεσία, τις δραστηριότητες και τις υπηρεσίες ενός τόπο προορισμού.

Ο Schafer (2005) δημιούργησε ένα σύστημα συστάσεων το οποίο επιτρέπει στον χρήστη να αξιολογεί το είδος της ταινίας, το μήκος και το cast και να επιλέγει πιο από τα κριτήρια αυτά είναι το πιο σπουδαίο γι' αυτόν. Για παράδειγμα αν ένας χρήστης θεωρήσει ότι το πιο σπουδαίο

κριτήριο γι'αυτόν είναι ο τύπος της ταινίας π.χ κωμωδία, τότε το σύστημα θα φιλτράρει τις ταινίες για να βρει τις κωμωδίες μόνο.

Συμπερασματικά βλέπουμε ότι υπάρχουν αρκετές προσεγγίσεις με βάση την πολυκριτηριακή ανάλυση δεδομένων σε συστήματα συστάσεων, καθώς και μεγάλος τομέας αυτών έχει μείνει ανεξερεύνητος. Μερικά από τα κενά αυτά θα προσπαθήσουμε να κλείσουμε στην διπλωματική αυτή εργασία.

2.6 Πλεονεκτήματα της Πολυκριτηριακής Ανάλυσης σε Recommendation Systems

Τα περισσότερα recommendation systems αδυνατούν να προσαρμόσουν τις συστάσεις τους σύμφωνα με τις απαιτήσεις των χρηστών. Με άλλα λόγια τα συστήματα αυτά δημιουργούνται για τις απαιτήσεις όλων των χρηστών συνολικά και όχι για την προσωποποιημένη αναζήτηση κάθε χρήστη συγκεκριμένα.

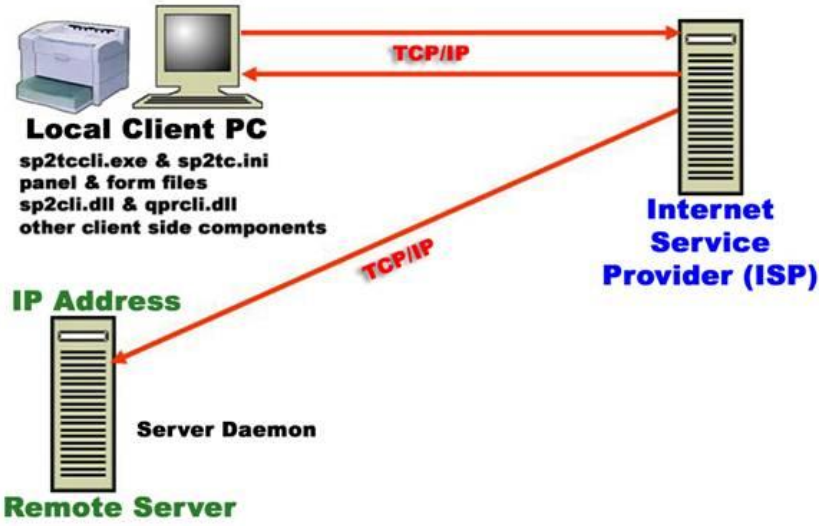
Υπήρξαν διάφορες προσπάθειες για φιλτράρισμα συστάσεων όμως οι περισσότερες εστίασαν στο φιλτράρισμα πληροφορίας για το αντικείμενο-προϊόν και όχι για τον χρήστη. (Schafer 2005). Για παράδειγμα σε ένα σύστημα σύστασης ταινιών συνήθως οι χρήστες μπορούν να επιλέξουν την αξιολόγηση της ταινίας από το είδος, τη διάρκεια, τον σκηνοθέτη κτλ. Παρόλα αυτά σε ένα σύστημα συστάσεων βασισμένο στην πολυκριτηριακή ανάλυση, ένας συγκεκριμένος χρήστης μπορεί να θέλει να επιλέξει μια ταινία με βάση κάποιο άλλο κριτήριο π.χ την πλοκή. Το κριτήριο αυτό μπορεί να είναι υποκειμενικό για τον κάθε χρήστη. Η πολυκριτηριακή ανάλυση επιτρέπει στην πληροφορία να κατανεμηθεί ανάλογα με τα ενδιαφέροντα ενός χρήστη συγκεκριμένα και όχι γενικά όλων των χρηστών μαζί σε έναν πιο προσωποποιημένο τρόπο και να προσαρμόσει τις συστάσεις ανάλογα.

ΚΕΦΑΛΑΙΟ 3

ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ: ΠΟΛΥΚΡΙΤΗΡΙΑΚΗ ΑΝΑΛΥΣΗ ΕΥΡΕΤΙΚΟΙ/ΜΕΘΕΥΡΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ

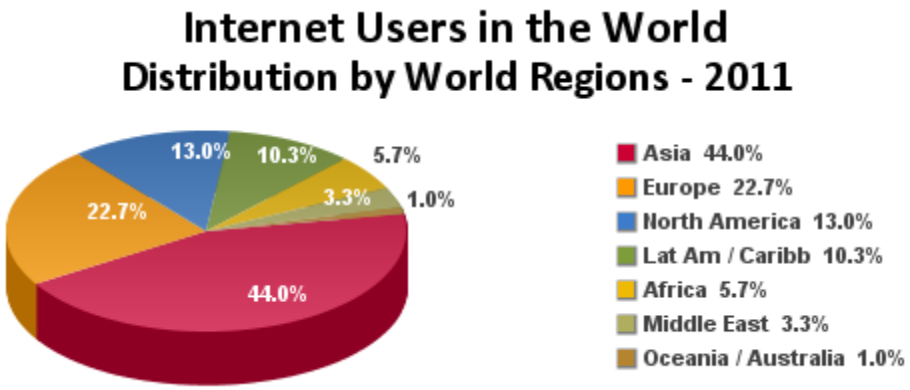
3.1 Ο παγκόσμιος ιστός

Ο παγκόσμιος ιστός (web) έκανε τα πρώτα του βήματα στις αρχές της δεκαετίας του 90. Η τεχνολογία η οποία υλοποιεί τον παγκόσμιο αυτό ιστό μπορεί να χαρακτηριστεί ως ένα σύστημα πληροφορίας το οποίο συνθέτουν πράκτορες. (Architecture of the World Wide Web,2004) Οι πράκτορες αυτοί είναι προγράμματα τα οποία δρουν εκ μέρους των χρηστών. Τα κύρια είδη των πρακτόρων αυτών είναι οι πράκτορες- εξυπηρετητές (server agents) και οι πράκτορες-πελάτες (client agents). Ο πράκτορας εξυπηρετητής προσφέρει υπηρεσίες οι οποίες χρησιμοποιούνται από τον πράκτορα πελάτη. Π.χ. όταν ο χρήστης ακολουθεί ένα link σε μια ιστοσελίδα στον browser, ο τελευταίος στέλνει ένα αίτημα στον εξυπηρετητή, ο οποίος με την σειρά του απαντάει εμφανίζοντας την ζητούμενη ιστοσελίδα.

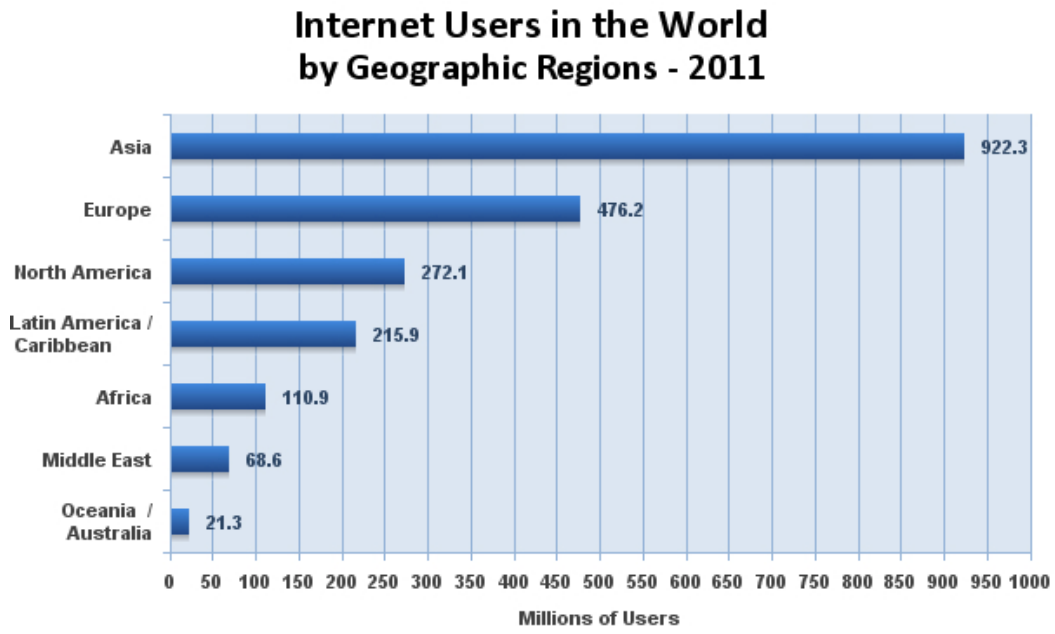


Σχήμα 1
Μοντέλο client-agent
Πηγή: www.flexus.com

Γενικότερα, στόχος του παγκόσμιου ιστού είναι η ανάπτυξη και η διατήρηση πρωτόκολλων τα οποία θα προσδιορίζουν τα πρότυπα που ενεργοποιούν τους υπολογιστές στο διαδίκτυο με σκοπό την αποθήκευση και την επικοινωνία διαφορετικών μορφών πληροφορίας.



Σχήμα 2
Κατανομή χρηστών internet παγκοσμίως
Πηγή: Internet World Stats-www.internetworldstats.com/stats.html
[16]



Σχήμα 3
Αριθμητική κατανομή χρηστών Internet
Πηγή: Internet World Stats-www.internetworldstats.com/stats.html

Τα πρότυπα που χρησιμοποιούνται είναι τα εξής:

❖ Hypertext Transfer Protocol (HTTP) (T. Berners-Lee, R. T. Fielding, and H. Frystyk *1945*) :είναι ένα πρωτόκολλο μεταφοράς το οποίο προσδιορίζει την επικοινωνία εξυπηρετητών-πελατών μεταξύ τους. Όταν ο χρήστης πληκτρολογεί το URL μιας ιστοσελίδας ή ακολουθεί ένα link σε μια άλλη ιστοσελίδα, ο browser του χρήστη στέλνει ένα HTTP αίτημα στον εξυπηρετητή. Ο τελευταίος αποκρίνεται και στέλνει με την σειρά του το περιεχόμενο της ιστοσελίδας που ζήτησε ο χρήστης.

❖ Hypertext Mark-up Language (HTML) (D. W. Connolly and L. Masinter, 2000) : χρησιμοποιείται για να προσδιοριστεί η δομή των ιστοσελίδων. Η γλώσσα αυτή περιέχει έννοιες για την ενσωμάτωση αναφορών σε άλλα έγγραφα. Αυτές οι αναφορές εμφανίζονται στις ιστοσελίδες ως υπερσύνδεσμοι (Hyperlinks) τους οποίους οι χρήστες μπορούν να διαλέξουν ώστε να εμφανίσουν την εν λόγω ιστοσελίδα. Πρόσφατα μια διαφορετική

γλώσσα, η XML δημιουργήθηκε ώστε να διευκολύνει τους χρήστες να μοιράζονται δεδομένα μέσω διαφορετικών συστημάτων πληροφορίας.

❖ Uniform Resource Locator (URL): σύστημα για την αναφορά πηγών στο web.(T. Berners-Lee, L. Masinter, and M. McChahill, 1994)

Όλα μαζί τα παραπάνω πρότυπα συνθέτουν μια απλή και αποτελεσματική πλατφόρμα για να μοιράζονται πληροφορίες οι διάφοροι χρήστες. Εξαιτίας της αναγκαιότητας αλλά και της διαθεσιμότητας των ηλεκτρονικών υπολογιστών και του Internet, ο παγκόσμιος ιστός γνώρισε μια τεράστια αύξηση, τόσο στον αριθμό των ηλεκτρονικών υπολογιστών, όσο και στον αριθμό των χρηστών του με αποτέλεσμα την μεγάλη παράλληλη αύξηση και ανάπτυξη των ιστοσελίδων. Φυσική απόρροια των παραπάνω ήταν η δημιουργία διαφόρων τεχνικών και συστημάτων τα οποία βοηθούν τους χρήστες στην εύρεση πληροφορίας στο παγκόσμιο ιστό. Αυτές οι τεχνικές ονομάζονται ανάκτηση πληροφορίας (information retrieval) και φιλτράρισμα πληροφορίας (information filtering).

3.2 Ανάκτηση και φιλτράρισμα πληροφορίας

Οι χρήστες του web έχουν στη διάθεση τους ένα τεράστιο ποσό πληροφορίας το οποίο πολλές φορές δεν μπορούν να το διαχειριστούν κατάλληλα. Αυτή η συσσώρευση πληροφορίας είναι και ο λόγος ο οποίος διάφορες τεχνικές όπως η ανάκτηση και το φιλτράρισμα πληροφορίας αναπτύχθηκαν.

Ανάκτηση πληροφορίας

Η ανάκτηση πληροφορίας (information retrieval) η οποία συνήθως σχετίζεται με την εξεύρεση δεδομένων, είναι μια τεχνολογία η οποία περιλαμβάνει τα εξής:

- Crawling : πρόσβαση σε εξυπηρετητές του παγκόσμιου ιστού και συστήματα αρχείων με σκοπό την προσκόμιση πληροφορίας
- Processing : προσαρμογή της πληροφορίας. Π.χ. διαγραφή, επεξεργασία ή αντιγραφή ενός αρχείου.
- Indexing : διεργασία η οποία δημιουργεί μια δομή δεδομένων η οποία ονομάζεται index.

- Queries : αιτήματα πληροφορίας. Συστήματα που χρησιμοποιούν τη μέθοδο ανάκτησης πληροφορίας επιτρέπουν στο χρήστη να γράψει ένα query το οποίο περιγράφει τη ζητούμενη πληροφορία. Υπάρχουν query interfaces μέσω των οποίων ο χρήστης μπορεί να επικοινωνήσει με το σύστημα. Υπάρχει επίσης query επεξεργαστής ο οποίος χρησιμοποιεί το index για να βρει παραπομπές βασιζόμενος στις λέξεις κλειδιά που πληκτρολόγησε ο χρήστης στο query με σκοπό την ανάλυση και ταυτοποίηση της πρόθεσης του χρήστη, καθώς και την επιστροφή των πιο σχετικών αποτελεσμάτων.

Το φιλτράρισμα της πληροφορίας στα συστήματα αυτά λειτουργεί με το να επιτρέπεται στον χρήστη να συγκεκριμενοποιεί τη πληροφορία που αναζητά μέσω της πληκτρολόγησης λέξεων-κλειδιών. Τα συστήματα αυτά είναι πολύ επιτυχημένα με χρήστες οι οποίοι ξέρουν ακριβώς να περιγράψουν ακριβώς τι θέλουν να βρουν.

Φιλτράρισμα πληροφορίας

Τα συστήματα που χρησιμοποιούν τη μέθοδο του φιλτραρίσματος πληροφορίας (information filtering) στοχεύουν στο φιλτράρισμα της πληροφορίας η οποία βασίζεται στο προφίλ του χρήστη. Το προφίλ αυτό μπορεί να διατηρηθεί επιτρέποντας τον χρήστη να προσδιορίσει και να συγκρίνει τα ενδιαφέροντα του με ακρίβεια ή αλλιώς επιτρέποντας στο σύστημα να κάνει την παραπάνω διεργασία.

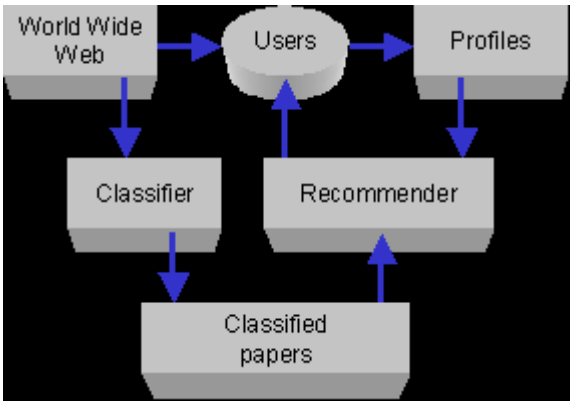
Το φιλτράρισμα σε αυτά τα συστήματα πραγματοποιείται όταν ο χρήστης αυτόματα λαμβάνει την πληροφορία που χρειάζεται η οποία βασίζεται στο προφίλ του. Η πληροφορία μεταφέρεται στον χρήστη δίνοντας του μια ειδοποίηση ή επιτρέποντας στο σύστημα να χρησιμοποιήσει την πληροφορία για λογαριασμό του. Το πλεονέκτημα αυτής της μεθόδου είναι η ικανότητα προσαρμογής στα μεγάλης διάρκειας αντικείμενα τα οποία έχει στο προφίλ του ο χρήστης.

Παρόμοιο με τι σύστημα φιλτραρίσματος πληροφορίας είναι το σύστημα το οποίο δρα ως ένας προσωπικός οδηγός αποφάσεων για τους χρήστες. Τα συστήματα αυτά λέγονται συστήματα αποφάσεων (recommendation systems).

3.3 Συστήματα συστάσεων

Τα τελευταία χρόνια τα online συστήματα συστάσεων συνήθως χρησιμοποιούνται είτε για να προβλέψουν αν ένας συγκεκριμένος χρήστης ενδιαφέρεται για κάποιο συγκεκριμένο αντικείμενο, είτε για να

προσδιορίσουν μια ομάδα N αντικειμένων τα οποία ενδεχομένως να ενδιαφέρουν έναν χρήστη. Τα συστήματα αυτά χρησιμοποιούνται σε πολλές εφαρμογές. Παραδείγματα αυτών των εφαρμογών είναι τα διαδικτυακά καταστήματα και τα διαδικτυακά μουσικά δισκοπωλεία. Παραδείγματα τέτοιων συστημάτων είναι το Recommender, το Firefly, το Ripper, το Fab και το Ringo. Τα συστήματα συστάσεων χρησιμοποιούνται έτσι ώστε να προτείνουν και επιπλέον να παρέχουν πληροφορία στους χρήστες τους, έτσι ώστε οι τελευταίοι να κάνουν εύκολα τις αγορές τους. Για την πραγματοποίηση των αγορών αυτών τα προϊόντα τα οποία είναι προς πώληση μπορούν να ταξινομηθούν με διάφορους τρόπους ανάλογα με την προηγούμενη διαδικτυακή συμπεριφορά του χρήστη ή με βάση τα προσωπικά χαρακτηριστικά του.



Σχήμα 4
Παράδειγμα recommendation System για άρθρα
Πηγή: Exploiting Synergy Between Ontologies and Recommender Systems

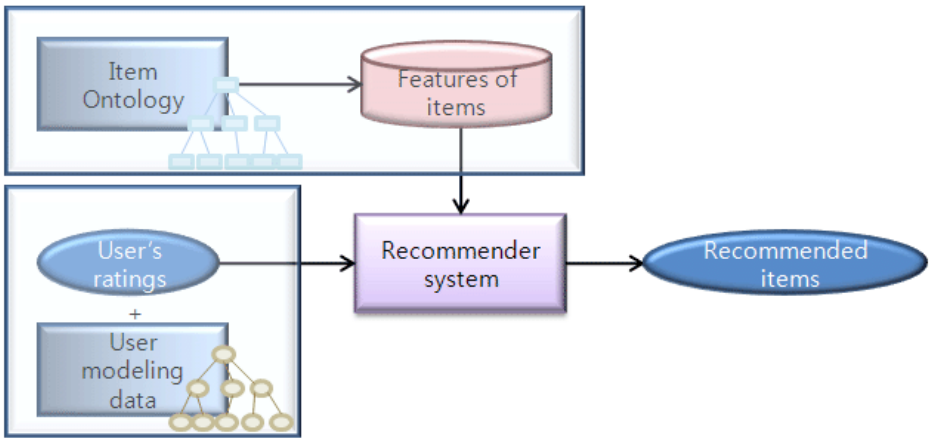
3.3.1 Content- Based Filtering

Τεχνικές οι οποίες αντιμετωπίζουν την υπερφόρτωση πληροφορίας συχνά χρησιμοποιούν το content-based φιλτράρισμα ή αλλιώς cognitive filtering (Malone et al, 1987). Το φιλτράρισμα αυτό βασίζεται στον ακριβή προσδιορισμό των χαρακτηριστικών του περιεχομένου των αντικειμένων καθώς και οποιαδήποτε άλλης πληροφορίας έτσι ώστε να ταιριάζει το κάθε αντικείμενο με πιθανούς μελλοντικούς χρήστες του συστήματος. Με άλλα

λόγια οι content-based τεχνικές συστήνουν αντικείμενα (προϊόντα) τα οποία ταιριάζουν με τα αντικείμενα για τα οποία ενδιαφέρεται κάποια συγκεκριμένη ομάδα χρηστών. Η τεχνική φιλτραρίσματος αυτή χρησιμοποιεί ταξινομητές ή μεθόδους όπως αυτή του κοντινότερου γείτονα (K-nearest neighbor). Όσον αφορά τους ταξινομητές κάθε χρήστης σχετίζεται με έναν από αυτούς και έτσι δημιουργείται το προφίλ του. Ο εκάστοτε ταξινομητής δέχεται σαν είσοδο κάποιο αντικείμενο και στη συνέχεια συμπεραίνει με βάση τα περιεχόμενα του αν κάποιος χρήστης μπορεί να ενδιαφέρεται για αυτό(Pazzani and Billsus, 2000). Τεχνικές που χρησιμοποιούν ταξινομητές σήμερα είναι τα Νευρωνικά δίκτυα (neural networks), τα δένδρα απόφασης (decision trees) καθώς και τα Μπευσιανά δίκτυα (Bayesian networks).

Σε αντίθεση με τους ταξινομητές η τεχνική φιλτραρίσματος η οποία βασίζεται στην μέθοδο του κοντινότερου γείτονα λειτουργεί ως εξής :

Αρχικά αποθηκεύει στο προφίλ του χρήστη όλα τα αντικείμενα τα οποία θεωρεί ενδιαφέροντα για αυτόν. Στη συνέχεια για να καθορίσει αν ένα αντικείμενο μπορεί να ενδιαφέρει τον χρήστη στο μέλλον ελέγχει το περιεχόμενο των αντικειμένων τα οποία είναι αποθηκευμένα στο προφίλ του χρήστη. Αν το περιεχόμενο κάποιου αντικειμένου στο προφίλ του χρήστη είναι παρόμοιο ή κοντά στο περιεχόμενο του αντικειμένου που θέλουμε να εξετάσουμε τότε το κατατάσσουμε ως ενδιαφέρον για τον χρήστη (Montaner et al, 2003; Pazzani and Billsus).



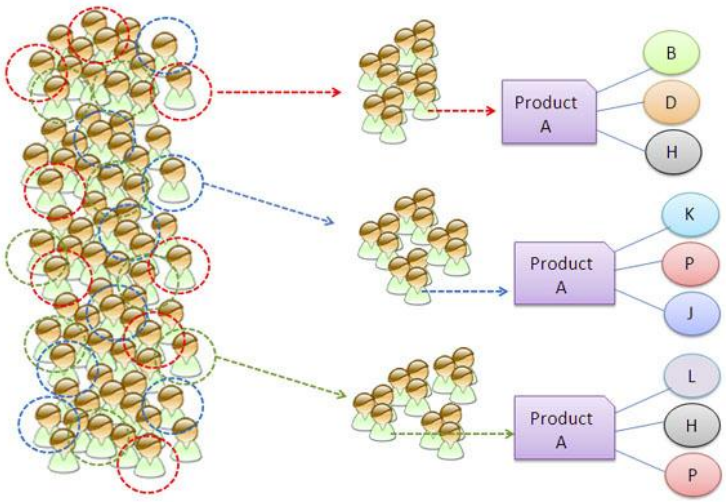
Σχήμα 5
Δομή Content-based Συστήματος
Πηγή : http://nlp.postech.ac.kr/research/previous_research/ocr

Τα κύρια πλεονεκτήματα του content-based filtering είναι πρώτον ότι το περιεχόμενο των αντικειμένων είναι γνωστό και έτσι ο χρήστης μπορεί πλήρως να κατανοήσει γιατί τα αντικείμενα αυτά επιλέχθηκαν για αυτόν(Montaner et al, 2003). Δεύτερο πλεονέκτημα είναι ότι οι τεχνικές που χρησιμοποιούν το content-based filtering επηρεάζονται λιγότερο από το cold-start problem το οποίο είναι το κύριο μειονέκτημα του collaborative filtering. Η μέθοδος φιλτραρίσματος αυτή παρουσιάζει όμως και αρκετά μειονεκτήματα. Ένα μειονέκτημα είναι η λάθος επιλογή κριτηρίων για τον χρήστη. Για παράδειγμα η επιλογή μπορεί να βασιστεί στη τιμή ενός προϊόντος ενώ ο χρήστης να θεωρεί πιο σημαντική την ποιότητα του προϊόντος αυτού. Ένα άλλο μειονέκτημα που παρουσιάζει η μέθοδος αυτή είναι ότι προτείνει στον χρήστη προϊόντα τα οποία έχει ήδη δει ενώ πολλές φορές ο χρήστης ενδιαφέρεται για κάτι διαφορετικό (over-specialization problem)(Resnick and Varian,1997, Schafer et al, 2000).
Επίσης στο content-based filtering τα αντικείμενα αναπαρίστανται σε τέτοια μορφή έτσι ώστε η σημασιολογία των διαφόρων χαρακτηριστικών τους να μην μπορεί να κατηγοριοποιηθεί εύκολα. Αυτό έχει ως αποτέλεσμα τα χαρακτηριστικά αυτά να πρέπει να ομαδοποιηθούν χειροκίνητα(Shardanand and Maes,1995). Το τελευταίο μειονέκτημα το οποίο παρατηρείται στο content- based φιλτράρισμα είναι η πολλές φορές ανακρίβεια των συστάσεων για χρήστες που έχουν λίγες προτιμήσεις έως τώρα και έτσι δεν μπορούν να συλλεγούν τα κατάλληλα χαρακτηριστικά από τα αντικείμενα που ήδη βρίσκονται στο προφίλ τους.

3.3.2 Collaborative Filtering

Το collaborative filtering σύστημα ή αλλιώς social filtering είναι ευρέως γνωστό για την χρήση του σε διαδεδομένα e-commerce sites όπως το Amazon.com, το imdb.com και το Netflix.com . Άλλα παραδείγματα τέτοιων συστημάτων είναι το Ringo,το Jester και το Video Recommender. Ένα σύστημα σαν και τα παραπάνω κάνει συστάσεις στους χρήστες του βασιζόμενο στη γνώμη και στις προτιμήσεις άλλων χρηστών με παρόμοια γούστα και ενδιαφέροντα(Breese et al, 1998). Σε σχέση με τις άλλες τεχνικές φιλτραρίσματος έχει τρία σημαντικά πλεονεκτήματα :
1^{ον} δέχεται ως είσοδο αντικειμενικές πληροφορίες για το εκάστοτε αντικείμενο-προϊόν ,(βασίζεται στα χαρακτηριστικά του αντικειμένου όπως το

στυλ, η ποιότητα κτλ.) πετυχαίνοντας έτσι καλύτερη ποιότητα συστάσεων. (Montaner et al, 2003)
2ον λαμβάνοντας υπόψη τα ενδιαφέροντα άλλων χρηστών με τις ίδιες προτιμήσεις προτείνει μεγαλύτερη ποικιλία αντικειμένων τα οποία μπορεί να μην είναι ήδη γνωστά στον χρήστη. (Montaner et al, 2003)
3ον λειτουργεί πολύ καλά για αντικείμενα με περίπλοκο περιεχόμενο και χαρακτηριστικά όπως η μουσική και οι ταινίες.(Burke, 2002)
Τα πιο διαδεδομένα collaborative filtering συστήματα είναι τα Tapestry και Grouplens. Αν και ευρέως διαδεδομένα τα συστήματα αποφάσεων αυτά που χρησιμοποιούν τη τεχνική του collaborative φιλτραρίσματος έχουν και αυτά κάποια μειονεκτήματα.



Σχήμα 6
δομή ενός collaborative συστήματος
Πηγή : http://www.bridgewell.com/recommendationEngine.html

Το πρώτο και κύριο μειονέκτημα τους είναι ότι επηρεάζονται πολύ από το cold-start problem το οποίο χωρίζεται σε δυο υποκατηγορίες την **new-system cold-start problem** και στην **new-user cold-start problem**. Στην πρώτη υποκατηγορία το πρόβλημα υφίστανται όταν κάποιο καινούριο σύστημα έχει λίγα προφίλ χρηστών με αποτέλεσμα να μην μπορεί να αναλύσει χρήστες με παρόμοιες προτιμήσεις έτσι ώστε να κάνει μια ακριβής σύσταση σε κάποιον άλλο χρήστη. Στην δεύτερη υποκατηγορία το πρόβλημα υφίστανται όταν κάποιος καινούριος χρήστης εισέρχεται στο σύστημα. Ο χρήστης αυτός έχει λίγες προτιμήσεις και έτσι δεν μπορεί να αξιολογηθεί κατάλληλα ώστε το σύστημα να μπορέσει να του κάνει κάποια σύσταση.(Middleton et al, 2002)

Ένα άλλο μειονέκτημα των συστημάτων που χρησιμοποιούν collaborative φιλτράρισμα είναι το "early-rater problem". Σε αυτό το πρόβλημα αν κάποιο αντικείμενο εισαχθεί στο σύστημα δεν μπορεί να προταθεί σε κάποιον χρήστη αν πρώτα κάποιος άλλος χρήστης το αξιολογήσει και το θεωρήσει ενδιαφέρον ή όχι. Ένα επιπλέον μειονέκτημα είναι το "sparsity problem". Στην περίπτωση αυτή το πρόβλημα προκύπτει όταν σε κάποιο σύστημα υπάρχουν πολλά αντικείμενα τα οποία δεν βρίσκονται στα προφίλ κάποιον χρηστών με συνέπεια να μην μπορούν να προταθούν στη συνέχεια. Με άλλα λόγια το πρόβλημα προκύπτει όταν στο σύστημα έχουμε πολλά αντικείμενα και λίγους χρήστες με αποτέλεσμα κάποια αντικείμενα να μην βρίσκονται στην προτίμηση κανενός χρήστη και έτσι να μην μπορούν να προταθούν ποτέ.
Πολλές φορές προκύπτουν με βάση το φιλτράρισμα αυτό λάθος συστάσεις καθώς πολλοί χρήστες μπορεί να έχουν εντελώς διαφορετικές προτιμήσεις από τους άλλους χρήστες. Το πρόβλημα αυτό επίσης οδηγεί στη δυσκολία υλοποίησης συστάσεων για τους χρήστες των οποίων οι προτιμήσεις είναι πολύ διαφορετικές.
Τέλος ένα άλλο κύριο πρόβλημα των collaborative filtering συστημάτων είναι η επεκτασιμότητα (scalability). Για να έχουμε σωστές και απόλυτα ακριβείς συστάσεις στο σύστημα αυτό απαιτείται μεγάλος αριθμός τόσο χρηστών όσο και αντικειμένων συνολικά για κάθε χρήστη ξεχωριστά (αντικείμενα του ενδιαφέροντος του).

3.3.2.1 Item-Based Collaborative Filtering

Για την αποφυγή προβλημάτων όπως το scalability και το sparsity στα collaborative filtering συστήματα αποφάσεων δημιουργήθηκε ένα νέο μοντέλο συστήματος, το "item-based collaborative filtering"(Badrul et al, 2001, Deshpande and Karypis,2004, Linden et al, 2003). Το σύστημα αυτό σε αντίθεση με το collaborative filtering ερευνά τα αντικείμενα που κάποιος χρήστης έχει αξιολογήσει (έχει κάνει rating) και εξετάζει την ομοιότητα των αντικειμένων αυτών με τα αντικείμενα που δεν έχει αξιολογήσει κανένας χρήστης στη δεδομένη στιγμή. Στην ουσία το σύστημα αυτό καθορίζει αν δυο αντικείμενα είναι παρόμοια το ένα με το άλλο με βάση παρόμοιους βαθμούς αξιολόγησης. Μια παραλλαγή του συστήματος αυτού βασίζεται όχι στις ομοιότητες των αντικειμένων αλλά στις διαφορές μεταξύ τους σε συνδυασμό με το rating. Για παράδειγμα αν έχουμε δύο αντικείμενα Α και Β το σύστημα

αυτό εξετάζει πιο από τα δύο αντικείμενα έχει υψηλότερο rating. Αν υψηλότερο rating έχει το Α τότε οι χρήστες του συστήματος αυτού που έχουν εκδηλώσει ενδιαφέρον για το Β πιθανόν να ενδιαφέρονται και για το Α επίσης (Σχήμα 7).

	Star Wars	Jurassic Park	Terminator 2	Indep. Day
User 1	7	6	3	7
User 2	7	4	4	6
User 3	3	7	7	2
User 4	4	4	6	2
User 5	7	4	3	?

Similar attributes

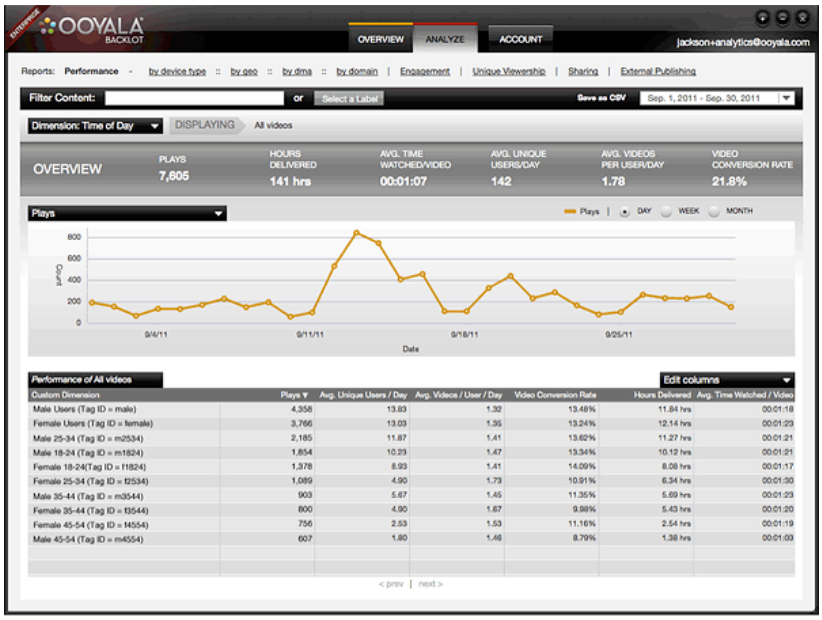
Σχήμα 7
παράδειγμα item-based collaborative συστήματος
Πηγή: google

Ο κύριος λόγος δημιουργίας του "item-based collaborative filtering" συστήματος είναι η αποφυγή του scalability και του sparsity προβλημάτων. Αυτό στη πράξη συμβαίνει γιατί υπάρχουν περισσότερα αντικείμενα που αξιολογούνται από πολλούς χρήστες παρότι χρήστες που αξιολογούν πολλά αντικείμενα. Επιπλέον η σχέση μεταξύ των αντικειμένων θεωρείται ως στατική σε αντίθεση με τις σχέσεις μεταξύ των χρηστών οι οποίες αλλάζουν δραστικά με το πέρασμα του χρόνου πράγμα που βοηθάει στην σύγκριση των αντικειμένων offline χωρίς την συμμετοχή των χρηστών, (λόγω της στατικότητας των αντικειμένων μεταξύ τους) πράγμα που αυξάνει την αποδοτικότητα του συστήματος και μειώνει το scalability problem.

3.3.3 Demographic Filtering

Τα συστήματα αποφάσεων τα οποία βασίζονται στο δημογραφικό φιλτράρισμα (demographic-based) χρησιμοποιούν ως βάση τους τα διάφορα δημογραφικά χαρακτηριστικά των χρηστών όπως π.χ. η ηλικία, το επάγγελμα,

η χώρα και η ηλικία, καθώς και τις προτιμήσεις τους ως παράγοντες σύστασης των αποφάσεων. (Σχήμα 8)
Στις μέρες μας δεν υπάρχουν πολλά συστήματα αποφάσεων τα οποία χρησιμοποιούν τη μέθοδο του demographic filtering για τον λόγω ότι τέτοιου είδους πληροφορία είναι δύσκολο να συλλεχθεί. Ο κάθε χρήστης είναι επιφυλακτικός στο να μοιραστεί πληροφορίες όπως το τηλέφωνο του ή η διεύθυνση κατοικίας του σε ένα τέτοιο σύστημα και στο διαδίκτυο γενικότερα. Παρόλα αυτά υπάρχουν διάφορα εμπορικά συστήματα τα οποία χρησιμοποιούν τις πιστωτικές κάρτες των χρηστών για παραγγελίες στο διαδίκτυο τα οποία ξεπερνούν το παραπάνω πρόβλημα.

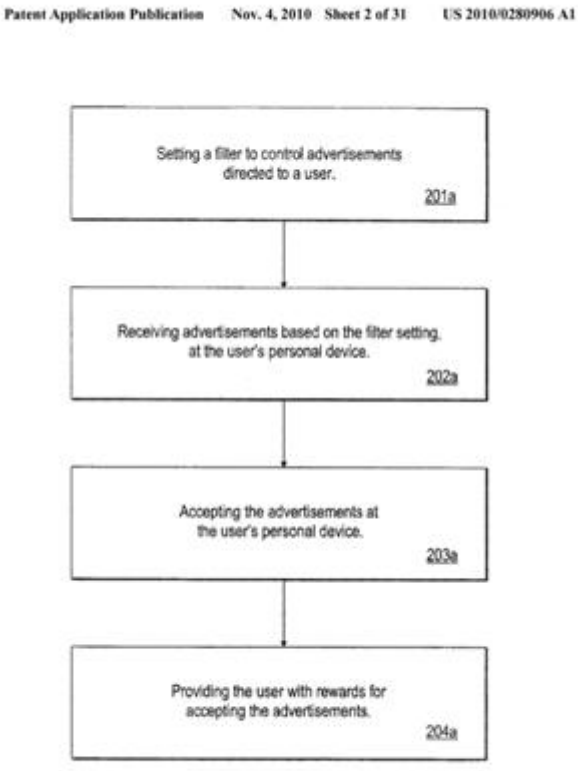


Σχήμα 8
παράδειγμα demographic συστήματος
Πηγή: http://www.reelseo.com/ooyala-custom-video-analytics

Ένα κύριο πρόβλημα του συστήματος αυτού είναι η αναξιοπιστία των συστάσεων οι οποίες βασίζονται μόνο στα δημογραφικά χαρακτηριστικά του κάθε χρήστη τα οποία δεν μπορούν πάντα να αναλύσουν επακριβώς τα ενδιαφέροντα του. Επιπλέον στο σύστημα αυτό υπάρχει η αδυναμία ελέγχου των αλλαγών των προτιμήσεων των χρηστών με το πέρασμα του χρόνου. Με άλλα λόγια ο κάθε χρήστης δεσμεύεται με τις αρχικές προτιμήσεις του λόγω της στατικότητας των μεταβλητών με βάση των οποίων το σύστημα κάνει συστάσεις.

3.3.4 Economic Filtering

Αυτού του είδους τα φίλτρα σε ένα σύστημα συστάσεων ειδικεύονται στο φιλτράρισμα πληροφορίας μεταξύ δύο χρηστών οι οποίοι επικοινωνούν μεταξύ τους. Τα συστήματα που χρησιμοποιούν τα φίλτρα αυτά εισάγουν την έννοια του κόστους ανάλογα με το μέγεθος της σύστασης υποθέτοντας ότι ο ένας χρήστης συστήνει κάποιο αντικείμενο σε κάποιον άλλο. Η τεχνική αυτή δεν εφαρμόζεται σε πολλά συστήματα για το λόγω ότι οι χρήστες πρέπει να αποκτούν κίνητρα για να κάνουν μόνοι τους συστάσεις και όχι να αποτρέπονται από αυτές. Κίνητρα για τη μορφή των συστάσεων αυτών μεταξύ των χρηστών μπορεί να είναι κάποια μορφή αμοιβής για παράδειγμα ένα είδος ηλεκτρονικού χρήματος για πληρωμή και επιβράβευση όπως χρησιμοποιεί το σύστημα Knowledge Pump το οποίο καλεί την αμοιβή αυτή "chit".(Glance, N., Arregui, D. And Dardenne, M. 1999).



Σχήμα 9
Παράδειγμα Economic Συστήματος

Πηγή : Economic Filtering System for Delivery of Permission Based, Targeted, Incentivized Advertising
http://www.faqs.org

3.3.5 Knowledge-based recommendations

Τα συστήματα αυτά χρησιμοποιούν την εκ των προτέρων γνώση στο ‘πάντρεμα’ των αντικειμένων που συστήνονται με τις ανάγκες του κάθε χρήστη. Το κύριο πλεονέκτημα των συστημάτων αυτών είναι η αποφυγή του προβλήματος εκκίνησης (bootstrapping problem). Στα συστήματα αυτά δεν χρειάζεται κάποιος χρόνος μάθησης των αντικειμένων προκειμένου να γίνει μια καλή και αξιόπιστη σύσταση αφού τα αντικείμενα αξιολογούνται πριν εισαχθούν στο σύστημα.
Κύριο μειονέκτημα των συστημάτων που χρησιμοποιούν την τεχνική αυτή είναι η αδυναμία δημιουργίας μη τυποποιημένων συστάσεων οι οποίες δεν προσαρμόζονται κατάλληλα στον κάθε χρήστη ξεχωριστά.

3.3.6 Υβριδικές Τεχνικές

Από τις τεχνικές συστάσεων τις οποίες είδαμε στις προηγούμενες υποενότητες του κεφαλαίου αυτού παρατηρούμε ότι καθεμία από αυτές έχει τα πλεονεκτήματα και τα μειονεκτήματά της και ότι καμία δεν είναι η βέλτιστη λύση για την εξυπηρέτηση όλων των χρηστών συνολικά (Wei et al, 2005). Η ανάγκη αυτή εξυπηρέτηση ενός μεγαλύτερου αριθμού χρηστών από τα συστήματα συστάσεων οδήγησε στη δημιουργία υβριδικών συστημάτων, Τα συστήματα αυτά αποτελούνται από δύο ή περισσότερες γνωστές τεχνικές φιλτραρίσματος για την υλοποίηση τους. Παραδείγματα τέτοιων συστημάτων είναι το Quickstep και το Foxtrot. Η λογική δημιουργίας ενός τέτοιου υβριδικού συστήματος είναι η απόκτηση περισσότερων πλεονεκτημάτων και παράλληλα λιγότερων μειονεκτημάτων από τις τεχνικές τις οποίες αποτελείται με σκοπό τη μεγαλύτερη ακρίβεια και ποιότητα των συστάσεων (Schafer et al, 2000). Οι τεχνικές από τις οποίες αποτελείται ένα τέτοιο σύστημα μπορεί να ποικίλουν. Ένα παράδειγμα ενός υβριδικού συστήματος μπορεί να χρησιμοποιεί και collaborative και content-based filtering για να βρει

ομοιότητες μεταξύ των χρηστών και να προτείνει συστάσεις με βάση αυτές. Το πλεονέκτημα του συστήματος αυτού είναι το χαμηλό κόστος και η προσπάθεια εκτέλεσής του καθώς και η δυνατότητα εύκολης προσαρμογής του στις απαιτήσεις των χρηστών. Τα υβριδικά συστήματα προσφέρουν την δυνατότητα εναλλαγής ανάμεσα στις τεχνικές που το αποτελούν με σκοπό την καλύτερη εξυπηρέτηση του χρήστη. Για παράδειγμα το υβριδικό σύστημα συστάσεων DailyLearner πετυχαίνει να ξεπεράσει το cold-start problem χρησιμοποιώντας αρχικά τη μέθοδο content-based και στη συνέχεια την collaborative μέθοδο.(Billsus et al, 2000)

Η λειτουργία του συστήματος αυτού είναι η συλλογή επιμέρους συστάσεων από τις τεχνικές τις οποίες αποτελείται και στην συνέχεια η ανάλυσή τους για την δημιουργία πιο ολοκληρωμένων συστάσεων. Έτσι όταν έχουμε μεγάλο αριθμό επιμέρους συστάσεων μια ολοκληρωμένη και ακριβής σύσταση μπορεί εύκολα να επιτευχθεί. Τέτοια συστήματα εφαρμόζονται εύκολα στην πράξη γιατί δεν χρειάζεται να ολοκληρωθεί μέχρι τέλους καθεμία από τις επιμέρους τεχνικές που το αποτελούν.

Μια άλλη παρατήρηση πάνω στα υβριδικά αυτά συστήματα είναι η εξής: εφόσον αποτελούνται από δύο και πάνω τεχνικές είναι δυνατόν τα δεδομένα ή η πληροφορία η οποία χρησιμοποιείται σαν είσοδος σε μια τεχνική να είναι έξοδος από μια άλλη. Για παράδειγμα αν χρησιμοποιήσουμε το collaborative filtering σε ένα hybrid σύστημα και παραχθούν συστάσεις από αυτό των οποίων η ποιότητα και η ακρίβεια είναι χαμηλή, τότε μπορεί να δεχθεί σαν είσοδο τις συστάσεις αυτές μια content-based filtering τεχνική και να φιλτράρει τις ανεπιθύμητες συστάσεις.

Μια επιπλέον εφαρμογή είναι η χρησιμοποίηση όχι μόνο των συστάσεων που παράγονται από μια συγκεκριμένη τεχνική π.χ. content-based αλλά η χρησιμοποίηση ολόκληρης της τεχνικής αυτής σαν είσοδο σε μια άλλη π.χ. collaborative.

Πλεονέκτημα της εφαρμογής αυτής είναι η λύση του sparsity problem. Γενικά τα υβριδικά συστήματα κληρονομούν τα πλεονεκτήματα των τεχνικών από τις οποίες αποτελούνται. Πολλές φορές όμως κληρονομούν τα μειονεκτήματα και τους περιορισμούς τους οποίους μπορεί να έχουν οι επιμέρους αυτές τεχνικές. Για τον λόγο αυτό στην δημιουργία τέτοιον υβριδικών συστημάτων συστάσεων πρέπει να είμαστε αρκετά προσεκτικοί.

3.4 Μέθοδοι και τεχνικές μοντελοποίησης του προφίλ των χρηστών

Η μοντελοποίηση του προφίλ του χρήστη έχει ως σκοπό την διαχείρηση μεγάλου όγκου πληροφοριών έτσι ώστε να προσφέρει στον χρήστη μόνο εκείνες τις πληροφορίες που τον αφορούν. Η μοντελοποίηση αυτή επιτυγχάνεται με συστήματα τα οποία μπορούν να διαχειριστούν μεγάλο πλήθος δεδομένων, βασίζουν την λειτουργία τους στο προφίλ του χρήστη και έχουν ως κύριο χαρακτηριστικό τους την αποτροπή στον χρήστη μη χρήσιμων γι' αυτόν πληροφοριών, τις οποίες αυτόματα απομακρύνουν.
Αυτό επιτυγχάνεται με τέσσερις παράγοντες.

1. Την αρχική λειτουργιας τους βάση της οποίας διαχωρίζονται σε ενεργά (active) και παθητικά (passive). Τα ενεργά συστήματα δρουν αυτόματα και πραγματοποιούν την αναζήτηση σχετικών με το προφίλ των χρηστών πληροφοριών. Ένα κύριο μειονέκτημά τους είναι η μεταφορά στον χρήστη και ανεπιθύμητων πληροφοριών λόγω της μη ορθής δημιουργίας του προφίλ του. Τα παθητικά συστήματα δεν εμφανίζουν στον χρήστη πληροφορίες τις οποίες δεν χρειάζεται με βάση το προφίλ του. Υπάρχουν συστήματα που φιλτράρουν όλες τις πληροφορίες ενώ άλλα που τις εμφανίζουν όλες με σειρά προτεραιότητας ανάλογα με την χρησιμότητα τους.
2. Την λειτουργία τους. Διαχωρίζονται σε information source, filtering servers και users sites. Στην πρώτη περίπτωση ο χρήστης αποστέλει το προφίλ και τα ερωτήματά του σε έναν παροχέα πληροφοριών με αποτέλεσμα την τροφοδότηση πληροφοριών που συμβαδίζουν με αυτά που έστειλε ο χρήστης. Στη δεύτερη περίπτωση η υλοποίηση γίνεται σε ενδιάμεσους servers. Όπως και στην προηγούμενη περίπτωση ο χρήστης αποστέλει προφίλ και ερωτήματα αλλά αυτή τη φορά στον server ο οποίος με την σειρά του φιλτράρει και στέλνει πίσω τα αποτελέσματα. Στην τελευταία περίπτωση έχουμε την αξιολόγηση των εισερχόμενων δεδομένων από το τοπικό υποσύστημα φιλτραρίσματος και την απομάκρυνση των τυχόν αδιάφορων προς το προφίλ του χρήστη πληροφοριών. Παράλληλα εμφανίζονται με σειρά σχετικότητας τα δεδομένα που ταιριάζουν στο προφίλ του χρήστη.
3. Τη μεθοδολογία φιλτραρίσματος βάση της οποίας διαχωρίζονται σε γνωστικά (cognitive) και κοινωνικά (social). Στην πρώτη περίπτωση το φιλτράρισμα πραγματοποιείται σύμφωνα με το περιεχόμενο της πληροφορίας αλλά και βάσει του γνωστικού υπόβαθρου του χρήστη, την προσωπικότητα και τα πλάνα του. Στη δεύτερη περίπτωση το φιλτράρισμα

στηρίζεται στα ιδιαίτερα χαρακτηριστικά και ιδιότητες του χρήστη. Τα συστήματα που χρησιμοποιούν collaborative filtering διαδικασία προτείνουν μια πληροφορία σε ένα χρήστη με τη λογική ότι έχει παρόμοιες συνήθειες και παρόμοιο προφίλ με κάποιον άλλο χρήστη.

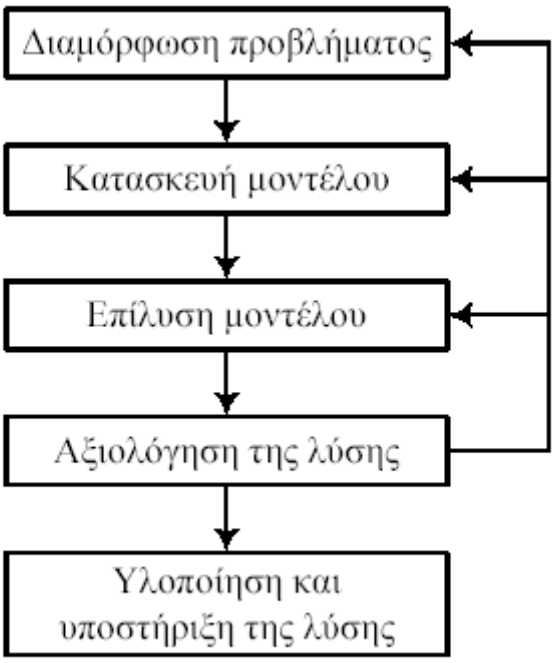
4. Τη μεθοδολογία απόκτησης γνώσης για τους χρήστες βάση της οποίας διαχωρίζονται σε άμεσα (implicit) και έμμεσα (explicit) ή και τα δύο μαζί. Στην πρώτη περίπτωση η απόκτηση της γνώσης ολοκληρώνεται κυρίως από την πραγματοποίηση μιας μικρής συνέντευξης (user interrogation) με τον χρήστη η οποία είναι με την μορφή ερωτηματολογίου το οποίο πρέπει να συμπληρώσει ο χρήστης. Ένας άλλος τρόπος είναι η προβολή προκαθορισμένων προφίλ από τα οποία ο χρήστης πρέπει να επιλέξει αυτά που του ταιρι΄ζουν περισσότερο. Στα explicit η απόκτηση γνώσης επιτυγχάνεται με την καταγραφή αντιδράσεων του χρήστη κατά την παρουσίαση πληροφοριών με σκοπό την εκπαίδευση του συστήματος για την σχετικότητα παρόμοιων πληροφοριών. Στην τελευταία περίπτωση έχουμε δύο προσεγγίσεις. Την document space στην οποία δημιουργείται ένα δείγμα πληροφοριών που ο χρήστης το έχει προκαθορίσει. Κάθε εισερχόμενη πληροφορία ελέγχεται για τυχόν ομοιότητες με το δείγμα του χρήστη. Αν η ομοιότητα ξεπερνά ένα προκαθορισμένο κατώφλι σχετικότητας τότε θεωρείται σχετική και εγκρίνεται. Η επικινδυνότητα της μεθόδου αυτής αφορά το πόσο το δείγμα του χρήστη είναι επαρκές ή όχι. Η δεύτερη προσέγγιση είναι η Stereotypes Inference στην οποία ο χρήστης καλείται να δώσει άμεσες πληροφορίες για τον εαυτό του για να μπορέσει το σύστημα να συσχετίσει τους χρήστες με προκαθορισμένα στερεότυπα τα οποία εκφράζουν δεδομένες πληροφορίες για ομάδες χρηστών.Η χρήση στερεοτύπων είναι πολύ συνηθισμένη στο user modeling.

3.5 Πολυκριτηριακή ανάλυση αποφάσεων

Η διαπίστωση ότι η επίλυση πολύπλοκων και σημαντικών προβλημάτων λήψης αποφάσεων δεν μπορεί να πραγματοποιείται μέσω μίας μονόπλευρης και μονοδιάστατης ανάλυσης, οδήγησε στην ανάπτυξη και διάδοση της πολυκριτήριας ανάλυσης αποφάσεων (multicriteria decision aid, MCDA, ή multicriteria decision making, MCDM). Η πολυκριτήρια ανάλυση αποφάσεων έχει ως βασικό αντικείμενο την αντιμετώπιση ενός προβλήματος που παρουσιάζεται κατά την προσπάθεια εξέτασης όλων των παραμέτρων ενός προβλήματος και των κριτηρίων που επηρεάζουν

τη λήψη της κατάλληλης απόφασης. Το πρόβλημα αυτό αφορά τον τρόπο με τον οποίο μπορεί να γίνει η σύνθεση όλων των παραμέτρων ώστε να επιτευχθεί η λήψη ορθολογικών αποφάσεων. Βασικό χαρακτηριστικό και σημαντική διαφορά της πολυκριτήριας ανάλυσης από άλλες εναλλακτικές προσεγγίσεις είναι ότι η αναγκαία σύνθεση πραγματοποιείται υπό το πρίσμα της πολιτικής λήψης των αποφάσεων και του συστήματος προτιμήσεων και αξιών, το οποίο χρησιμοποιείται συνειδητά ή ασυνείδητα από αυτόν που αποφασίζει. Έτσι, η πολυκριτήρια ανάλυση αποφάσεων ενσωματώνει τον αποφασίζοντα και τις προτιμήσεις του στη διαδικασία ανάπτυξης των υποδειγμάτων, χωρίς να προσδίδει απλά στον αποφασίζοντα έναν παθητικό ρόλο, που τον περιορίζει στην παρακολούθηση και εφαρμογή των αποτελεσμάτων μαθηματικών υποδειγμάτων. Όπως φαίνεται, λοιπόν, η πολυκριτήρια ανάλυση ενδιαφέρεται ιδιαίτερα για την εξέταση θεμάτων που αφορούν την ανάλυση, μαθηματική μοντελοποίηση και αναπαράσταση των προτιμήσεων που διέπουν την πολιτική λήψης αποφάσεων από την πλευρά αυτού που αποφασίζει. Βασικός στόχος της ανάλυσης είναι η παροχή των αναγκαίων πληροφοριών για την υποστήριξη της διαδικασίας λήψης των αποφάσεων, συμβάλλοντας στον εντοπισμό των βασικών χαρακτηριστικών του εξεταζόμενου προβλήματος και των ιδιαιτεροτήτων των διαθέσιμων εναλλακτικών λύσεων.

Το μεθοδολογικό πλαίσιο της επιχειρησιακής έρευνας στην ‘παραδοσιακή’ του μορφή φαίνεται στο Σχήμα 10.



Σχήμα 10

Στο πρώτο στάδιο πραγματοποιείται η διαμόρφωση του προβλήματος, που περιλαμβάνει τον καθορισμό των μεταβλητών απόφασης (decision variables), τον προσδιορισμό του στόχου του προβλήματος (objective) και τον προσδιορισμό του χώρου των εφικτών λύσεων (feasible solutions). Το δεύτερο στάδιο αφορά την κατασκευή του κατάλληλου μοντέλου που περιγράφει το πρόβλημα. Ως μοντέλο ορίζεται η μαθηματική αναπαράσταση του προβλήματος που αποτυπώνει όλες τις μεταβλητές απόφασης, τους στόχους και τους περιορισμούς, και η κατασκευή του βασίζεται σε κάποιες υποθέσεις, οι οποίες όσο πιο ρεαλιστικές είναι τόσο αυξάνεται η πιθανότητα το μοντέλο να συμβάλει με επιτυχία στην αντιμετώπιση του προβλήματος που ερευνάται. Το τρίτο στάδιο, αφορά την επίλυση του μοντέλου με την κατάλληλη μαθηματική μέθοδο έτσι ώστε να προσδιοριστούν οι τιμές των μεταβλητών απόφασης οι οποίες αντιστοιχούν σε μία εφικτή λύση που βελτιστοποιεί τον στόχο του προβλήματος. Η επόμενη φάση ερευνά την ποιότητα της λύσης (ευαισθησία, ευστάθεια κλπ) σε συνάρτηση με τις παραμέτρους του μοντέλου, τις υποθέσεις και των δεδομένων του προβλήματος. Τέλος,

στο τελευταίο στάδιο περιλαμβάνεται η υλοποίηση της λύσης και η υποστήριξη – αιτιολόγηση της σε περίπτωση που χρειαστεί .

Η λήψη αποφάσεων με το ‘παραδοσιακό’ μεθοδολογικό πλαίσιο παρουσιάζει διάφορα προβλήματα, βασικότερα χαρακτηριστικά των οποίων είναι :

Η ύπαρξη πολλαπλών κριτηρίων οδηγεί σε αντικρουόμενα αποτελέσματα, καθώς η επιλογή που θεωρείται ως βέλτιστη με βάση ένα κριτήριο δεν είναι απαραίτητα βέλτιστη και σε σχέση με τα υπόλοιπα κριτήρια της ανάλυσης.

Δεδομένης της αντικρουόμενης φύσης των κριτηρίων δεν είναι δυνατός ο εντοπισμός μίας βέλτιστης λύσης.

Η επιλογή της κατάλληλης λύσης είναι υποκειμενική και βασίζεται στην πολιτική λήψης αποφάσεων που ακολουθεί αυτός που αποφασίζει. Αναλυτική μελέτη πολυκριτηριακών μεθόδων λήψης αποφάσεων,(Σταμάτης Κ. Σπανός)

3.5.1 Κύρια χαρακτηριστικά των πολυκριτηριακών μεθόδων λήψης αποφάσεων

Η πολυκριτηριακή ανάλυση δημιουργεί σχέσεις προτίμησης μεταξύ των επιλογών μέσω αναφοράς σε ένα εκτεταμένο σύνολο στόχων τους οποίους έχει εντοπίσει ο λήπτης αποφάσεων και για τους οποίους έχει ορίσει μετρήσιμα κριτήρια για να αξιολογήσει το βαθμό στον οποίο οι στόχοι αυτοί έχουν επιτευχθεί. Σε απλές συνθήκες, η διαδικασία εντοπισμού των στόχων και των κριτηρίων μπορεί να παρέχει από μόνη της αρκετές πληροφορίες για τους λήπτες αποφάσεων. Ωστόσο, όταν απαιτείται ένα επίπεδο λεπτομέρειας αρκετά συναφές με την ανάλυση κόστους-κέρδους (CBA), η πολυκριτηριακή ανάλυση προσφέρει μία πληθώρα τρόπων συγκέντρωσης των στοιχείων για κάθε κριτήριο ξεχωριστά ώστε να παρέχει δείκτες για τη συνολική απόδοση των επιλογών.

Ένα κύριο χαρακτηριστικό της πολυκριτηριακής ανάλυσης είναι η έμφαση που δίνεται στην κρίση της ομάδας των ληπτών απόφασης για τον καθορισμό των στόχων και των κριτηρίων, εκτιμώντας παράλληλα και τη σχετική σημαντικότητα των βαρών καθώς και για την κριτική της συνεισφοράς κάθε επιλογής σε κάθε κριτήριο απόδοσης. Η υποκειμενικότητα που χαρακτηρίζει τη διαδικασία αυτή μπορεί να προκαλέσει κάποια ανησυχία. Η βάση της κατά κύριο λόγο αποτελείται από τις προσωπικές επιλογές στόχων, κριτηρίων, βαρών και αξιολογήσεων εκπλήρωσης των στόχων των ληπτών αποφάσεων, παρόλο που «αντικειμενικά» στοιχεία, όπως οι τιμές που έχουν παρατηρηθεί, μπορούν επίσης να συμπεριληφθούν. Η πολυκριτηριακή ανάλυση, ωστόσο, προσφέρει ένα βαθμό δομής, ανάλυσης και ανοίγματος σε κατηγορίες αποφάσεων οι οποίες εμπίπτουν πέραν της πρακτικής εφαρμογής της .

Η πολυκριτηριακή ανάλυση μπορεί να ορισθεί ως μία συστηματική και μαθηματικά τυποποιημένη προσπάθεια επίλυσης προβλημάτων που προκύπτουν από αντικρουόμενους στόχους. Η ικανοποίηση των στόχων αυτών δεν μπορεί να είναι πλήρης. Οι διαθέσιμες επιλογές σε ένα τέτοιο πρόβλημα παρουσιάζουν άριστη επίδοση μόνο ως προς έναν ή περισσότερους – αλλά ποτέ ως προς όλους – τους στόχους, γιατί τότε δε θα υπήρχε πρόβλημα απόφασης: η επιλογή που θα ικανοποιούσε μια τέτοια συνθήκη θα ήταν η άριστη. Είναι αναγκαίος λοιπόν ένας συμβιβασμός μεταξύ των αλληλοσυγκρουόμενων στόχων. Πρέπει δηλαδή ο υπεύθυνος για τη λήψη της απόφασης να επιλέξει τον ή τους στόχους, τους οποίους επιθυμεί να μεγιστοποιήσει, καθώς και τις αντισταθμιστικές απώλειες που είναι διατεθειμένος να αποδεχθεί ως προς τους υπόλοιπους στόχους. Η έννοια του συμβιβασμού και κατ᾽ επέκταση της συμβιβαστικής λύσης – σε αντιδιαστολή προς την άριστη λύση – δηλώνει το χαρακτήρα των αποφάσεων – λύσεων, που αναζητούνται στα πολυκριτηριακά προβλήματα. Οι λύσεις αυτές είναι άριστες μόνο κατά την άποψη του ατόμου που αποφασίζει για την επιλογή.

Η επιστημονική περιοχή της πολυκριτηριακής ανάλυσης περιλαμβάνει κατ᾽ αρχήν ένα θεωρητικό υπόβαθρο, στο οποίο αναπτύσσεται η βασική λογική για την προσέγγιση τέτοιου είδους προβλημάτων. Ακόμη προσδιορίζονται τα κύρια δομικά στοιχεία του προβλήματος και

αναλύονται οι βασικές τους ιδιότητες. Με βάση αυτό το θεωρητικό υπόβαθρο έχει αναπτυχθεί ένα πλήθος τεχνικών, κατάλληλων για την αντιμετώπιση ενός μεγάλου εύρους προβλημάτων που προκύπτουν στην πράξη. Αν και η ταξινόμηση των τεχνικών αυτών σε ιδιαίτερες κατηγορίες δεν είναι αυστηρή, διακρίνονται τρεις βασικές ομάδες μεθόδων:

- Πολυκριτηριακή ιεράρχηση επιλογών
- Πολυκριτηριακός μαθηματικός προγραμματισμός
- Πολυκριτηριακή θεωρία χρησιμότητας

Το βασικό στοιχείο που διαφοροποιεί τις δύο πρώτες κατηγορίες είναι το είδος του συνόλου των επιλογών. Συγκεκριμένα, η πρώτη κατηγορία εφαρμόζεται σε προβλήματα που εξετάζουν ένα πεπερασμένο σύνολο διακριτών επιλογών, ενώ η δεύτερη σε προβλήματα με συνεχές σύνολο άπειρου αριθμού επιλογών, στα οποία κατ' αναλογία με τα προβλήματα γραμμικού μονοκριτηριακού προγραμματισμού, οι μεταβλητές απόφασης μπορεί να παίρνουν οποιαδήποτε τιμή εντός ενός καθορισμένου πεδίου. Τέλος, η τρίτη κατηγορία μεθόδων εφαρμόζεται και σε συνεχές και σε διακριτό σύνολο επιλογών και στηρίζεται στη λογική της αναγωγής του πολυκριτηριακού σε μονοκριτηριακό πρόβλημα μέσω του προσδιορισμού μιας συνολικής συνάρτησης χρησιμότητας που συνθέτει τις επιμέρους (ανά κριτήριο) προτιμήσεις του αποφασίζοντα σε ένα ενιαίο μέτρο με βάση το οποίο προχωράει στη λήψη της απόφασης.

Όσον αφορά στην ταυτοποίηση προβλημάτων πολυκριτηριακής ανάλυσης επισημαίνεται το εξής: Κάθε πρόβλημα προσδιορίζεται από ορισμένα δομικά χαρακτηριστικά, που απορρέουν είτε από την ίδια τη φύση του προβλήματος είτε από τις απόψεις και τις προτιμήσεις του αποφασίζοντα. Η ταυτοποίηση του αντικειμένου της πολυκριτηριακής ανάλυσης ως προς τα χαρακτηριστικά αυτά αποτελεί ένα πρώτο στάδιο της αναλυτικής διαδικασίας, που διευκολύνει την κατανόηση του προβλήματος και επιτρέπει την επιλογή της κατάλληλης μεθόδου επίλυσης. Ιδιαίτερη έμφαση δίνεται:

- <u>Στο στάδιο δόμησης του προβλήματος:</u>
  - καθορισμός του προβλήματος και επιλογή των πιθανών εναλλακτικών σεναρίων,

- επιλογή των κριτηρίων,
- μέτρηση των επιδόσεων και ταξινόμηση των κριτηρίων,
- εκτίμηση της βαρύτητας του κάθε κριτηρίου,
- δημιουργία του μοντέλου αξιολόγησης,
- καθορισμός των πιθανών περιοριστικών παραμέτρων ανάλογα με το αντικείμενο του εξεταζόμενου προβλήματος,
- τελική ταξινόμηση των εξεταζόμενων σεναρίων κατά σειρά βαθμολογίας με βάση τα χαρακτηριστικά του μοντέλου που θα επιλεχθεί (το σενάριο με την υψηλότερη βαθμολογία αντιστοιχεί στην ευνοϊκότερη περίπτωση).

♦ <u>Στο στάδιο ανάλυσης των αποτελεσμάτων:</u>
  - ανάλυση ευαισθησίας της λύσης,
  - προσδιορισμός της σύγκρουσης των κριτηρίων.

♦ Το μαθηματικό μοντέλο υποβοηθά τον αποφασίζοντα στην αναζήτηση της βέλτιστης λύσης και στην καλύτερη κατανόηση της διαδικασίας και των συνεπειών της απόφασής του.

Ορισμένα χαρακτηριστικά σημεία που πρέπει να αναφερθούν σε σχέση με το πρόβλημα είναι τα εξής:

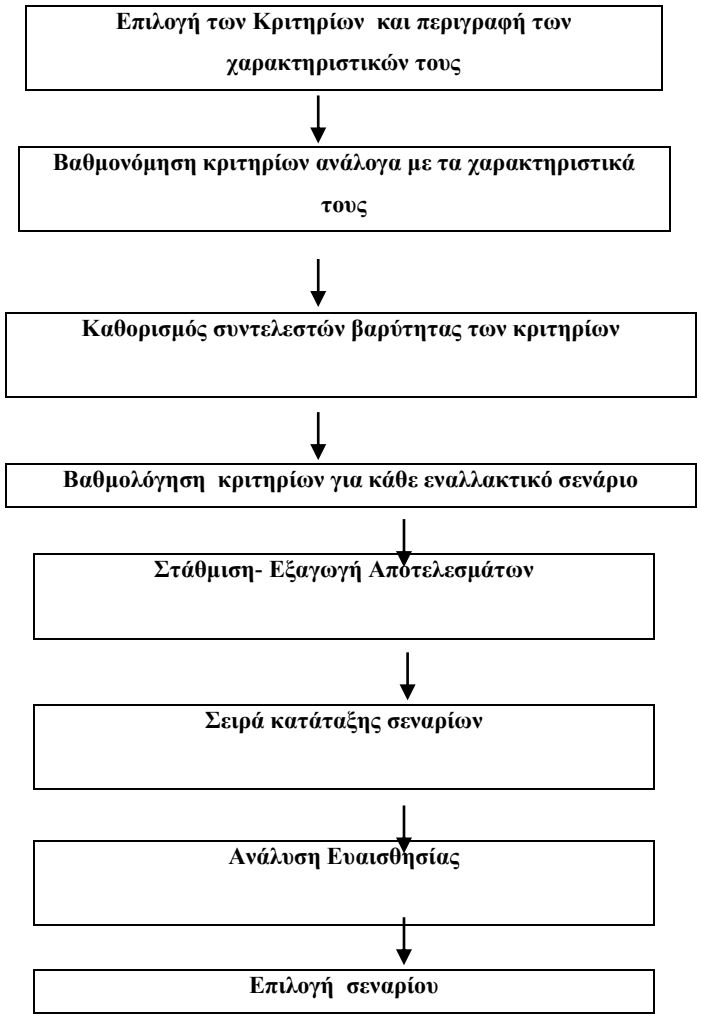
♦ Τα βασικά στοιχεία του προβλήματος είναι η μήτρα αξιολόγησης που περιλαμβάνει ένα σύνολο διακριτών επιλογών, ένα σύνολο κριτηρίων αξιολόγησης και την επίδοση της κάθε επιλογής στο αντίστοιχο κριτήριο και το σύστημα προτιμήσεων του αποφασίζοντα που εμπεριέχει τη σχετική βαρύτητα των κριτηρίων, την κατεύθυνση προτίμησης των επιδόσεων (ελάχιστο ή μέγιστο) και τα όρια ανοχής.

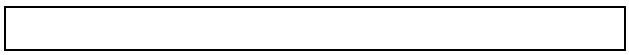
♦ Το ζητούμενο από την επίλυση του προβλήματος είναι:

- ο προσδιορισμός της σχετικά βέλτιστης λύσης,
- η ιεράρχηση του συνόλου των λύσεων,
- η ταξινόμηση των λύσεων σε ομάδες.

♦ Η μέθοδος επίλυσης του προβλήματος:

- μέθοδοι σύνθεσης των επιδόσεων: αναγωγή σε μονοκριτηριακό πρόβλημα, όπου το ένα κριτήριο εκφράζει τη συνολική χρησιμότητα της επιλογής,
- μέθοδοι ιεράρχησης των επιλογών: δυαδική σύγκριση των επιλογών σε κάθε κριτήριο και διατύπωση σχέσεων επικράτησης.





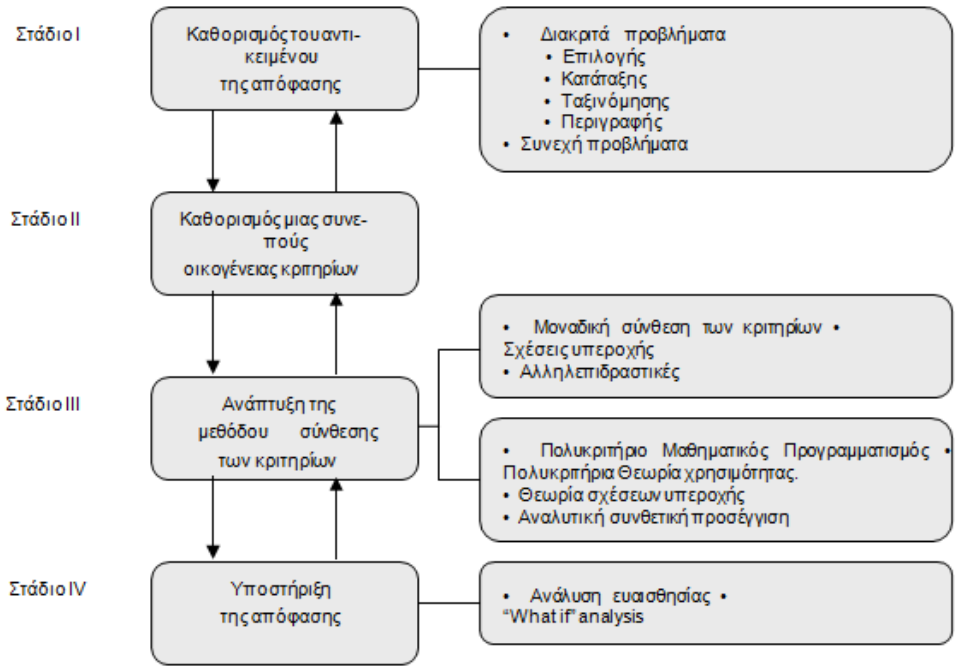
Σχήμα 11

3.5.2 Μεθοδολογικό πλαίσιο της πολυκριτήριας ανάλυσης αποφάσεων

Για να πάρει κάποιος μια απόφαση πρέπει κατ'αρχήν να γνωρίζει το πρόβλημα, την ανάγκη και το σκοπό της απόφασης, τα κριτήρια αλλά και τα υποκριτήρια της απόφασης, τους παράγοντες που επηρεάζονται από την απόφασή μας, αλλά και τις εναλλακτικές ενέργειες που έχουμε στη διάθεσή μας. Κύριο αντικείμενο της πολυκριτήριας ανάλυσης αποφάσεων και κοινό στοιχείο όλων των μεθοδολογικών προσεγγίσεων του χώρου αυτού είναι η ανάπτυξη και χρήση υποδειγμάτων σύνθεσης όλων των βασικών παραμέτρων ενός προβλήματος, έτσι ώστε να υποστηριχθεί ο αποφασίζοντας στη λήψη ορθολογικών αποφάσεων στη βάση του συστήματος αξιών και προτιμήσεων που τον διέπει. Η επίτευξη αυτού του στόχου είναι μια ιδιαίτερα περίπλοκη και σύνθετη διαδικασία, η οποία δεν οδηγεί σε βέλτιστες λύσεις και αποφάσεις αλλά σε ικανοποιητικές λύσεις οι οποίες ανταποκρίνονται στη γενικότερη πολιτική που ακολουθεί ο αποφασίζων. Παρά τις διαφορές που υπάρχουν ανάμεσα στις διάφορες μεθόδους και τεχνικές της πολυκριτήριας ανάλυσης αποφάσεων, τα βασικά συστατικά τους είναι πολύ απλά: ένα πεπερασμένο ή άπειρο σύνολο ενεργειών (εναλλακτικές, λύσεις, σχέδια δράσεις, κλπ.), τουλάχιστον δύο κριτήρια και προφανώς τουλάχιστον ένας αποφασίζων. Λαμβάνοντας υπόψη τα βασικά αυτά συστατικά η πολυκριτήρια ανάλυση αποφάσεων είναι μια μεθοδολογία η οποία βοηθάει τη λήψη αποφάσεων κυρίως σε όρους επιλογής, αξιολόγησης και ταξινόμησης των διαφόρων ενεργειών.

Ένας από τους θεμελιωτές της σύγχρονης θεωρίας της πολυκριτήριας ανάλυσης, ο Bernard Roy (1985), προκειμένου να οριοθετήσει σε βάθος το σύνολο των δραστηριοτήτων του αναλυτή, παρουσίασε στα μέσα της δεκαετίας του 1970 ένα γενικό μεθοδολογικό πλαίσιο αντιμετώπισης πολυδιάστατων προβλημάτων λήψης αποφάσεων. Το πλαίσιο αντιμετώπισης πολυδιάστατων προβλημάτων λήψης αποφάσεων. Το πλαίσιο αυτό ουσιαστικά αποτελεί τη ραχοκοκαλιά κάθε πολυκριτήριας προσέγγισης, χαρακτηρίζει απόλυτα τη φιλοσοφία όλων των μεθοδολογιών του χώρου, και

στα τριάντα και πλέον χρόνια ζωής του, αντιμετώπισε προβλήματα του μάνατζμεντ από τα πιο απλά ως τα πλέον σύνθετα.
Το μοντέλο το Roy όπως φαίνεται και στο Σχήμα 12 αποτελείται από τέσσερα διαδοχικά αλλά αλληλεπιδρώντα στάδια, και η υλοποίηση της μεθοδολογίας στα προβλήματα δεν γίνεται απαραίτητα μέσα από μια γραμμική διαδικασία, αφού ο αναλυτής, διατρέχοντας τα στάδια είναι δυνατό να διαπιστώσει έλλειψη πληροφόρησης ή λάθη τα οποία μπορεί να διορθώσει ανατρέχοντας σε προηγούμενα στάδια. Παρακάτω εξετάζονται όλα τα στάδια αναλυτικότερα, το καθένα ξεχωριστά.



Σχήμα 12

ΣΤΑΔΙΟ Ι : Αντικείμενο της απόφασης.
Στο στάδιο αυτό, είναι απαραίτητο να εντοπιστεί και να οριστεί το σύνολο Α των εφικτών λύσεων και δυνατών δραστηριοτήτων (actions) ή αλλιώς αποφάσεων και παράλληλα να καθοριστεί το αντικείμενο του προβλήματος. Το σύνολο Α μπορεί να είναι συνεχές ή διακριτό. Στην πρώτη περίπτωση, το σύνολο Α των εφικτών λύσεων, ορίζεται έμμεσα από μαθηματικές σχέσεις (γραμμικές ανισοεξισώσεις) που έχουν το ρόλο των περιορισμών, ως υπερπολύεδρο

του πραγματικού χώρου l διαστάσεων (όσες και οι μεταβλητές απόφασης). Αντίθετα στην περίπτωση που το σύνολο Α είναι διακριτό, θεωρείται ότι υπάρχει ένα σαφές σύνολο εναλλακτικών δραστηριοτήτων, οι οποίες αφού καταγραφούν μπορούν να αναλυθούν ώστε να ληφθεί η κατάλληλη απόφαση. Με τον εντοπισμό του συνόλου Α καθορίζεται και το αντικείμενο της απόφασης, δηλαδή ο τρόπος με τον οποίο θα πρέπει να εξεταστούν οι εναλλακτικές δραστηριότητες ώστε το αποτέλεσμα της ανάλυσης να απαντά με σαφήνεια στο εξεταζόμενο πρόβλημα. Η εξέταση των εναλλακτικών δραστηριοτήτων μπορεί να πραγματοποιηθεί με μια εκ των ακόλουθων τεσσάρων προβληματικών :

Προβληματική α (επιλογή choice): Η προβληματική τύπου α αναφέρεται στην επιλογή μίας ή περισσοτέρων εναλλακτικών του συνόλου Α οι οποίες θεωρούνται ως οι πλέον κατάλληλες. Για παράδειγμα, κατά την χωροθέτηση ενός εργοστασίου η προβληματική αφορά την επιλογή της πλέον κατάλληλης τοποθεσίας.

- Προβληματική β (ταξινόμηση, classification/sorting): Η προβληματική τύπου β αναφέρεται στην ταξινόμηση των εναλλακτικών δραστηριοτήτων σε προκαθορισμένες ομοιογενείς κατηγορίες. Για παράδειγμα κατά την αξιολόγηση μιας αίτησης αδείας, το αντικείμενο της ανάλυσης αφορά την αξιολόγηση του αιτούντα και την ταξινόμησή του είτε στην κατηγορία των αποδεκτών αιτήσεων είτε στην κατηγορία των απορριπτέων αιτήσεων.
- Προβληματική γ (κατάταξη, ranking): Η προβληματική τύπου γ αναφέρεται στην κατάταξη των εναλλακτικών δραστηριοτήτων από τις καλύτερες προς τις χειρότερες. Για παράδειγμα στη F1 η σειρά εκκίνησης των μονοθέσιων στον Κυριακάτικο αγώνα καθορίζεται από τα αποτελέσματα των προκριματικών ή αλλιώς των κατατακτήριων δοκιμών του Σαββάτου. Στην περίπτωση αυτή απαιτείται η κατάταξη των μονοθέσιων βάσει του χρόνου που έχουν επιτύχει στα προκριματικά.
- Προβληματική δ (περιγραφή, description): Η προβληματική τύπου δ αναφέρεται στην περιγραφή των εναλλακτικών δραστηριοτήτων βάσει των επιδόσεών τους στα επιμέρους κριτήρια αξιολόγησης.Την επιλογή της καταλληλότερης προβληματικής, την καθορίζει η φύση του προβλήματος που εξετάζεται, χωρίς όμως η προβληματική που έχει υιοθετηθεί να παραμένει αναγκαστικά σταθερή σε όλη τη διάρκεια της διαδικασίας απόφασης, αφού μπορεί να μεταβάλλεται ανάλογα με την πολυπλοκότητα του προβλήματος. Αυτό σημαίνει ότι σε ορισμένες

περιπτώσεις, πιθανόν να απαιτείται ο συνδυασμός δύο προβληματικών για την καλύτερη αντιμετώπιση του προβλήματος. Για παράδειγμα, στην περίπτωση αξιολόγησης υποψηφίων για μια θέση εργασίας σε μια εταιρία, όπου ο αριθμός των υποψηφίων είναι μεγάλος, αρχικά μπορεί να εφαρμοστεί η προβληματική β , για το χωρισμό των υποψηφίων σε εκείνους που δεν πληρούν και σε εκείνους που πληρούν κάποια ελάχιστα κριτήρια για τους οποίους θα εφαρμοστεί, σε δεύτερη φάση, η προβληματική α για την τελική επιλογή.

ΣΤΑΔΙΟ ΙΙ: Συνεπής Οικογένεια Κριτηρίων
οι παράγοντες οι οποίοι επιδρούν στο αποτέλεσμα της ανάλυσης των εναλλακτικών δραστηριοτήτων του συνόλου Α , εντοπίζονται στο δεύτερο στάδιο της διαδικασίας. Σύμφωνα με την πολυκριτήρια ανάλυση αποφάσεων, κριτήριο θεωρείται κάθε παράγντας που επιδρά στη λήψη μιας απόφασης. Ως κριτήριο ορίζεται κάθε πραγματική συνάρτηση g η οποία αποτυπώνει τη συμπεριφορά των εναλλακτικών δραστηριοτήτων σε έναν πραγματικό αριθμό, έτσι ώστε για οποιεσδήποτε δύο εναλλακτικές δραστηριότητες x και y να ισχύουν :

$g(x) \succ g(y) \Leftrightarrow x \succ y$ (η x προτιμάται της y)
$g(x) = g(y) \Leftrightarrow x \sim y$ (η x είναι αδιάφορη της y)

Αυτό που κάνει την έννοια του κριτηρίου να διαφέρει από την έννοια του χαρακτηριστικού (attribute) που χρησιμοποιείται από άλλες μεθοδολογικές προσεγγίσεις, είναι αυτές ακριβώς οι δύο παραπάνω ιδιότητες. Το χαρακτηριστικό δεν εκφράζει καμιά σμπεριφορά προτίμησης όπως αυτή αναπαρίσταται μέσω των παραπάνω δύο βασικών σχέσεων. Το γεγονός ότι η αριθμητική περιγραφή μιας εναλλακτικής δραστηριότητας x σε ένα χαρακτηριστικό είναι μεγαλύτερη σε σχέση με την αντίστοιχη αριθμητική περιγραφή μιας άλλης εναλλακτικής δραστηριότητας y δεν σημαίνει με κανέναν τρόπο ότι η x υπερέχει της y.Αποτέλεσμα των διαδικασιών σε αυτό το στάδιο της διαδικασίας ανάλυσης ενός προβλήματος, είναι η κατασκευή ενός συστήματος κριτηρίων G = {g1,g2,...,gn} πουονομάζεται «συνεπής οικογένεια κριτηρίων» (consistent family of criteria), και το οποίοσύστημα διαθέτει τις ακόλουθες βασικές ιδιότητες:

1. Συνέπεια ή μονοτονία (monotonicity / cohesiveness): Ένα σύνολο κριτηρίων θεωρείται ότι διαθέτει την ιδιότητα της μονοτονίας εάν και μόνο εάν για κάθε ζεύγος εναλλακτικών x και y για τις οποίες υπάρχει

κάποιο κριτήριο gi∈ G τέτοιο ώστε $g_i(x) \succ g_i(y)$ και $g_j(x) = g_j(y)$ για κάθε $j \neq i$, συμπεραίνεται ότι $x \succ y$.

2. Επάρκεια (exhaustivity): Ένα σύνολο κριτηρίων θεωρείται ότι διαθέτει την ιδιότητα της επάρκειας εάν και μόνο εάν για κάθε ζεύγος εναλλακτικών x και y τέτοιες ώστε gi(x) = gi(y) για κάθε κριτήριο gi∈ G , συμπεραίνεται ότι $x \sim y$.

3. Μη πλεονασμός (non-redundancy): Ένα σύνολο κριτηρίων θεωρείται ότι διαθέτει την ιδιότητα του μη πλεονασμού εάν και μόνο εάν η διαγραφή ενός οποιουδήποτε κριτηρίου οδηγεί σε παραβίαση των ιδιοτήτων της μονοτονίας ή της επάρκειας.

ΣΤΑΔΙΟ ΙΙΙ: Μοντέλο ολικής προτίμησης

Στο τρίτο στάδιο της διαδικασίας ανάλυσης ενός προβλήματος και αφού πλέον έχει καθοριστεί το σύνολο των κριτηρίων, καθορίζεται η μορφή του υποδείγματος σύνθεσης των κριτηρίων βάσει του οποίου θα αντιμετωπιστεί το αντικείμενο του προβλήματος, όπως αυτό καθορίστηκε στο πρώτο στάδιο (επιλογή, κατάταξη, ταξινόμηση, περιγρφή). Συνήθως ένα μοντέλο ολικής προτίμησης αποτελεί τον κανόνα σύνθεσης των κρι τηρίων, των μοντέλων μερικής προτίμησης δηλαδή. Οι δράσεις του συνόλου A συγκρίνονται λοιπόν συνολικά με βάση το μοντέλο αυτό και τον τύπο προβληματικής που έχει οριστεί στο στάδιο Ι. Στο στάδιο ΙΙΙ, ο αναλυτής πρέπει να καθορίσει μια μέθοδο πολυκριτήριας σύνθεσης η οποία θα επιτρέψει τη σύγκριση των δράσεων του συνόλου A , λαμβάνοντας υπόψη συνολικά όλες τις τιμές των δράσεων πάνω στα κριτήρια της συνεπούς οικογένειας κριτηρίων. Σε γενικές γραμμές πάντως, τα μοντέλα σύνθεσης πολλαπλών κριτηρίων χωρίζονται σε δύο κατηγορίες :

- Αντισταθμιστικά μοντέλα (compensatory models): Πρόκειται για μοντέλα στα οποία η υποβάθμιση ενός κριτηρίου είναι δυνατό να αποζημιωθεί από την βελτίωση της τιμής ενός άλλου κριτηρίου.
- Μη αντισταθμιστικά μοντέλα (non compensatory models): Πρόκειται για μοντέλα στα οποία η αντιστάθμιση ενός κριτηρίου από ένα άλλο δεν είναι επιτρεπτή.

Οι κυριότερες κατηγορίες πολυκριτήριων μεθόδων, από θεωρητικής απόψεως είναι τρεις:

1. Συναρτησιακές μέθοδοι: Η σύνθεση των κριτηρίων επιτυγχάνεται μέσω μιας ή περισσοτέρων συναρτήσεων αξίας ή χρησιμότητας.
2. Σχεσιακές μέθοδοι: Η σύνθεση των κριτηρίων επιτυγχάνεται μέσω μιας ή περισσοτέρων σχέσεων υπεροχής.
3. Αναλυτικές μέθοδοι: Το μοντέλο σύνθεσης των κριτηρίων συμπεραίνεται έμμεσα από δεδομένα ολικής προτίμησης του αποφασίζοντος.
4. Στις παραπάνω κατηγορίες εντάσσονται επίσης και τα μοντέλα αποφάσεων υπό αβεβαιότητα με ένα ή πολλαπλά κριτήρια απόφασης.

ΣΤΑΔΙΟ IV: Υποστήριξη της απόφασης
Στο στάδιο αυτό της διαδικασίας λαμβάνουν χώρα όλες εκείνες οι δραστηριότητες οι οποίες θα βοηθήσουν τον αποφασίζοντα να κατανοήσει τα αποτελέσματα του υποδείγματος σύνθεσης των κριτηρίων που καθορίστηκε στο τρίτο στάδιο καθώς και τη διαδικασία με την οποία εξάχθηκαν τα αποτελέσματα αυτά. Ο αναλυτής του προβλήματος είναι σε θέση να υλοποιήσει με επιτυχία τα αποτελέσματα της ανάλυσης, να αναζητήσει και να οργανώσει τα στοιχεία απάντησης σε συγκεκριμένα ερωτήματα που θέτουν ή ενδέχεται να θέσουν κάποιοι εμπλεκόμενοι στην διαδικασία της απόφασης και κυρίως ο αποφασίζων.
Το τέταρτο αυτό στάδιο, πρόκειται για συμπληρωματικό στάδιο του προηγουμένου, του οποίου ο κυριότερος λόγος ύπαρξης οφείλεται στο γεγονός ότι μια λύση που δίνει ένα μοντέλο δεν είναι άμεσα εκμεταλλεύσιμη στα πεδία λήψης αποφάσεων. Οι τεχνικές που συμβάλλουν στην αρτιότερη υποστήριξη ή τεκμηρίωση διαφόρων επιλογών εξαρτώνται κάθε φορά από το μοντέλο ολικής προτίμησης το οποίο έχει επιλεγεί στο στάδιο ΙΙΙ.

3.6 Μέθοδος Maut (Multi-Attribute Utility Theory)

Η μέθοδος MAUT χρησιμοποιείται για να βοηθήσει τους λήπτες αποφάσεων να αποκτήσουν διορατικότητα στις αποφάσεις (πχ

παράγοντες και προτεραιότητες). Η μέθοδος δεν αποσκοπεί στην ανακάλυψη ή την απόδειξη της ¨αλήθειας¨.

Η μέθοδος MAUT ως μαθηματικό μοντέλο

Τα βασικά στοιχεία μίας πολυκριτήριας μεθόδου περιλαμβάνουν :

-Μία αριθμητική τιμή της συνολικής χρησιμότητας μίας επιλογής

-Βάρη καθορισμένα σε μεμονωμένα χαρακτηριστικά

-Μέτρα της απόδοσης των επιλογών έναντι των χαρακτηριστικών

-Ένα προσθετικό κανόνα που να περικλείει όλα τα μέτρα απόδοσης

Έτσι, είναι:

$$U_Y = \sum_i w_i u_{i,Y}$$

όπου U_Y είναι η συνολική χρησιμότητα (ή τιμή) του προϊόντος Y, Σ ο προσθετικός κανόνας (που δεν είναι πάντοτε ένα άθροισμα), w_i το βάρος του χαρακτηριστικού i, και $u_{i,Y}$ η χρησιμότητα του προϊόντος Y σε σχέση με το i. Η U_Y είναι στην ουσία η συνάρτηση που υπολογίζει την περιοχή που ¨ταιριάζει¨ στα κριτήρια αξιολόγησης. Αυτή η κεντρική ιδέα έχει πολλές παραλλαγές.

Όσον αφορά την χρησιμότητα, αυτή για να αθροιστεί πρέπει πρώτα να προσδιοριστεί ποσοτικά:

Πιο συχνά μία απλή συνάρτηση από μέτρα χρησιμότητας επαρκεί

Η διαδικασία αυτή βασίζεται στην κρίση του καθενός

Πειθαρχία στην απόκτηση των μέτρων των προϊόντων και συνεπής χρήση των συναρτήσεων μετατροπής μπορεί να αποτρέψει την ομοιότητα της λογικής. (Wallnau,1998)

3.7 Μέθοδος Mavt (Multiattribute Value Theory)

Η μέθοδος MAVT είναι η πιο ευρέως χρησιμοποιούμενη προσέγγιση για την λύση πολυκριτηριακά προβλήματα κατάταξης. Το βασικό μοντέλο MAVT παρουσιάζεται παρακάτω:

$$V(X_j) = \sum_{i=1}^{n} w_i v_i(x_{ij})$$

όπου V(Xj) είναι η συνολική συνάρτηση προσθετικής αξίας για το υποψήφια εναλλακτική j, w_i το βάρος που καθορίζεται για το κριτήριο I, v_i η (Μία λεπτομερής ανάλυση για την πολυκριτηριακή λήψη αποφάσεων βρίσκεται στο Perlack et al., Prototype Framework for R&D Decisions, Working Draft, Oak Ridge National Laboratory, December 1994) συνάρτηση που χαρακτηρίζει το xi, όπου xi είναι το μέτρο του χαρακτηριστικού I για την εναλλακτική j, και n ο αριθμός των κριτηρίων.

Το μοντέλο γίνεται λειτουργικό με τα ακόλουθα βήματα :

-Ορισμός υποψήφιων εναλλακτικών και κριτηρίων απόφασης (ιεραρχία)

-Αξιολόγηση κάθε εναλλακτικής ξεχωριστά για κάθε κριτήριο

-Καθορισμός βαρών για τα κριτήρια

Άθροισμα των βαρών των κριτηρίων και των αξιολογήσεων των εναλλακτικών για κάθε κριτήριο ξεχωριστά, έτσι ώστε να αποκτηθεί ένα συνολικό μέτρο της τιμής ή της αξίας (πχ συνάρτηση προσθετικής αξίας)

Διεξαγωγή αναλύσεων ευαισθησίας

Κατάταξη των εναλλακτικών και καθορισμός του πιθανού μεγέθους αγοράς

Για παράδειγμα, στην έρευνα Determination of the Potential Market Size and Opportunities for Biomass-to-Electricity Projects in China που έγινε από τον Perlack (1995), τέσσερα κριτήρια χρησιμοποιούνται :

Ο εσωτερικός συντελεστής απόδοσης

Ένας δείκτης πηγών βιομάζας

Ένας δείκτης που να εκφράζει την ανάγκη για ενέργεια

Μία ποιοτική εκτίμηση των τοπικών δυνατοτήτων για οργάνωση και εφαρμογή της εναλλακτικής.

Όταν κατάλληλες συναρτήσεις κλίμακας αναπτυχθούν και κάθε εναλλακτική αξιολογηθεί, καθορίζονται τα βάρη για τα κριτήρια. Η αρχική κατάταξη που προκύπτει μπορεί να αλλάξει με αλλαγή και καθαρισμό των δεικτών των κριτηρίων. Επίσης, οι κατατάξεις είναι ευαίσθητες στην επιλογή των βαρών. Διαφορετικά βάρη θα οδηγήσουν σε διαφορετικές κατατάξεις.

3.8 Μέθοδος Uta (Utility Theory Additive)

Η μέθοδος UTA, όπως αρχικά προτάθηκε από τους Jacquet-Lagrèze και Sisko (1978, 1982), βασίζεται σε ένα μοντέλο προγραμματισμού στόχων για να εκτιμήσει μία προσθετική, μη γραμμική συνάρτηση χρησιμότητας από μία κατάταξη προτίμησης των εναλλακτικών.

Το πρόβλημα είναι η σύγκριση, κατάταξη και αποτίμηση ενός συνόλου πράξεων, ή η επιλογή εναλλακτικών, σχετικά προς N κριτήρια που μετρούν την χρησιμότητα αυτών των εναλλακτικών. Οι μετρήσεις αυτών των εναλλακτικών δίνονται από το διάνυσμα g(a) = (g1(a),g2(a),...,gN(a)) για κάθε εναλλακτική a που ανήκει στο A. Για παράδειγμα, για μία εναλλακτική για αυτοκινητόδρομο, τα gi(a) θα μπορούσαν να είναι η αναλογία κόστους-οφέλους, η επίδρασή του στην ασφάλεια, στο περιβάλλον κλπ.

Το μοντέλο υποθέτει την ύπαρξη μίας συνάρτησης χρησιμότητας :

U(g(a))=U(g1(a),g2(a),...,gN(a))

η οποία ικανοποιεί τα κλασσικά αξιώματα της θεωρίας αποφάσεων, που είναι ονομαστικά τα αξιώματα της συγκρισιμότητας, αντανακλαστικότητας, μεταβατικότητας των επιλογών, συνέχειας και αυστηρής υπεροχής.

Η συνάρτηση χρησιμότητας είναι προσθετική,

$$U(g(a)) = \sum_{i=1}^{N} u_i(g_i(a))$$

με $u_i(g_i) \geq 0 et \frac{du_i}{dg_i} > 0.$

Η προσθετική συνάρτηση υποδηλώνει συγκεκριμένα ότι η μερική χρησιμότητα ενός κριτηρίου $u_i(g_i(a))$ εξαρτάται μόνο στο επίπεδο αυτού του συγκεκριμένου κριτηρίου. Αυτή η συνάρτηση παρέχει ένα άθροισμα των κριτηρίων σε ένα κοινό δείκτη ώστε να συγκρίνει και να αποτιμήσει τις εναλλακτικές που λαμβάνονται υπόψη.

Κατατάσσει την εναλλακτική σε μία ολοκληρωμένη ¨αδύναμη¨ σειρά R : εάν το P υποδεικνύει μία αυστηρή προτίμηση και I την αδιαφορία μεταξύ δύο εναλλακτικών a και b, τότε :

$U[g(a)] > U[g(b)] \Leftrightarrow a\ P\ b$

$U[g(a)] = U[g(b)] \Leftrightarrow a\ I\ b$

Σύμφωνα με τους Jacquet-Lagrèze και Sisko(1978, 1982), η μέθοδος UTA(α) εκτιμάει την συνάρτηση U στο σύνολο αναφερόμενων εναλλακτικών A, με βάση την μέθοδο γραμμικού προγραμματισμού στόχων. Η μέθοδος γραμμικού προγραμματισμού στόχων προτάθηκε από τους Charnes και Cooper (1961, 1977), και παρέχει μία προσέγγιση μίας μη γραμμικής συνάρτησης από γραμμικά διαστήματα.

Για να εφαρμοστεί αυτή η μέθοδος, το πεδίο παρέκκλισης κάθε κριτηρίου $[g_{i^*}, g_i^*]$, ορισμένο από την λιγότερο ικανοποιητική τιμή του κριτηρίου (g_{i^*}) και την καλύτερη τιμή του (g_i^*), διαιρείται σε α_i ίσα διαστήματα $[g_i^j, g_i^{j+1}]$. Οι μεταβλητές που εκτιμούνται από τη μέθοδο είναι οι μερικές χρησιμότητες σε αυτά τα όρια, έστω $u_i(g_i^j)$. Η χρησιμότητα σε μέσες τιμές των κριτηρίων δίνονται από γραμμική παρεμβολή. Κατά συνέπεια, για $g_i(a) \in [g_i^j, g_i^{j+1}]$ είναι:

$$u_i[g_i(a)] = u_i(g_i^j) + \frac{g_i(a) - g_i^j}{g_i^{j+1} - g_i^j}[u_i(g_i^{j+1}) - u_i(g_i^j)].$$

Για κάθε ζεύγος εναλλακτικών (a, b) που ανήκουν στο Α΄, ο λήπτης αποφάσεων πρέπει να εκφράσει τις προτιμήσεις του ή την αδιαφορία του. Όπως αναφέρεται στη μελέτη A multi-criteria analysis of stated preferences among freight transport alternatives (2003), σύμφωνα με την εκδοχή που προτάθηκε από τους Despotis et al. (1990), η ονομαζόμενη UTASTAR εκδοχή, τα αποτελέσματα αυτών των συγκρίσεων εισάγονται ως περιορισμοί σύμφωνα με τις ακόλουθες συνθήκες:

$$\sum_{i=1}^{N}\{u_i(g_i(a)) - u_i(g_i(b))\} + \sigma^+(a) - \sigma^-(a) - \sigma^+(b) + \sigma^-(b) \geq \delta \Leftrightarrow a\mathrm{P}b$$

$$\sum_{i=1}^{N}\{u_i(g_i(a)) - u_i(g_i(b))\} + \sigma^+(a) - \sigma^-(a) - \sigma^+(b) + \sigma^-(b) = 0 \Leftrightarrow a\mathrm{I}b$$

με όλα τα σ^+ και $\sigma^- \geq 0$.

Το σ^+ αντιπροσωπεύει σε ένα θετικό λάθος σε σχέση με τη διαφορά μεταξύ των επιπέδων χρησιμότητας, ενώ το σ^- υποδεικνύει ένα αρνητικό λάθος. Αυτά τα λάθη είναι όλα μη αρνητικά, αντιπροσωπεύουν τα πιθανά λάθη μίας εκτίμησης της χρησιμότητας μίας πράξης. Η αντικειμενική χρησιμότητα F που ελαχιστοποιείται είναι το σύνολο αυτών των λαθών :

$$F = \sum_{a \in A'} \sigma^+(a) + \sigma^-(a)$$

Η παράμετρος δ που χρησιμοποιήθηκε παραπάνω πρέπει να είναι αυστηρά θετική. Η τιμή του μπορεί να επηρεάσει την λύση του προγράμματος. Ως εκ τούτου, σύμφωνα με τους Beuthe και Scannella (2001), δεν πρέπει να πάρει αρχικά υψηλή τιμή. Η υπόθεση ότι οι μερικές

χρησιμότητες αυξάνουν με την τιμή των κριτηρίων, επιβάλλει μία σειρά από επιπρόσθετους περιορισμούς:

$u_i\,(g_i^{j+1} - u_i(g_i^j)) \geq s_i,$ j=1,2,…,α_i, i=1,2,..,N

όπου si πρέπει να είναι (αυστηρά) θετικός. Όπως για το δ, είναι καλύτερο να δοθεί αρχικά μία μικρή τιμή.

Τέλος, οι μερικές χρησιμότητες ομαλοποιούνται με βάση τις συνθήκες:

$$\sum_{i=1}^{N} u_i(g_i^*) = 1$$

και

$$u_i(g_{i^*}) = 0 \forall i$$

Η πρώτη εξίσωση δείχνει ότι οι τιμές των $u_i(g_i^*)$, οι χρησιμότητες των κριτηρίων στα υψηλότερα επίπεδα τους αντιπροσωπεύουν τα σχετικά βάρη των κριτηρίων στη συνάρτηση χρησιμότητας.

Βάζοντας μαζί όλα τα παραπάνω,καταλήγουμε:

$$\text{MIN F} = \sum_{a \in A'} [\sigma^+(a) + \sigma^-(a)]$$

και

$\sigma^+(a) \geq 0, \sigma^-(a) \geq 0 \ \geq 0 \ \forall a \in A$

$u_i(g_i^j \geq 0 \ \forall i, \forall j$

όπου χρησιμοποιούνται για να υπολογιστούν οι χρησιμότητες των gi(a) μεταξύ δύο διαδοχικών ορίων. Αυτή είναι το βασικό UTA-UTASTAR μοντέλο. Siskos, Y., E. Grigoroudis, N.F. Matsatsinis (2005), UTA methods, in: J. Figueira, S. Greco, M. Ehrgott (eds.), Multiple Criteria Decision Analysis, - State of the Art - Surveys, International Series in Operations Research and Management Science, pp. 297-344, Springer,.Περισσότερες λεπτομέρειες καθώς και για συγκριτικές προσομοιώσεις μπορούν να βρεθούν στο Beuthe και Scannella (2001). Μερικές από αυτές τις εξειδικεύσεις περιλαμβάνουν επιπρόσθετους περιορισμούς για χειρισμό περισσότερων πληροφοριών που μπορεί να δοθούν από τον λήπτη αποφάσεων.

Μπορεί να έχουμε δύο τύπους λύσεων: είτε όλα τα λάθη έχουν μηδενικές τιμές και F= 0, είτε κάποια λάθη είναι θετικά και F > 0. Στην δεύτερη περίπτωση, δεν υπάρχει μία μη γραμμική προσθετική συνάρτηση χρησιμότητας που να παρουσιάζει τέλεια τις προτιμήσεις που εκφράζονται από των λήπτη αποφάσεων. Εάν αποκλείσουμε την περίπτωση ενός λήπτη αποφάσεων ο οποίος δεν θα μπορούσε να συγκρίνει με λογική τα σχέδια ή θα ήταν ανορθολογικός με την έννοια της παρουσίασης αμετάβατων προτιμήσεων, η παρουσία των λαθών μπορεί να υποδεικνύει ότι οι προτιμήσεις του λήπτη αποφάσεων χαρακτηρίζονται από μερικές χρησιμότητες οι οποίες δεν είναι ανεξάρτητες μεταξύ τους ή δεν αυξάνονται μονότονα. Όμως, μπορεί να υπάρχει και η πιο απλή περίπτωση όπου τα επιλεγμένα διαστήματα θα έπρεπε να είναι πιο πολυάριθμα ή να οριστούν με άλλο τρόπο.

Ο καθορισμός μίας προσθετικής συνάρτησης και η προέλευση της από ξεχωριστές εκτιμήσεις των μερικών συναρτήσεων, προϋποθέτει μία προνομιακή ανεξαρτησία. Αυτό σημαίνει ότι, εάν δύο εναλλακτικές χαρακτηρίζονται από τις ίδιες τιμές για κάποια κριτήρια, οι προτιμήσεις μεταξύ τους εξαρτώνται μόνο από τις τιμές που παίρνουν από τα άλλα κριτήρια. Το κατά πόσο αυτή η υπόθεση είναι αποδεκτή σε πρακτικές εφαρμογές , διαφέρει ανάλογα με την κάθε περίπτωση. Οι Von Winterfeldt και Edwards (1973) έχουν την άποψη ότι μία προσθετική συνάρτηση μπορεί να χρησιμοποιηθεί ως μία καλή προσέγγιση. Πράγματι, μία προσθετική συνάρτηση μη γραμμικών συναρτήσεων μερικής χρησιμότητας είναι ένας ευέλικτος προσδιορισμός και μπορεί να παρέχει μία εκτίμηση που λαμβάνει υπόψη έναν ορισμένο βαθμό ανεξαρτησίας μεταξύ των κριτηρίων. Ο Stewart (1995) εμπειρικά απέδειξε ότι ήταν όντως ένας ισχυρός προσδιορισμός. Επιπλέον, με τη χρήση ενός συνόλου προσομοιώσεων. Οι Beuthe και Scannella (2001) έδειξαν ότι με το μοντέλο UTA είναι δυνατόν να αποκτηθούν χρήσιμα αποτελέσματα με το F ίσο ή κοντά στο 0, ακόμα και στην περίπτωση ανεξαρτησίας μεταξύ των κριτηρίων.

Όταν το F ισούται με μηδέν ή όχι, η λύση του προγράμματος μπορεί να μην είναι μοναδική, όπως συχνά συμβαίνει στην περίπτωση του γραμμικού προγραμματισμού. Αυτό το πρόβλημα πρέπει να λυθεί, λοιπόν, με μία post-optimality ανάλυση των αποτελεσμάτων. Οι Jacquet-Lagrèze και Siskos (1978) πρότειναν απλά να χρησιμοποιηθεί μία συνάρτηση η οποία είναι ο μέσος όρος των ακραίων βέλτιστων συναρτήσεων που αποκτήθηκαν από την ανάλυση ευαισθησίας που εφαρμόστηκε στα τελευταία όρια κάθε κριτηρίου. Αυτή η ανάλυση ευαισθησίας γίνεται με ένα επιπρόσθετο περιορισμό

$$\sum_{a \in A'} \sigma(a) \leq F^* + \theta$$

όπου θ είναι ένας μικρός θετικός αριθμός. Αυτή η διαδικασία παρουσιάστηκε για να δώσει μία πρακτική και αποτελεσματική μέθοδο εκτίμησης, και ένα παράδειγμα της χρήσης της βρίσκεται στην ανάλυση A multi-criteria analysis of stated preferences among freight transport alternatives (2003).

3.9 Καθορισμός συντελεστών βαρύτητας

Ο βαθμός σπουδαιότητας των εφαρμοζόμενων κριτηρίων για την αξιολόγηση των διαφόρων εναλλακτικών σεναρίων καθορίζεται από το συντελεστή βαρύτητας που αποδίδεται στα κριτήρια αυτά. Ανάλογα με την περίπτωση, χρησιμοποιούνται είτε άμεσοι συντελεστές βαρύτητας είτε έμμεσοι. Οι άμεσοι συντελεστές βαρύτητας χρησιμοποιούνται στην περίπτωση που ο αριθμός των κριτηρίων είναι μικρός και είναι δυνατή η επιλογή συντελεστών βαρύτητας. Οι έμμεσοι συντελεστές βαρύτητας προσδιορίζονται με την ταξινόμηση των κριτηρίων κατά σειρά σπουδαιότητας, την απόδοση ενός συνολικού συντελεστή βαρύτητας ή ενός μέγιστου συντελεστή βαρύτητας και στη συνέχεια τον προσδιορισμό των συντελεστών βαρύτητας σε σχέση με το άθροισμα όλων των συντελεστών βαρύτητας ή σε σχέση με το μεγαλύτερο συντελεστή. Επιπλέον, είναι δυνατή η χρήση κριτηρίων, στα οποία δεν έχει αποδοθεί συντελεστής βαρύτητας.

Οι συντελεστές βαρύτητας αντικατοπτρίζουν το σύστημα αξιών και προτιμήσεων του αποφασίζοντα. Δηλαδή, ο προσδιορισμός της σπουδαιότητας του κάθε κριτηρίου βασίζεται στην ιδιαίτερη σημασία που δίνουν οι ενδιαφερόμενοι φορείς για κάθε κριτήριο. Συνεπώς, ανάλογα με το είδος του προβλήματος είναι δυνατό να παρουσιάζουν μεγαλύτερη σημασία για τους ενδιαφερόμενους φορείς τα περιβαλλοντικά κριτήρια σε σχέση με τα οικονομικά ή και το αντίστροφο. Έτσι, για τον προσδιορισμό των συντελεστών βαρύτητας απαιτείται η προσεκτική ιεραρχική ταξινόμηση των διαφόρων κριτηρίων από τους ενδιαφερόμενους φορείς.

3.10 Επιλογή του βέλτιστου σεναρίου

Έχει αναπτυχθεί μεγάλος αριθμός μεθόδων και υπολογιστικών προγραμμάτων, τα οποία είναι δυνατό να προσδιορίσουν το βέλτιστο σενάριο για κάθε διαχειριστικό πρόβλημα. Οι μέθοδοι αυτές βασίζονται στην εκτίμηση της συνολικής απόδοσης ενός σεναρίου με βάση τις επιμέρους επιδόσεις σε κάθε κριτήριο και μπορούν να ταξινομηθούν ως εξής:

1. Υπολογισμός της συνολικής προτίμησης για κάθε σενάριο. Στην περίπτωση αυτή, η επιλογή του βέλτιστου σεναρίου βασίζεται στην επιλογή του σεναρίου, που παρουσιάζει την υψηλότερη βαθμολογία ανεξάρτητα από τα επιμέρους κριτήρια.
2. Προσέγγιση της προτίμησης ενός σεναρίου σε σχέση με ένα άλλο, η οποία βασίζεται στη δοκιμή της υπόθεσης, ότι ένα σενάριο (α) είναι καλύτερο από ένα σενάριο (β), εφόσον το σενάριο (α) είναι τουλάχιστον τόσο καλό (ή όχι χειρότερο) από το σενάριο (β). Η προσέγγιση αυτή στηρίζεται στη δυαδική σύγκριση των επιλογών σε κάθε μεμονωμένο κριτήριο. Στην περίπτωση αυτή, πριν τη συγκριτική ταξινόμηση των κριτηρίων ανάλογα με τη βαθμολογία τους τίθενται κάποιοι περιοριστικοί όροι, οι οποίοι εκφράζουν την προτίμηση σε κάποια κριτήρια σε σχέση με άλλα. Με τη χρήση της μεθόδου αυτής η εύρεση του βέλτιστου σεναρίου βασίζεται εν μέρει στον προσδιορισμό της συνολικής βαθμολογίας για κάθε σενάριο και περισσότερο στη σύγκριση μεταξύ των επιμέρους σεναρίων.
3. Διαδραστική προσέγγιση, όπου τα μοντέλα, που χρησιμοποιούνται για την εκτίμηση του βέλτιστου σεναρίου, βασίζονται σε επαναληπτικές μεθόδους.

3.11 Σύστημα λήψης αποφάσεων με χρήση αθροιστικής συνάρτησης ομάδων κριτηρίων (Πολυκριτηριακή θεωρία αξίας ή χρησιμότητας - Multi – Attribute Value or Utility Theory)

Σ' αυτό το σύστημα της πολυκριτηριακής ανάλυσης η συγκριτική αξιολόγηση των εναλλακτικών σεναρίων ακολουθεί τα εξής στάδια:

1^{ο} Στάδιο: Αρχικά, γίνεται η επιλογή των κριτηρίων, τα οποία θα πρέπει να καλύπτουν όλες τις πλευρές του εξεταζόμενου προβλήματος και να μπορούν να βαθμολογηθούν σε κατάλληλη κλίμακα. Μετά, ακολουθεί η ταξινόμηση των κριτηρίων σε ομάδες. Καθεμιά από αυτές τις ομάδες χαρακτηρίζεται από ένα συντελεστή βαρύτητας, που δηλώνει το "βάρος" της στο κάθε σενάριο και προσδιορίζεται μετά από συζητήσεις με όλους τους εμπλεκόμενους φορείς, λαμβάνοντας υπόψη και δεδομένα ανάλογων περιπτώσεων. Το άθροισμα των συντελεστών αυτών θα πρέπει να είναι ίσο με 100%. Κατόπιν, βάσει των παραπάνω προκύπτει η αντίστοιχη αθροιστική συνάρτηση, η οποία θα έχει τη μορφή:

F (O)= Σ Ai *Oi
όπου:

O_i είναι οι επιμέρους ομάδες κριτηρίων

A_i είναι ο συντελεστής βαρύτητας κάθε μίας από τις ομάδες κριτηρίων O_i και το άθροισμα των συντελεστών βαρύτητας πρέπει να ισούται με 1 (100%), $\Sigma A_i = 1$

2^{ο} Στάδιο: Οι ομάδες κριτηρίων αναλύονται στα επιμέρους κριτήρια αξιολόγησης, για τα οποία επίσης καθορίζεται η σχετική σπουδαιότητά τους μέσα στην ομάδα κριτηρίων με τη βοήθεια κατάλληλων συντελεστών βαρύτητας. Το άθροισμα των συντελεστών βαρύτητας των επιμέρους κριτηρίων μέσα σε κάθε ομάδα είναι επίσης 100%.

3^{ο} Στάδιο: Πραγματοποιείται ανάλυση όλων των εναλλακτικών χαρακτηριστικών κάθε επιμέρους κριτηρίου τα οποία στη συνέχεια ποσοτικοποιούνται βάσει κλίμακας 1-10, όπου οι μικρότερες τιμές αφορούν στις δυσμενέστερες αποδόσεις των χαρακτηριστικών του

κριτηρίου και οι μεγαλύτερες τιμές στις ευνοϊκότερες (καλύπτοντας με τον τρόπο αυτό όλες τις πιθανές περιπτώσεις).

4^{ο} Στάδιο: Αρχικά γίνεται αποτύπωση των χαρακτηριστικών κάθε επιμέρους κριτηρίου για κάθε εναλλακτικό σενάριο και αφού γίνει σύγκριση τους με την κλίμακα που αναπτύσσεται στο 3^{ο} στάδιο, λαμβάνει μία συγκεκριμένη τιμή απόδοσης σε κλίμακα από 1 –10. Στη συνέχεια, οι τιμές που προκύπτουν, πολλαπλασιάζονται με το σχετικό συντελεστή βαρύτητας που έχει καθένα από τα κριτήρια σε κάθε ομάδα. Ακολούθως, προστίθενται τα αντίστοιχα γινόμενα για την κάθε ομάδα και με τον τρόπο αυτό ποσοτικοποιείται κάθε ομάδα κριτηρίων. Μετά, ο βαθμός κάθε ομάδας πολλαπλασιάζεται με τον αντίστοιχο συντελεστή βαρύτητάς της, κι έτσι προκύπτει μέσω της αθροιστικής συνάρτησης ένα μέτρο της συνολικής αποτελεσματικότητας κάθε επιλογής. Με βάση τη βαθμολογία αυτή γίνεται κατάταξη των εναλλακτικών σεναρίων, με ευνοϊκότερο, αυτό που έχει την υψηλότερη επίδοση.

3.12 Γραμμικά Πρόσθετα Μοντέλα

Εάν μπορεί να αποδειχτεί, ή λογικά να θεωρηθεί, ότι τα κριτήρια είναι με βάση την προτίμηση ανεξάρτητα μεταξύ τους και αν η αβεβαιότητα δεν έχει ληφθεί υπόψη από το μοντέλο πολυκριτηριακής ανάλυσης, τότε το απλό γραμμικό προσθετικό μοντέλο εφαρμόζεται. Το γραμμικό μοντέλο δείχνει πως οι τιμές μίας επιλογής στα διάφορα κριτήρια μπορεί να συνδυαστεί σε μία συνολική τιμή. Αυτό γίνεται πολλαπλασιάζοντας την τιμή για κάθε κριτήριο με το βάρος αυτού του κριτηρίου, και στη συνέχεια προσθέτοντας όλα τα αυτά σταθμισμένα σκορ μαζί. Παρόλα αυτά, αυτό το απλό αριθμητικό μοντέλο είναι κατάλληλο εάν τα κριτήρια είναι αμοιβαίως ανεξάρτητα με βάση την προτίμηση.

Τα μοντέλα αυτής της μορφής έχουν ένα καλά εδραιωμένο ιστορικό παροχής ισχυρής και αποτελεσματικής υποστήριξης στους λήπτες αποφάσεων για ένα ευρύ πεδίο προβλημάτων και για διάφορες περιστάσεις.

3.13 Ευρετικοί αλγόριθμοι (Heuristics)

Οι ευρετικοί αλγόριθμοι επικεντρώνονται και αυτοί στην επίλυση προβλημάτων παραγωγής. Οι συνηθισμένοι αλγόριθμοι προχωρούν στη διαδικασία επίλυσης χωρίς να γνωρίζουν εκ των προτέρων ποιά από τα μονοπάτια που ακολουθούν τους οδηγούν σε αδιέξοδο. Σε πραγματικά προβλήματα όμως, ο αριθμός των συνδυασμών που καταφθάνουν σε τερματικές καταστάσεις είναι αρκετά μεγάλος, φαινόμενο που μεταμορφώνει την επιλυσή τους σε μία χρονοβόρα διαδικασία. Για αυτό ακριβώς το λόγο χρησιμοποιούνται οι αλγόριθμοι ευρετικής αναζήτησης οι οποίοι στοχεύουν τόσο στη μείωση του χρόνου επίλυσης όσο και στη μείωση του αριθμού καταστάσεων που εξετάζει ένας αλγόριθμος. Ο Ευρετικός μηχανισμός (heuristic) είναι η τεχνική που στηρίζεται στην πληροφορία εκείνη η οποία αξιολογεί ποιο μονοπάτι δεν οδηγεί σε θεμιτό αποτέλεσμα. [Michalevicz&Fogel,1999]. Με τη χρησιμοποίηση του συγκεκριμένου μηχανισμού, ο χρόνος που ξοδεύεται μπορεί να μειωθεί σημαντικά, εφόσον δίνεται το πλεονέκτημα της δυνατότητας πρόβλεψης των καταστάσεων που δεν οδηγούν πουθενά και για αυτό τον λόγο μπορούν να κλαδευτούν. Λαμβάνοντας υπόψη την αποδοτικότητα των μηχανών, τα στοιχεία πρέπει να φορτωθούν πρώτα στη μηχανή που μπορεί να τελειώσει πιο γρήγορα. Είναι επιθυμητό να αυξηθεί η χρησιμότητα της μηχανής ή να ελαχιστοποιηθεί ο συνολικός χρόνο ροής.

Η διαδικασία Shifting Bottleneck Procedure (SBP), που δημιούργησε ο Adams το 1988 [Anant&Meeran,1998] εμπνεύστηκε από τη θεωρία των περιορισμών (Theory of constraints), είναι η επικρατέστερη τεχνική στη οποία βασίζονται οι ευρετικοί αλγόριθμοι. Η βασική ιδέα είναι η αναγωγή ενός προβλήματος m μηχανών σε επιμέρους m προβλήματα μιας μηχανής. Το κάθε υποπρόβλημα λύνεται μεμονωμένα και στη συνέχεια ανάλογα τη λύση οι αντίστοιχες μηχανές ιεραρχούνται σύμφωνα με την επίδοσή τους. Η μηχανή με την καλύτερη δυνατή λύση χαρακτηρίζεται ως «bottleneck» και έχει μεγαλύτερη προτεραιότητα από τις άλλες. Στη συνέχεια η μηχανή αυτή θεωρείται ως μηχανή αναφοράς και επομένως όλες οι προηγούμενες εργασίες επαναπροσδιορίζονται με τα νέα όμως δεδομένα. Η ίδια διαδικασία επαναλαμβάνεται και για τις υπόλοιπες μηχανές. Αξίζει να αναφερθεί ότι η I/O bottleneck διαδικασία στα παράλληλα συστήματα υπολογιστών έχει πρόσφατα ξεκινήσει να δέχεται αυξανόμενο ενδιαφέρον και το μεγαλύτερο ποσοστό της

προσοχής επικεντρώνεται στη βελτίωση της απόδοσης των I/O συσκευών χρησιμοποιώντας χαμηλού επιπέδου παραλληλισμό.

Στη συνέχεια περιγράφονται δυο διαφορετικές ευρετικές μέθοδοι οι οποίες αντιπροσωπεύουν και τις κύριες κατηγορίες ευρετικών αλγορίθμων: τους στατικούς και τους δυναμικούς ευρετικούς αλγόριθμους. Αυτό που άμεσα ενδιαφέρει είναι πως αυτοί χρησιμοποιούνται σε ένα πραγματικό πρόβλημα όπως είναι για παράδειγμα η επεξεργασία παρτίδων σε μια διαδικασία κατασκευής [Michalevicz&Fogel,1999].

3.14 Στατικοί ευρετικοί αλγόριθμοι (Static Heuristics)

Ο συγκεκριμένος αλγόριθμος χρησιμοποιεί μια τροποποίηση της γνωστού κανόνα φαινομενικής βραδύτερης περάτωσης του κόστους (Apparent Tardiness Cost Rule ATC) το οποίο αναπτύχθηκε το 1987 από τους Vepsalainen και Morton . Ο κανόνας αυτός προσαρμόστηκε από τον Mason et al το 2002 στο πρόβλημα χρονοπρογραμματισμού των μηχανών δεμάτων. Ο δείκτης του κανόνα ATC I stat , ij(t) για την εργασία i που ανήκει στην οικογένεια j και η οποία υπολογίζεται στο χρόνο t δίνεται:

$$I_{ij,stat}(t) = \frac{w_{ij}}{p_j} \exp\left(- \frac{(d_{ij} - p_j - t)^+}{k\bar{p}} \right)$$

Ως k δηλώνεται μια κλιμακόμενη παράμετρο και $\bar{p}$ είναι ο μέσος επεξεργάσιμος χρόνος των υπόλοιπων εργασιών ενώ γίνεται η συντόμευση $x^+ := max(x,0)$. Για παράδειγμα, εάν ο στόχος είναι να διαμορφωθούν οι παρτίδες σε μια διαδικασία κατασκευής, τοποθετούνται σε σειρά οι ποσότητες που περιμένουν σε κατιούσα θέση όπως ακριβώς ο αντίστοιχος δείκτης τους I stat , ij(t) προστάζει. Τότε θεωρείται ως G_{il} η αντιστάθμιση για τη ποσότητα i που ανήκει στην οικογένεια j ενώ ως d_{ij} η ημερομηνία παράδοσης μιας ποσότητας i. Τέλος ως p_j δηλώνεται ο χρόνος επεξεργασίας των παρτίδων ενώ ως r_{ij} ο χρόνος προετοιμασίας

τους. Αρχικά, λαμβάνονται οι B πρώτες ποσότητες από αυτή τη σειρά ώστε να σχηματιστεί η παρτίδα η οποία πρέπει να τεθεί σε επεξεργασία στη συνέχεια. Οι υπολογισμοί επαναλαμβάνονται αρκετές φορές για διαφορετικές τιμές της παραμέτρου k με σκοπό να βρεθεί η βέλτιστη, παίρνοντας υπόψην την ολική τιμή βραδύτητας για τις ποσότητες που αναμένουν την επεξεργασία τους. Η στρατηγική αυτή όμως δεν λαμβάνει υπόψην τυχόν μελλοντικές αφίξεις ποσοτήτων γεγονός που την καθιστά πιο αποτελεσματική σε στατικό παράλληλο περιβάλλον μηχανών, ενώ έχει αντίθετα αποτέλεσματα σε δυναμικό περιβάλλον.

3.15 Δυναμικοί ευρετικοί αλγόριθμοι (Dynamic Heuristics)

Ο δεύτερος ευρετικός αλγόριθμος, σε αντίθεση με τον πρώτο, λαμβάνει υπόψην μελλοντικές αφίξεις ποσοτήτων. Για αυτό ακριβώς τον λόγο, θεωρείται μία χρονική προθήκη (t,t+Δt). Δηλώνεται ως:

$$M(i,t,\Delta t) = \left\{ ij \mid r_{ij} \le t + \Delta t \right\}$$

το σύνολο των ποσοτήτων που είναι έτοιμα για επεξεργασία τη χρονική στιγμή t ή εκείνα που θα καταφθάσουν το χρονικό διάστημα (t,t+Δt). Συνήθως επιλέγεται ένα συγκεκριμένο τμήμα του μέσου χρόνου επεξεργασίας των αναμενόμενων ποσοτήτων, Δt . Τα στοιχεία του $M(i,t,\Delta t)$ ταξινομούνται σε κατιούσα σειρά λαμβάνοντας υπόψην τον παρακάτω δείκτη:

$$I_{ij,dyn}(t) = \frac{w_{ij}}{p_j} \exp\left(- \frac{(d_{ij} - p_j - t + (r_{ij} - t)^+)^+}{k\bar{p}} \right)$$

Επόμενο βήμα αποτελεί η λήψη των p πρώτων ποσοτήτων που μεγιστοποιούν το $M(i,t,\Delta t)$ όπου το p είναι μια σταθερή ποσότητα που καθορίζεται από τους ευρετικούς αλγόριθμους. Στο συγκεκριμένο

παράδειγμα που αναφέρθηκε παραπάνω για μια πιθανή σχηματισμένη παρτίδα ο υπολογιζόμενος δείκτης της παρτίδας είναι:

$$I_{bj1}(t) = \frac{w_{bj}}{p_j} \exp\left[- \frac{(sl + rt)^+}{k\bar{p}} \right] \min\left[\frac{n_{bj}}{B}, 1 \right]$$

όπου :

$d_{bj} = \min_{j \in bj}(d_{ij})$ η ελάχιστη ημερομηνία διάρκειας ενδιάμεσα στις εργασίες,

$r_{bj} = \max_{j \in bj}(r_{ij})$ ο μέγιστος χρόνος ετοιμασίας των εργασιών,

w_{bj} η μέση αντιστάθμιση των ποσοτήτων που σχηματίζουν τη παρτίδα,

n_{bj} ο αριθμός των ποσοτήτων σε μία παρτίδα,

ενώ ισχύουν οι συντομεύσεις $sl = d_{bj} - p_j - t$ και $rt = (r_{bj} - t)^+$.

Τέλος, ένας άλλος γνωστός ευρετικός αλγόριθμος ο οποίος αξίζει να αναφερθεί είναι ο NEH που δημιουργήθηκε από τους Nawaz, Enscore και HamCampbell. [Hoon&Dong1998] Η τεχνική αυτή ακολουθεί ένα συγκεκριμένο μοντέλο που μοιάζει σε αρκετά σημεία με τη διαδικασία Shifting Bottleneck Procedure με τη διαφορά ότι αυτός θεωρεί ότι μεγαλύτερη προτεραιότητα έχει η εργασία με τον μεγαλύτερο συνολικό χρόνο διεξαγωγής. Τα βήματα που ακολουθεί είναι τα εξής:

Κατανομή των εργασιών κατά φθίνουσα σειρά με βάση το άθροισμα του χρόνου διεργασίας στην κάθε μηχανή.

Θέτει K=2

Διαλέγει τις δυο πρώτες εργασίες από την κατανεμημένη λίστα και τις προγραμματίζει με σκοπό την ελαχιστοποίηση του συνολικού χρόνου, σαν να υπήρχαν μόνο αυτές οι δυο εργασίες.

Θέτει K=K+1

Γενικεύει την προηγούμενη λειτουργία παίρνοντας την Κ εργασία από την λίστα και την τοποθετεί σε κάθε δυνατή θέση, διαλέγοντας τελικά αυτή που ελαχιστοποιεί τον συνολικό χρόνο.

Κρατάει αυτό που προκύπτει σαν την βέλτιστη λύση μέχρι αυτό το σημείο.

Αν Κ=Ν (όπου Ν το σύνολο των εργασιών) σταματάει, αλλιώς επιστρέφει στο τρίτο βήμα.

3.16 Περιεχόμενο, Απόκτηση, Μάθηση &Ανανέωση Μοντέλου Χρήστη

3.16.1 Περιεχόμενο ενός Μοντέλου Χρήστη

Υπάρχουν τρεις προσεγγίσεις για το περιεχόμενο του μοντέλου του χρήστη: η γνωσιακή (cognitive), η κοινωνική (social) και η υβριδική (hybrid).

Γνωσιακή Προσέγγιση (cognitive)

Στην γνωσιακή (cognitive) προσέγγιση το μοντέλο του χρήστη απεικονίζει τα ενδιαφέροντά του. Στο γνωσιακό προφίλ οι τομείς ενδιαφέροντος παρουσιάζονται με την μορφή λέξεων-κλειδιά (keywords). Οι λέξεις αυτές συνήθως προκύπτουν είτε από τις ανακτώμενες πληροφορίες είτε από μια σύντομη συνέντευξη.

Η διαδικασία που ακολουθείται για την ανάκτηση και το φιλτράρισμα των πληροφοριών με τη βοήθεια ενός γνωσιακού προφίλ είναι η εξής:

Η αναπαράσταση ενός κειμένου αντλείται από το περιεχόμενό του και για το λόγο αυτό πραγματοποιείται μια συντακτική και σημασιολογική ανάλυση, η οποία συνήθως βασίζεται στη συχνότητα με την οποία οι λέξεις εμφανίζονται σε κάθε κείμενο.
Η συνάρτηση σύγκρισης συνήθως έχει ως αποτέλεσμα την

κατάταξη των κειμένων βάση της σχετικότητάς τους με το προφίλ του χρήστη. Στις περισσότερες εφαρμογές τόσο το προφίλ όσο και η αναπαράσταση του περιεχομένου ενός κειμένου έχουν την μορφή διανυσμάτων, για το λόγο αυτό για τη μέτρηση της σχετικότητας των δύο διανυσμάτων χρησιμοποιείται η γωνία που αυτά σχηματίζουν. Συνεπώς το τετράγωνο του συνημίτονου της γωνίας αυτής (το οποίο υπολογίζεται απλά με το κανονικοποιημένο εσωτερικό γινόμενο των δύο διανυσμάτων) είναι ο πιο απλός τρόπος για την κατάταξη αυτών των κειμένων βάση του προφίλ του χρήστη.
Τα αποτελέσματα του φιλτραρίσματος με τη χρήση του γνωστικού μοντέλου μπορούν να βελτιωθούν με τη στάθμιση του διανύσματος πληροφοριακών αναγκών δηλαδή με την εισαγωγή βαρών στις λέξεις που περιγράφουν τα ενδιαφέροντα του χρήστη.
Τα ενδιαφέροντα του χρήστη παρέχουν ασαφή δεδομένα για τη δημιουργία ενός ακριβούς μοντέλου και οδηγούνται στην ανεύρεση επιπλέον στοιχείων, τα οποία μπορούν να βελτιώσουν περισσότερο τα αποτελέσματα του φιλτραρίσματος. Πιο συγκεκριμένα τα δεδομένα τα οποία μπορούν να αποτελέσουν ένα πιο εμπλουτισμένο μοντέλο χρήστη μπορεί να έχουν σχέση με τις γνώσεις και την εμπειρία που έχει ο χρήστης στο συγκεκριμένο τομέα αναζήτησης, αν δηλαδή δεν κατέχει καθόλου σχετικές πληροφορίες με το θέμα αυτό ή αν ασχολείται ήδη αρκετά χρόνια και έχει αρκετά καλή γνώση του αντικειμένου, το σκοπό και τους στόχους που τον ωθούν να αναζητήσει πληροφορίες, τις προτιμήσεις του ως προς κάποιες πηγές ή κάποιους συγγραφείς είτε και άλλες παραμέτρους της αναζήτησης αντίστοιχης λογικής.

Κοινωνική Προσέγγιση (social/collaborative)

Στην κοινωνική (social ή collaborative) προσέγγιση τα μοντέλα που δημιουργούνται περιέχουν κυρίως προσωπικά στοιχεία του χρήστη. Τέτοια στοιχεία είναι η ηλικία, το εκπαιδευτικό υπόβαθρο, η επαγγελματική κατάσταση, η ύπαρξη εξειδικευμένων γνώσεων, τα γενικότερα ενδιαφέροντα, η γνώση ξένων γλωσσών κ.α. Τα στοιχεία τα οποία περιέχονται σε ένα κοινωνικό μοντέλο χρήστη εξαρτώνται σε πολύ μεγάλο βαθμό από τη φύση των αναζητούμενων πληροφοριών και συνεπώς ο αριθμός τους μπορεί να αυξάνει σημαντικά κάθε φορά.

Υβριδική Προσέγγιση (hybrid)

Στην υβριδική μέθοδο το περιεχόμενο που προφίλ προκύπτει από το συνδυασμό των περιεχομένων της γνωσιακής και της κοινωνικής προσέγγισης. Ο συνδυασμός αυτός θεωρείται αρκετά αποτελεσματικός γιατί λαμβάνει υπόψη του τόσο δεδομένα για τα ενδιαφέροντα του χρήστη που έχουν να κάνουν με το περιεχόμενο όσο και προσωπικά στοιχεία που επηρεάζουν σημαντικά τα αποτελέσματα αναζήτησης.
Μια διαφορετική προσέγγιση στη δομή του περιεχομένου του μοντέλου του χρήστη περιέχει δύο κατηγορίες δεδομένων, τα σταθερά και τα εξαρτώμενα ή δυναμικά.
Στην πρώτη κατηγορία περιλαμβάνονται δεδομένα για το χρήστη τα οποία δεν μεταβάλλονται ή οι μεταβολές τους είναι ελάχιστες με την πάροδο του χρόνου. Τέτοια δεδομένα θεωρούνται το εκπαιδευτικό υπόβαθρο του χρήστη, οι γνώσεις του, το επαγγελματικό του προφίλ κα. Τα δεδομένα αυτά για να μεταβληθούν απαιτείται άμεσα επέμβαση από το χρήστη. Συνήθως τα σταθερά δεδομένα ταυτίζονται με αυτά τα οποία περιέχονται σε ένα προφίλ που ακολουθεί την κοινωνική προσέγγιση.
Στη δεύτερη κατηγορία περιλαμβάνονται δεδομένα τα οποία έχουν σχέση με τις αναζητούμενες πληροφορίες καθώς και με τις προτιμήσεις και τα πιστεύω του χρήστη. Η ανανέωση αυτών των πληροφοριών είναι απαραίτητη για την ύπαρξη ενός συνεπούς προφίλ.

3.16.2 Τρόποι Απόκτησης του Περιεχομένου ενός Μοντέλου Χρήστη

Ο Rich (1979) οριοθέτησε τη διαφορά ανάμεσα στα άμεσα μοντέλα χρήστη, τα οποία δημιουργούνται απευθείας από το χρήστη, και στα έμμεσα, τα οποία δημιουργούνται από το σύστημα βάση της συμπεριφοράς του.
Η λογική του Rich συμπληρώνεται με την άμεση ή έμμεση δυνατότητα μεταβολής και συντήρησης των μοντέλων καθώς και τη συνύπαρξη των δύο προσεγγίσεων.

Άμεση Προσέγγιση

Κατά τη δημιουργία μοντέλων χρηστών που ακολουθούν την άμεση προσέγγιση απόκτησης δεδομένων το σύστημα αλληλεπιδρά με το

χρήστη συνήθως μέσω της διεξαγωγής μιας μικρής συνέντευξης. Η συνέντευξη αυτή τις περισσότερες φορές έχει τη μορφή ερωτηματολογίου το οποίο συμπληρώνει ο χρήστης καταγράφοντας τα ενδιαφέροντα του, και άλλες κρίσιμες παραμέτρους που μπορούν να οδηγήσουν στην δημιουργία του μοντέλου του. Ένας άλλος τρόπος της μεθόδου αυτής είναι η προβολή προκαθορισμένων προφίλ από τα οποία ο χρήστης πρέπει να επιλέξει αυτό ή σε κάποιες περιπτώσεις αυτά που του ταιριάζουν περισσότερο.

Ο χρήστης μπορεί να αλληλεπιδρά άμεσα βελτιώνοντας και προσαρμόζοντας το προφίλ του, μεταβάλλοντας τα σημεία που υπάρχουν διαφορές ή προσθέτοντας παραμέτρους που δεν υπάρχουν καθόλου. Επίσης για τα προφίλ που είναι βασισμένα σε κανόνες (rule-based) και τα οποία ακολουθούν την άμεση προσέγγιση ο καθορισμός των κανόνων φιλτραρίσματος συνήθως πραγματοποιείται με ένα εργαλείο που βοηθάει το χρήστη να καθορίσει τους κανόνες που θα δομήσουν το προφίλ του.

Η άμεση προσέγγιση ακολουθείται όταν η ακρίβεια των παρεχόμενων από το χρήστη πληροφοριών είναι σημαντικότερος παράγοντας από την ευκολία του χρήστη εφόσον η αυτόματη εξαγωγή συμπερασμάτων που προσφέρουν οι έμμεσες προσεγγίσεις δεν μπορούν να θεωρηθούν αρκετά ακριβείς.

Έμμεση Προσέγγιση

Η έμμεση προσέγγιση υλοποιείται από την εξαγωγή συμπερασμάτων μέσω της καταγραφής των αντιδράσεων του χρήστη καθώς και από ένα σύνολο άλλων παρατηρήσεων με σκοπό την εκπαίδευση του συστήματος για τη σχετικότητα των πληροφοριών. Οι (Morita and Shinoda, 1994) θεωρούν πως τα συμπεράσματα για το χρήστη μπορούν να προκύψουν από τη συμπεριφορά του (αποθήκευση, εκτύπωση, προώθηση ή ακόμα άμεση απόρριψη), ενώ οι (Stefani and Strapparava, 1999) πιστεύουν πως η ανάλυση του περιβάλλοντος του χρήστη (bookmarks, αποθηκευμένα αρχεία κ.α) μπορεί να οδηγήσει σε συμπεράσματα για το χρήστη μπορούν να προκύψουν από την ανίχνευση του περιβάλλοντος του. Η λογική της έμμεσης απόκτησης πληροφοριών στηρίζεται στο γεγονός ότι οι περισσότεροι χρήστες δεν θέλουν να απασχολούνται με την παροχή δεδομένων και αναδράσεων ακόμα και αν για το λόγο αυτό απαιτείται μια και

μόνο κίνηση.

Το μοντέλο του χρήστη ανανεώνεται με την ανίχνευση αλλαγών που προκύπτουν από την παρατήρηση. Η έμμεση εξαγωγή συμπερασμάτων χρησιμοποιείται πολύ συχνά κατά την υλοποίηση εμπορικών εφαρμογών ανάκτησης πληροφοριών οπού είναι επιθυμητή η συλλογή στοιχείων για το χρήστη χωρίς αυτός να το αντιλαμβάνεται. Γενικά όμως οι έμμεσοι μέθοδοι απόκτησης της γνώσης μπορεί να οδηγήσουν σε λανθασμένα συμπεράσματα και συνεπώς θα πρέπει να χρησιμοποιούνται ως συμπληρωματικό στοιχείο για τον εμπλουτισμό και την ανανέωση του προφίλ του χρήστη το οποίο πρέπει βασικά να προκύπτει από άμεσους μεθόδους.

Συνδυαστική Προσέγγιση

Ο συνδυασμός των δύο παραπάνω προσεγγίσεων δείχνει να είναι η πιο σωστή και έγκυρη λύση. Από τη μια ο χρήστης παρέχοντας δεδομένα εξασφαλίζει την ακρίβεια και από την άλλη δεν απασχολείται τόσο πολύ από το σύστημα. Οι πληροφορίες για τη δημιουργία των μοντέλων με αυτή την προσέγγιση μπορεί να αποκτηθούν με διάφορους τρόπους για παράδειγμα, δημιουργώντας μια ομάδα- δείγμα πληροφοριών που ο χρήστης προκαθορίζει ως σχετικές. Βάση των πληροφοριών αυτών σχηματίζεται ένα υποτυπώδες προφίλ. Στη συνέχεια κάθε νέα εισερχόμενη πληροφορία ελέγχεται για τυχόν ομοιότητες με το δείγμα. Αν η ομοιότητα που παρουσιάζουν οι πληροφορίες ξεπερνούν ένα προκαθορισμένο κατώφλι σχετικότητας θεωρούνται ενδιαφέρουσες για τον χρήστη. Βέβαια η προσέγγιση αυτή μπορεί να οδηγήσει σε σφάλμα σε περίπτωση που το δείγμα δεν είναι επαρκές. Για να είναι επαρκές ένα δείγμα θα πρέπει να καλύπτει όλους τους δυνατούς τομείς ενδιαφέροντος του χρήστη γεγονός που είναι τις περισσότερες φορές αδύνατο αποδυναμώνοντας αυτή την προσέγγιση.

Μια άλλη μέθοδος δημιουργίας προφίλ που ακολουθεί τη συνδυαστική προσέγγιση είναι η μοντελοποίηση βάση στερεοτύπων που αναφέρθηκε και στην προηγούμενη παράγραφο. Ο χρήστης καλείται να δώσει άμεσες πληροφορίες για τον εαυτό του έτσι ώστε να δημιουργηθούν ομάδες προκαθορισμένων στερεοτύπων. Τα στερεότυπα εκφράζουν δεδομένες πληροφορίες για ομάδες χρηστών. Συνεπώς στην

προσέγγιση αυτή έχουμε άμεση συλλογή πληροφοριών από το χρήστη και έμμεση κατηγοριοποίηση του σε δεδομένα στερεότυπα

3.16.3 Μάθηση και Ανανέωση των Μοντέλων

Εξαιτίας της δυσκολίας μοντελοποίησης των χρηστών καθώς και των συνεχών αλλαγών των πληροφοριακών τους αναγκών, η ύπαρξη διαδικασίας μάθησης θεωρείται απαραίτητη για τον εντοπισμό των αλλαγών στα ενδιαφέροντα τους και την ανανέωση των μοντέλων. Η έλλειψη διαδικασίας μάθησης προκαλεί ανακρίβειες και επηρεάζει σημαντικά τα αποτελέσματα του φιλτραρίσματος.
Δύο είναι τα βασικά αντικείμενα έρευνας για τη διαδικασία μάθησης και ανανέωσης των μοντέλων, ο τρόπος και η συχνότητα.

Τρόποι Μάθησης

Ανατρέχοντας στη βιβλιογραφία (Hanani et al, 2000) τρεις είναι οι τρόποι μάθησης και ανανέωσης του μοντέλου του χρήστη: η παρατήρηση, η ανάδραση και η εκπαίδευση.

Παρατήρηση

Στη μάθηση με τη μέθοδο της παρατήρησης οι καταστάσεις που προκαλούν δράσεις απομνημονεύονται. Όταν νέες καταστάσεις εμφανίζονται τότε αυτές συγκρίνονται με τις ήδη υπάρχουσες έτσι ώστε να αποφασιστεί το πεδίο δράσης ή να προταθούν κάποιες ενέργειες. Στην περίπτωση της ανάκτησης και του φιλτραρίσματος πληροφοριών το σύστημα παρατηρεί τη συμπεριφορά του χρήστη για διαφορετικά δεδομένα και μπορεί να συγκρίνει τις νέες πληροφορίες με τις ήδη γνωστές έτσι ώστε να προτείνει τις ενέργειες που πρέπει να πραγματοποιηθούν. Η μέθοδος αυτή χρησιμοποιείται συνήθως με την βοήθεια ευφυών πρακτόρων όπως στο σύστημα EUN (Newell, 1997) και στο σύστημα News Dude (Billsus & Pazzani,1999).

Ανάδραση

Στη μέθοδο αυτή ο χρήστης παρέχει ανάδραση άμεσα ή έμμεσα. Στην πρώτη περίπτωση δηλώνοντας τον τρόπο δράσης σε παρόμοιες καταστάσεις και στη δεύτερη παρέχοντας πληροφορίες στο σύστημα βαθμολογώντας ή κατατάσσοντας τα αποτελέσματα. Στην περίπτωση της ανάκτησης και του φιλτραρίσματος πληροφοριών το σύστημα μαθαίνει για την σχετικότητα των νέων δεδομένων από την ανάδραση του χρήστη που σχετίζεται με παρόμοια δεδομένα.

Εκπαίδευση

Στη μέθοδο αυτή ο χρήστης εισάγει υποθετικές καταστάσεις και δράσεις δημιουργώντας μια βάση δεδομένων με πιθανά σενάρια. Το σύστημα χρησιμοποιεί τα σενάρια αυτά για τη λήψη αποφάσεων σε μελλοντικές καταστάσεις. Στην περίπτωση της ανάκτησης και του φιλτραρίσματος πληροφοριών ο χρήστης εισάγει κατά τακτά χρονικά διαστήματα ένα σύνολο αξιολογημένων δεδομένων έτσι ώστε να ανανεωθεί το προφίλ του. Η μέθοδος αυτή μπορεί να χρησιμοποιηθεί και κατά τη δημιουργία του μοντέλου του χρήστη.

3.17 Μοντελοποίηση Χρήστη-(User Modeling)

Τα τελευταία χρόνια σημαντική ανάπτυξη έχει γνωρίσει ο τομέας της μοντελοποίησης χρήστη (User Modeling). Ως κύριο αίτιο της ανάπτυξης αυτής μπορεί να θεωρηθεί η αναδυόμενη ανάγκη για συστήματα προσωποποιημένης αναζήτησης τα οποία να μπορούν να προσαρμοσθούν στις ανάγκες του χρήστη. Πολλοί είναι οι τομείς εφαρμογής των εν λόγω συστημάτων:

- Συστήματα συμβουλευτικής
- Συστήματα παροχής βοηθητικών ενεργειών
- Συστήματα συστάσεων (recommendation systems).
- Συστήματα που προσαρμόζουν το αποτέλεσμά τους στις ιδιαίτερες απαιτήσεις του χρήστη
- Συστήματα e-commerce

•Ευφυή συστήματα αναζήτησης

Η προσωποποιημένη αναζήτηση στηρίζεται στην μοντελοποίηση του προφίλ του χρήστη η οποία είναι ουσιαστικά η διαδικασία δημιουργίας και εφαρμογής μοντέλων χρήστη. Σύμφωνα με τους Kass and Finin (1989), "τα μοντέλα χρήστη παρέχουν στα συστήματα διεπαφής των πληροφοριακών συστημάτων τις αναγκαίες πληροφορίες για να προσαρμόσουν την συμπεριφορά τους με τελικό σκοπό την υποστήριξη μεγάλου αριθμού χρηστών. Ειδικότερα, τα μοντέλα χρήστη είναι χρήσιμα στα πληροφοριακά συστήματα, τα οποία ζητούν να προσαρμόσουν την συμπεριφορά τους σε μεμονωμένους χρήστες ή να προβούν σε ενέργειες οι οποίες θα εξασφαλίσουν την επιτυχή επικοινωνία συστήματος - χρήστη".

Πολλοί επιστημονικοί κλάδοι οι οποίοι έχουν ως αντικείμενο την ανάπτυξη συστημάτων που χρησιμοποιούνται από ετερογενείς πληθυσμούς χρηστών πραγματοποιούν ερευνητικό έργο πάνω στο ζήτημα της μοντελοποίησης χρήστη.

Η μοντελοποίηση χρήστη είναι σύμφωνα με τον Kobsa (1993) "..η διαδικασία δημιουργίας μοντέλων χρηστών τα οποία παρέχουν πληροφόρηση σχετικά με τα χαρακτηριστικά του χρήστη όπως το γνωσιακό υπόβαθρό του, οι στόχοι του, τα σχέδιά του, οι προτιμήσεις του, τα ενδιαφέροντά του και οι τυχόν απορίες του".

Τα μοντέλα χρήστη μπορούν να κατασκευαστούν με δύο τρόπους. Ο πρώτος τρόπος είναι με την απόκτηση πληροφοριών μέσω κάποιων στερεότυπων και ο δεύτερος με την ανάλυση του τρόπου διεπαφής συστήματος-χρήστη στο παρόν και στο παρελθόν. Η απόκτηση γνώσης για τα ιδιαίτερα χαρακτηριστικά του χρήστη είναι μια σημαντική και δύσκολη διαδικασία με αρκετά προβλήματα:

1. Η προβληματική απόκτηση γνώσης από τα γνωσιακά μοντέλα/ βάσεις δεδομένων που χρησιμοποιούνται. Λύση στο

συγκεκριμένο πρόβλημα μπορεί να αποτελέσει η υιοθέτηση τεχνικών λήψης των κατάλληλων πληροφοριών βασισμένων στους ευφυείς πράκτορες.

2. Η ενυπάρχουσα αβεβαιότητα στην όλη διαδικασία μοντελοποίησης του προφίλ του χρήστη.

3. Η απειρία μεγάλου αριθμού χρηστών σε συστήματα μοντελοποίησης του προφίλ τους.

3.17.1 Τεχνικές Μοντελοποίησης

Σύμφωνα με την Kay (2001) η μοντελοποίηση του χρήστη μπορεί να γίνει ακολουθώντας μια από τις παρακάτω μεθοδολογίες.

Αρχικά, υπάρχει η μέθοδος της απόκτησης της απαραίτητης πληροφορίας για την μοντελοποίηση του χρήστη. Θεωρείται η πιο ορθή μεθοδολογία απόκτησης γνώσης και ουσιαστικά ο χρήστης καλείται να δώσει πληροφορίες, μέσα από την συμπλήρωση μιας φόρμας, για τις προτιμήσεις και το γνωστικό του επίπεδο.

Στην συνέχεια υπάρχει η μοντελοποίηση μέσω της παρατήρησης του χρήστη, όπου υποσυστήματα καταγράφουν και παρατηρούν την δραστηριότητα του χρήστη δημιουργώντας έτσι μοντέλα που μπορούν να προβλέψουν την συμπεριφορά του.

Τα στερεότυπα αποτελούν μια ακόμα μεθοδολογία δημιουργίας μοντέλων χρήστη. Υιοθετούνται συχνά σε πολλές προτάσεις δημιουργίας συστημάτων μοντελοποίησης του χρήστη. Με την χρήση στερεοτύπων γίνεται προσπάθεια συγκέντρωσης αντιπροσωπευτικής πληροφορίας για ομάδες ανθρώπων με ως επί το πλείστων κοινά χαρακτηριστικά. Συνήθως, συστήματα μοντελοποίησης που χρησιμοποιούν στερεότυπα ενσωματώνουν στατιστικές τεχνικές δημιουργίας μοντέλων χρήστη.

Η Knowledge Based Reasoning (KBR) προσέγγιση χρησιμοποιείται σε αρκετές περιπτώσεις για την δημιουργία μοντέλων χρήστη και

βασίζεται σε στατιστική επεξεργασία στερεοτύπων και χρήση βάσεων γνώσης. Η τελευταία προσέγγιση αφορά την διαχείριση αβεβαιότητας και μεταβαλλόμενης πληροφόρησης σχετικά με τον χρήστη. Στη συγκεκριμένη περίπτωση, γίνεται προσπάθεια αντιμετώπισης της αβεβαιότητας σχετικά με τις προθέσεις και τις προτιμήσεις του χρήστη.

Τεχνικές Μοντελοποίησης Προφίλ Χρήστη

Τα τελευταία χρόνια, έχουν προταθεί πολλές τεχνικές κατασκευής μοντέλων χρήστη. Μέχρι πρόσφατα, στον χώρο της δημιουργίας μοντέλων χρήστη είχαν επικρατήσει τεχνικές όπου στατιστικά μοντέλα προσπαθούσαν να σκιαγραφήσουν το μοντέλο του χρήστη.

Στατιστικά μοντέλα δημιουργίας προφίλ χρήστη

Τα στατιστικά μοντέλα πρόβλεψης του προφίλ του χρήστη κατανέμονται σε 6 κύριες ομάδες (Zukerman and Albrecht (2001):

Γραμμικά μοντέλα

Έχουν απλή δομή η οποία τα κάνει κατανοητά και εύκολα επεκτάσιμα. Τα γραμμικά μοντέλα παίρνουν σταθμισμένα αθροίσματα γνωστών αξιών για να παράγουν μια τιμή για μια άγνωστη ποσότητα. Για παράδειγμα, έστω στην collaborative αναζήτηση, η προσπάθεια για τη δημιουργία ενός γραμμικού μοντέλου το οποίο θα προβλέπει την κατάταξη άρθρων από τον χρήστη. Για κάθε υποψήφιο προς κατάταξη άρθρο π.χ, οι γνωστές αξίες είναι οι αξιολογήσεις προηγούμενων χρηστών και τα βάρη, το μέτρο της ομοιότητας μεταξύ του χρήστη που κάνει το ερώτημα και των προηγούμενων χρηστών. Το γραμμικό μοντέλο που προκύπτει είναι το σταθμισμένο άθροισμα των αξιολογήσεων. Ένα τέτοιο μοντέλο είχε προταθεί από τον Resnick, 1994. Τα γραμμικά μοντέλα έχουν επίσης υιοθετηθεί και σε συστήματα που εφαρμόζουν την

content based προσέγγιση δημιουργίας μοντέλων χρήστη. Χαρακτηριστικότερα παραδείγματα τέτοιων συστημάτων είναι του (Orwant, 1995) όπου τα γραμμικά μοντέλα χρησιμοποιήθηκαν για την πρόβλεψη του χρόνου που απαιτείται για διαδοχικές επιτυχείς εγγραφές του χρήστη στο σύστημα, και των (Raskutti et al., 1997), όπου με τη χρήση γραμμικών μοντέλων επιτεύχθηκε η πρόβλεψη των αξιολογήσεων των χρηστών.

Μοντέλα TFIDF - (Term Frequency Inverse Document Frequency).

Είναι μια μέθοδος που χρησιμοποιείται κυρίως στον τομέα του Information Retrieval Filtering όπου και βοηθάει στην εύρεση επιστημονικών άρθρων που ταιριάζουν στο προφίλ της αναζήτησης του χρήστη. Κάθε άρθρο αναπαριστάται με ένα σύνολο διανυσμάτων, όπου το βάρος του κάθε διανύσματος αντιστοιχεί σε έναν όρο κλειδί του άρθρου. Οι Balabanovic (1998), Moukas & Maes (1998) και Basu et al (1999) εφάρμοσαν την μέθοδο αυτή σε συστήματα βασιζόμενα στο περιεχόμενο (content based) που έκαναν συστάσεις επιλογής άρθρων σε χρήστες χρησιμοποιώντας τον τρόπο επιλογής παρόμοιας θεματολογίας άρθρων από τους ίδιους χρήστες.

Μοντέλα Markov

Έχουν απλή δομή αφού στηρίζονται στο θεώρημα πως μελλοντικές καταστάσεις εξαρτώνται από πεπερασμένο αριθμό προηγούμενων καταστάσεων. Έτσι, η ύπαρξη κάποιων καταγεγραμμένων γεγονότων μπορεί να μας οδηγήσει σε κάποια πρόβλεψη για την μελλοντική εξέλιξη παρόμοιων καταστάσεων.
Αν για παράδειγμα, η προς ανάλυση διεργασία αφορά την πρόβλεψη των ιστοσελίδων που θα ζητηθούν από έναν χρήστη, τότε ως καταγεγραμμένο γεγονός θα ήταν οι τελευταίες σελίδες που ο ίδιος είχε επισκεφθεί ή το σύνολο των συνδέσεων που ακολούθησε για να φτάσει στην τελευταία ιστοσελίδα. Οι Bestavros (1996), Horvitz (1998) και Zukerman et al. (1999) χρησιμοποίησαν μοντέλα Markov, σε συστήματα που εφάρμοζαν την collaborative προσέγγιση, με σκοπό την πρόβλεψη των επιθυμιών του χρήστη στο διαδίκτυο. Το μοντέλο του Bestavros υπολόγιζε την πιθανότητα να ζητήσει ο χρήστης ένα συγκεκριμένο άρθρο στο μέλλον. Το μοντέλο του Horvitz υπολόγιζε την πιθανότητα ένας χρήστης να αναζητήσει ένα συγκεκριμένο άρθρο

στην ακριβώς επόμενη αναζήτησή του. Οι (Zukerman et al., 1999) συνέκριναν διάφορα μοντέλα Markov για την επιλογή του καταλληλότερου για την ίδια διαδικασία πρόβλεψης με το σύστημα του Horvitz. Οι προβλέψεις που παρήχθησαν με τα προαναφερθέντα μοντέλα χρησιμοποιήθηκαν από συστήματα τα οποία έστελναν στους χρήστες εκ των προτέρων άρθρα που ήταν πιθανό να αναζητήσουν (Bestavros, 1996; Albrecht et al., 1999)

Ταξινόμηση (Classification).

Οι μέθοδοι ταξινόμησης τμηματοποιούν ένα σύνολο αντικειμένων σε ομάδες σύμφωνα με τις ιδιοτιμές αυτών των αντικειμένων. Η τεχνική αυτή χρησιμοποιήθηκε από τους Perkowitz & Etzioni (2000) οι οποίοι χρησιμοποίησαν μια παραλλαγή της κλασικής μεθόδου clustering με σκοπό να δημιουργήσουν αυτόματα σελίδες που περιέχουν links για ιστοσελίδες οι οποίες σχετίζονται μεταξύ τους (και τις οποίες επισκέπτεται ο χρήστης κατά την αναζήτησή του). Η τεχνική αυτή ονομάστηκε cluster mining.

Rule Induction-RI (επαγωγή με χρήση κανόνων).

Η μεθοδολογία RI αποτελείται από ένα σύνολο κανόνων μάθησης οι οποίοι προβλέπουν την τάξη του αντικειμένου παρατηρώντας τις ιδιότητές του. Τεχνικές RI έχουν χρησιμοποιηθεί τόσο στην συνεργατική όσο και στην βάση του περιεχομένου προσέγγιση. Σύμφωνα με την προσέγγιση περιεχομένου, οι (Morales και Pain, 1999), χρησιμοποίησαν το σύστημα Ripper το οποίο μαθαίνει κανόνες από ένα σύνολο χαρακτηριστικών με συγκεκριμένες τιμές. Οι (Chiu και Webb, 1998) συνδύασαν την τεχνική επαγωγής κανόνων C4.5 για να δημιουργήσουν δέντρα απόφασης (Quinlan, 1993), μέσω της τεχνικής μοντελοποίησης Feature Based Modeling η οποία χρησιμοποιείται σε εφαρμογές ευφυούς μάθησης (Webb και Kuzmycz, 1996). Σκοπός τους ήταν η δημιουργία ενός συστήματος πρόβλεψης λαθών σε αριθμητικές πράξεις. Ο Joerding (1999) χρησιμοποίησε τον CDL4, έναν αλγόριθμο μάθησης κανόνων (Shen, 1997), για να δημιουργήσει ένα σύστημα μάθησης των προτιμήσεων των χρηστών σε ένα διαδικτυακό εμπορικό περιβάλλον. Οι Billsus &

Pazzani (1999), εφάρμοσαν ένα μίγμα κανόνων επαγωγής, μεθοδολογιών TFIDF και γραμμικών μοντέλων με σκοπό την σύσταση άρθρων στους χρήστες. Το παραπάνω σύστημά χρησιμοποίησε δύο μοντέλα για να εξετάσει αν ο χρήστης θα έβρισκε ενδιαφέρον το υποψήφιο άρθρο. Το ένα μοντέλο συντηρούσε μια TFIDF διανυσματική αναπαράσταση των άρθρων που βρίσκονταν στη βάση γνώσης του συστήματος, και χρησιμοποίησε μόνο εκείνα τα άρθρα τα οποία εμφάνιζαν ομοιότητες με το υποψήφιο προς ανάγνωση άρθρο. Στη συνέχεια, ένα γραμμικό μοντέλο κατασκευαζόταν το οποίο προέβλεπε αν ο χρήστης θα έβρισκε ενδιαφέρον το συγκεκριμένο άρθρο.

Δίκτυα Bayes.

Η χρήση Bayesian δικτύων και οι παραλλαγές τους έχουν αποκτήσει πολλούς υποστηρικτές τα τελευταία χρόνια και έχουν χρησιμοποιηθεί για πλήθος εργασιών μοντελοποίησης χρήστη (Jameson, 1995). Τα δίκτυα Bayes είναι προσανατολισμένα άκυκλα γραφήματα όπου οι κόμβοι αντιστοιχούν σε τυχαίες μεταβλητές.Τα δίκτυα Bayes και οι προεκτάσεις τους έχουν χρησιμοποιηθεί για πραγματοποίηση διάφορων εργασιών που απαιτούν πρόβλεψη. Οι Horvitz et al. (1998) χρησιμοποίησαν δίκτυα Bayes για να προβλέψουν το είδος της βοήθειας που θα χρειάζονταν χρήστες που έκαναν εργασίες χρησιμοποιώντας λογιστικά φύλλα.

Οι Gmytrasiewicz et al. (1998) και Jameson et al. (2000) χρησιμοποίησαν διαγράμματα επιρροής για να προβλέψουν την συμπεριφορά πρακτόρων. Το σύστημα των Gmytrasiewicz et al. (1998) λάμβανε υπόψη του διάφορα μοντέλα πρόβλεψης της συμπεριφοράς ενός πράκτορα σε ένα σενάριο αεράμυνας. Στο σύστημα των Jameson et al. (2000) γινόταν πρόβλεψη των ποσοστών λάθους των χρηστών οι οποίοι ακολουθούσαν εντολές που δίνονταν με ένα συγκεκριμένο τρόπο (π.χ. διάφορες ενέργειες μαζί σε αντιδιαστολή με μία κάθε φορά) και στην συνέχεια επέλεγε τον τρόπο εκφοράς της εντολής που ελαχιστοποιούσε το λάθος.

Τεχνικές Τεχνητής Νοημοσύνης δημιουργίας μοντέλων χρήστη

Καθώς οι απαιτήσεις για δημιουργία μοντέλων χρήστη με αρκετά μεγάλη ακρίβεια σε σχέση με το πραγματικό του προφίλ, αυξάνονται όλο και περισσότερο, νέες τεχνικές από τον χώρο της Τεχνητής Νοημοσύνης άρχισαν να χρησιμοποιούνται στα συστήματα user modeling.

Γενετικοί αλγόριθμοι.

Οι γενετικοί αλγόριθμοι αποτελούν μια αρκετά μοντέρνα αντίληψη στις τεχνικές δημιουργίας μοντέλων χρήστη. Χρησιμοποιώντας τις αρχές της γενετικής οι αλγόριθμοι μπορούν να δημιουργήσουν το προφίλ του χρήστη. Παράδειγμα τέτοιας εφαρμογής αποτελεί η πρόταση των Lee et al (2002) για ένα σύστημα προτάσεων αγοράς DVD σύμφωνα με το προφίλ του εκάστοτε χρήστη.

Νευρωνικά δίκτυα.

Τα νευρωνικά δίκτυα είναι ικανά να εκφράσουν μεγάλο πλήθος μη γραμμικών διαδικασιών απόφασης. Αυτό γίνεται μέσα από την δομή των νευρωνικών δικτύων, τα μη γραμμικά κατώφλια και τα βάρη που βρίσκονται ανάμεσα στους κόμβους. Από το 1993 οι Jennings & Higuchi χρησιμοποίησαν νευρωνικά δίκτυα, μέσα από την content-based προσέγγιση, για να αναπαραστήσουν τις προτιμήσεις του χρήστη στην επιλογή επιστημονικών άρθρων.

Case Based Reasoning (CBR).

Η τεχνική αυτή είναι επιτρέπει την χρήση της γνώσης από προηγούμενες καταστάσεις (cases) και επιλύει ένα νέο πρόβλημα με τη χρήση της γνώσης αυτής. Στην περίπτωση των μοντέλων χρήστη οι cases μπορούν να παρέχουν πληροφορία η οποία βοηθάει στον

καθορισμό των ιδιαίτερων χαρακτηριστικών της συμπεριφοράς ενός χρήστη.

Ασαφής Λογική (Fuzzy Logic).

Ένα παραδοσιακό σύστημα ασαφούς λογικής έχει τρία στάδια:

1. Ασαφοποίηση-(fuzzification)
2. δημιουργία λογικών συμπερασμάτων
3. Από-ασαφοποίηση (defuzzification)

Στην μοντελοποίηση χρήστη δεν είναι πάντοτε απαραίτητο να ακολουθηθούν τα παραπάνω στάδια αλλά πολλές φορές εφαρμόζονται τμήματα αυτών. Η ασαφής λογική έχει χρησιμοποιηθεί ως εργαλείο για την δημιουργία εφαρμογών που πραγματοποιούν συστάσεις στους χρήστες και σε αυτές τις περιπτώσεις κατέστη δυνατό να συγκεραστούν διαφορετικές προτιμήσεις και προφίλ χρηστών που ως ένα βαθμό είχαν κάποιες ομοιότητες. Στο σύστημα που προτείνεται από τους Ardissono et al (1997) η ασαφής λογική χρησιμοποιείται για τη δημιουργία του μοντέλου χρήστη δικτυακών τόπων ηλεκτρονικών αγορών.

Στην περίπτωση αυτή τα στερεότυπα που χαρακτηρίζουν τον χρήστη δημιουργούνται με την χρήση κατάλληλων συναρτήσεων. Οι Schmit et al (2003) παρουσίασαν ένα σύστημα το οποίο προορίζεται να κάνει συστάσεις προϊόντων σε ένα δικτυακό τόπο ηλεκτρονικού εμπορίου. Το σκορ των υπό πρόταση προϊόντων καθορίζεται από τον Ordered Weighted Averaging συντελεστή ο οποίος επιτρέπει την αναπαράσταση συνδέσεων μεταξύ της ασαφούς λογικής και του συνόλου των διαφορετικών προτιμήσεων των χρηστών.

Fuzzy Clustering.

Με την μέθοδο αυτή γίνεται χειρισμός σχετικών δεδομένων όταν η πληροφορία για τη δημιουργία στερεοτύπων δεν μπορεί να αναπαρασταθεί από αριθμητικά διανύσματα. Στις περιπτώσεις αυτές ο καθορισμός της απόστασης γίνεται με χρήση διανυσματικών αναπαραστάσεων των διεπαφών του χρήστη με το προσαρμοστικό σύστημα υπερμέσων. Κύρια χρήση των τεχνικών Fuzzy Clustering στην μοντελοποίηση χρήστη, είναι η πραγματοποίηση συστάσεων (Nasraoui et al,1999 , Nasraoui et al, 2000) , και (Nasraoui et al, 2003) και κατατάξεων προϊόντων (Joshi et al, 2000) .

Neuro Fuzzy.

Τα Νεύρο - Ασαφή συστήματα χρησιμοποιούν νευρωνικά δίκτυα για να δημιουργήσουν κανόνες και/ ή συναρτήσεις συμμετοχής από τα εισερχόμενα και εξερχόμενα δεδομένα των συστημάτων Fuzzy. Με την προσέγγιση αυτή εξαλείφονται σε κάποιο βαθμό τα εγγενή μειονεκτήματα τόσο των νευρωνικών δικτύων (black box νευρωνικών δικτύων) όσο και των συστημάτων ασαφούς λογικής (δυσκολία εύρεσης κατάλληλων membership τιμών). Ως αποτέλεσμα του συνδυασμού αυτού είναι η αυτοματοποίηση της διαδικασίας μεταφοράς της γνώσης του χρήστη – ειδικού στους κανόνες ασαφούς λογικής. Ένα από τα σημαντικότερα συστήματα της συγκεκριμένης κατηγορίας είναι το ANFIS των Jang et al (1993).

Ο συνδυασμός νευρωνικών δικτύων και συνόλων Fuzzy δημιουργεί μεθοδολογίες που μοντελοποιούν την ανθρώπινη συμπεριφορά και επιτρέπει στα συστήματα NF να χρησιμοποιηθούν σε πληθώρα εφαρμογών. Οι Stathacopoulou et al (2003) χρησιμοποιούν NFS για πραγματοποίηση συστάσεων σε δικτυακό τόπο ηλεκτρονικού εμπορίου και για επιλογή μαθημάτων από on line εκπαιδευτικό οργανισμό αντίστοιχα. Στο σύστημα που προτείνουν οι Magoulas et al (2001 χρησιμοποιούν νεύρο – ασαφή για να δημιουργήσουν μια μεθοδολογία λήψης αποφάσεων για το περιεχόμενο web based μαθημάτων ανάλογα με το γνωστικό επίπεδο του μαθητή – χρήστη.

3.18 Γενετικοί Αλγόριθμοι

3.18.1 Εισαγωγή

Η ανάπτυξη των γενετικών αλγορίθμων ξεκίνησε στην δεκαετία του 1960 από τον John Holland τους συνεργάτες του και τους φοιτητές του στο πανεπιστήμιο του Michigan.Οι σκοποί της έρευνάς τους είχαν διπλή κατεύθυνση:

- Να συνοψίσουν και να εξηγήσουν αυστηρά τις προσαρμοστικές και αναπαραγωγικές διαδικασίες φυσικών συστημάτων και
- Να σχεδιάσουν λογισμικό τεχνικών συστημάτων που να διατηρεί τους πιο σημαντικούς από τους μηχανισμούς των φυσικών συστημάτων.

Η βασική ιδέα των Γενετικών Αλγορίθμων (Genetic Algorithms, GA) είναι η μίμηση των μηχανισμών της βιολογικής εξέλιξης που απαντώνται στη φύση.Σαν επακόλουθο χρησιμοποιείται ορολογία η οποία είναι δανεισμένη από το χώρο της Φυσικής Γενετικής. Έτσι αναφερόμαστε σε άτομα ή γενοτύπους μέσα σε ένα πληθυσμό. Κάθε άτομο ή γενότυπος αποτελείται από χρωμοσώματα.Στους GA αναφερόμαστε σε άτομα με ένα χρωμόσωμα. Τα χρωμοσώματα αποτελούνται από γονίδια που είναι διατεταγμένα σε γραμμική ακολουθία. Κάθε γονίδιο επηρεάζει την κληρονομικότητα ενός ή περισσοτέρων χαρακτηριστικών. Τα γονίδια τα οποία επηρεάζουν συγκεκριμένα χαρακτηριστικά γνωρίσματα του ατόμου ονομάζονται loci και κάθε χαρακτηριστικό γνώρισμα έχει την δυνατότητα να εμφανιστεί με διάφορες μορφές ανάλογα με την κατάσταση στην οποία βρίσκεται το αντίστοιχο γονίδιο που την επηρεάζει. Οι διαφορετικές αυτές καταστάσεις που μπορεί να πάρει το γονίδιο ονομάζονται alleles.

Κάθε γενότυπος αναπαριστά μια πιθανή λύση στο πρόβλημά μας. Το αποκωδικοποιημένο περιεχόμενο ενός συγκεκριμένου χρωμοσώματος καλείται φαινότυπος.Μια διαδικασία εξέλιξης που εφαρμόζεται πάνω σε ένα πληθυσμό αντιστοιχεί σε ένα εκτενές ψάξιμο στο χώρο των πιθανών λύσεων και με την εξερεύνηση όλου του διαστήματος και την διατήρηση των πιθανών καλύτερων λύσεων θα έχουμε τα επιθυμητά αποτελέσματα.

Οι GA διατηρούν ένα πληθυσμό πιθανών λύσεων του προβλήματος που θέλουμε να επιλύσουμε πάνω στον οποίο δουλεύουν αντίθετα από άλλες μεθόδους αναζήτησης που επικεντρώνονται σε ένα μόνο σημείο του διαστήματος αναζήτησης. Έτσι πραγματοποιείται ταυτόχρονη αναζήτηση σε πολλές κατευθύνσεις και υποστηρίζει καταγραφή και ανταλλαγή πληροφοριών μεταξύ αυτών των κατευθύνσεων. Ο πληθυσμός υφίσταται μια προσομοιωμένη γενετική εξέλιξη οπότε σε κάθε γενιά οι σχετικά καλές λύσεις αναπαράγονται ενώ γενικά οι κακές απομακρύνονται ώστε στο τέλος να βρούμε την βέλτιστη λύση. Ο διαχωρισμός και η αποτίμηση των διαφόρων λύσεων γίνεται με τη βοήθεια μιας αντικειμενικής συνάρτησης η οποία παίζει τον ρόλο του περιβάλλοντος μέσα στο οποίο εξελίσσεται ο πληθυσμός.(Παλαιολόγος Κ,2009)

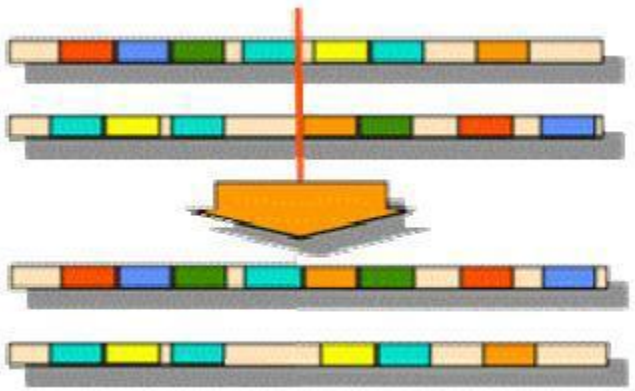
Συνοψίζοντας ένας GA για ένα πρόβλημα πρεέπει να αποτελείται από τα παρακάτω συστατικά:

- Μια γενετική αναπαράσταση των πιθανών λύσεων του προβλήματος
- Ένα τρόπο δημιουργίας ενάς αρχικού πληθυσμού από πιθανές λύσεις
- Μια αντικειμενική συνάρτηση αξιολόγησης των μελών του πληθυσμού που παίζει το ρόλο του περιβάλλοντος
- Γενετικούς τελεστές για την δημιουργία νέων μελών
- Τιμές για τις διάφορες παραμέτρους που χρησιμοποιεί ο GA

3.18.2 Μεθοδολογία Γενετικών Αλγορίθμων

Κατά την αρχικοποίηση του πληθυσμού δημιουργείται ένα σύνολο από χρωμοσώματα. Μετά την αρχικοποίηση επιλέγονται οι γονείς σύμφωνα με μια συνάρτηση πιθανότητας η οποία βασίζεται στην σχετική ποιότητα των ατόμων του πληθυσμού. Με άλλα λόγια τα άτομα με καλύτερη ποιότητα έχουν περισσότερες πιθανότητες να επιλεγούν ως γονείς. Όσο καλύτερη είναι η ποιότητα ενός ατόμου τόσο αυξάνονται οι πιθανότητες

να επιλεγεί σαν γονέας για την αναπαραγωγή παιδιών. Γενικά από Ν γονείς αναπαράγονται Ν παιδιά μέσω της διασταύρωσης (crossover) όπως ονομάζεται ο ανασυνδιασμός στην περίπτωση των GA. Τυπικά ακολουθεί η μετάλλαξη των παιδιών αντικαθιστώντας Ν γονείς του πληθυσμού και δημιουργώντας μια νέα γενιά.



Σχήμα 13

Διασταύρωση γενετικού αλγορίθμου

Πηγή: Google

Ο τελεστής διασταύρωσης μιμείται χονδρικά τον βιολογικό ανασυνδυασμό μεταξύ οργανισμών με μονό χρωμόσωμα. Μέσω της διασταύρωσης οι ΓΑ εκμεταλλεύονται αποτελεσματικά ιστορικές πληροφορίες για να κάνουν υποθέσεις πάνω σε νέα σημεία έρευνας με προσδωκόμενη βελτιωμένη απόδοση.

Τα βασικά βήματα ενός Γενετικού αλγορίθμου είναι τα εξής:

Αρχικοποίηση. Τυχαία δημιουργία αρχικού πληθυσμού λύσεων και εκτίμηση της συνάρτησης καταλληλότητας.
Νέος Πληθυσμός. Δημιουργία νέου πληθυσμού με επανάληψη των βημάτων μέχρι την ολοκλήρωση του νέου πληθυσμού:
Επιλογή. Επιλογή δύο γονέων από τον πληθυσμό ανάλογα με την καταλληλότητά τούς (μεγαλύτερη καταλληλότητα σημαίνει μεγαλύτερη

πιθανότητα επιλογής).
Διασταύρωση: Οι γονείς διασταυρώνονται με κάποια πιθανότητα για τη δημιουργία νέων απογόνων. Εάν οι γονείς δεν διασταυρωθούν τότε ο απόγονος είναι ακριβές αντίγραφο του γονέα.

Μεταλλαξη: Μεταλλάσσονται οι απόγονοι σε τυχαία επιλεγμένη Θέση.
Αποδοχή: Τοποθέτηση απογόνου στο νέο πληθυσμό.
Εκτίμηση: Υπολογισμός των τιμών καταλληλότητας του νέου πληθυσμού.
Δοκιμή: Εάν κάποιο κριτήριο τερματισμού ικανοποιείται, ο αλγόριθμος σταματά και επιστρέφει την καλύτερη λύση στον τρέχοντα πληθυσμό.
Επανάληψη: Εάν δεν ικανοποιείται κανένα κριτήριο τερματισμού, γύρνα στο Νέος Πληθυσμός.

3.18.3 Αξιολόγηση Γενετικών Αλγορίθμων

Οι γενετικοί αλγόριθμοι αποτελούν την πλέον εφαρμόσιμη τεχνική εξελικτικών αλγορίθμων που παρουσιάζει διάφορα πλεονεκτήματα και μειονεκτήματα σε ζητήματα υλοποίησης και εφαρμογών.

Πλεονεκτήματα:

- Οι γενετικοί αλγόριθμοι δεν έχουν ιδιαίτερες μαθηματικές απαιτήσεις για τα προβλήματα που καλούνται να βελτιστοποιήσουν.Επίσης είναι σε θέση να χειριστούν όλα τα είδη αντικειμενικών συναρτήσεων και όλους τους πιθανούς περιορισμούς που μπορεί να προϋποθέτει ο ορισμός του προβλήματος που μπορεί να είναι ορισμένοι σε διακριτό, συνεχές ή συνδυασμένο χώρο αναζήτησης.
- Χαρακτηρίζονται για την αξιόλογη ευελιξία τους που τους επιτρέπει να εφαρμοστούν σε διάφορες υβριδικές μορφές γενετικών αλγορίθμων με άλλες μεθόδους δίνοντας την δυνατότητα αποδοτικής υλοποίησης για ένα συγκεκριμένο πρόβλημα.
- Είναι ιδιαίτερα αποδοτικοί ως μέθοδοι γενικευμένης αναζήτησης. Οι κλασικές τεχνικές πραγματοποιούν συγκρίσεις που γίνονται ανάμεσα σε κοντινές τιμές και η διαδικασία οδηγείται σε τοπικά

βέλτιστα σημεία. Η ολική βέλτιστη λύση (global optimal solution) μπορεί να βρεθεί μόνο αν το πρόβλημα διαθέτει συγκεκριμένες ιδιότητας κυριότητας που εξασφαλίζουν πως το τοπικό ολικό είναι και ολικό βέλτιστο.

- Δίνουν τεράστιες δυνατότητες επέκτασης και διαμόρφωσης ανάλογα με τις ξεχωριστές ανάγκες της κάθε υλοποίησης. Πολλές από τις λειτουργίες που έχουν εφαρμοστεί σε ΓΑ διαφέρουν αρκετά από τις γενετικής εξέλιξης.

Μειονεκτήματα:

- Η διαδικασία εκτέλεσης ενός γενετικού αλγορίθμου απαιτεί επαναλαμβανόμενο υπολογισμό των συναρτήσεων καταλληλότητας γεγονός που κάνει ιδιαίτερα πολύπλοκη την εύρεση της βέλτιστης λύσης. Σε πολλά προβλήματα της πραγματικότητας ο συνεχής υπολογισμός της συνάρτησης καταλληλότητας μπρορεί να είναι απογορευτικά πολύπλοκος και χρονοβόρος. Η λύση σε αυτό το εμπόδιο είναι ο υπολογισμός και η εφαρμογή μιας μπροσεγγιστικής συνάρτησης καταλληλότητας.
- Η επιλογή της βέλτιστης λύσης του προβλήματος γίνεται σε σύγκριση με τις άλλες διαθέσιμες λύσεις. Συνεπώς υπάρχει ο κίνδυνος η συνθήκη τερματισμού να μην οδηγεί σε όλες τις περιπτώσεις στη συνολικά βέλτιστη λύση.
- Σε συγκεκριμένα προβλήματα προβλήματα αναζήτησης και βελτιστοποίησης οι εναλλακτικές τεχνικές θεωρούνται αποτελεσματικότερες σε σχέση με τους γενετικούς αλγορίθμους. Ακόμα και σήμερα είναι ανοιχτή η ερώτηση αν υπάρχουν προβλήματα που οι γενετικοί αλγόριθμοι είναι η βέλτιστη τεχνική για την βελτιστοποίηση τους.
- Σε προβλήματα που η μοναδική συνάρτηση καταλληλότητας είναι μια ένδειξη αποδεκτής ή μη αποδεκτής λύσης (π.χ προβλήματα απόφασης) καθώς δεν υπάρχει τρόπος να συγκλίνουν προς την λύση. Οι γενετικοί αλγόριθμοι δεν είναι ιδιαίτερα αποδοτικοί καθώς τεχνικές τυχαίας αναζήτησης μπορούν να οδηγήσουν σε αυτές τις περιπτώσεις σε ίδια αποτελέσματα με έναν γενετικό αλγόριθμο.

ΚΕΦΑΛΑΙΟ 4

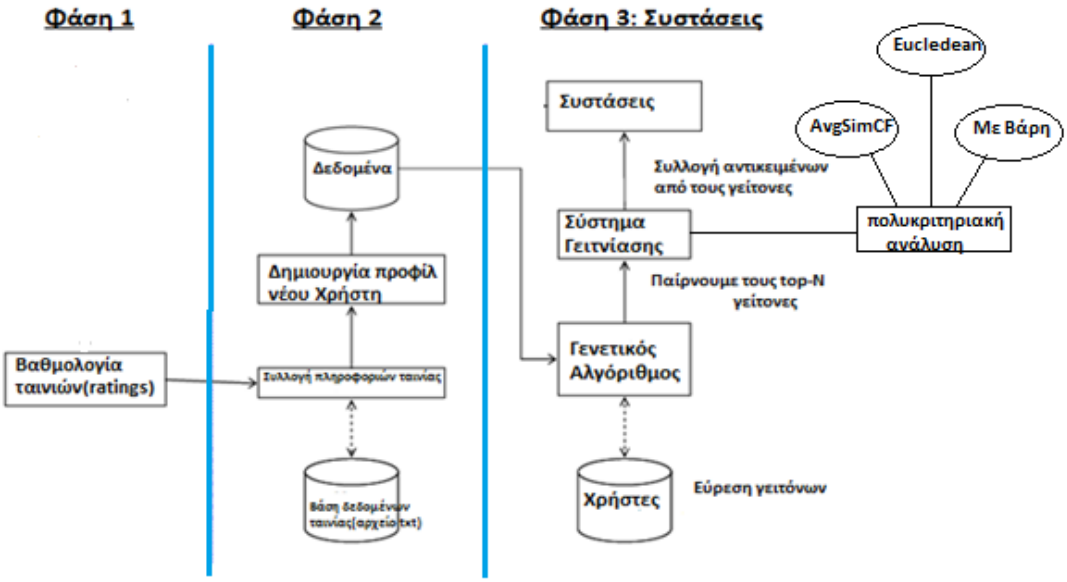
ΜΕΘΟΔΟΛΟΓΙΑ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ

Στο κεφάλαιο αυτό θα αναλύσουμε την μεθοδολογία που θα χρησιμοποιήσουμε με σκοπό τη σύσταση του recommendation system.

4.1 Εισαγωγή

Τα recommendation systems βασίζονται κυρίως στο content-based , στο collaborative, και στο hybrid filtering. Τα content-based recommendations προτείνουν αντικείμενα στους χρήστες τα οποία είναι παρόμοια με άλλα αντικείμενα για τα οποία οι χρήστες εκδήλωσαν ενδιαφέρων παλαιότερα. Στο collaborative filtering προτείνονται αντικείμενα τα οποία είναι παρόμοια μεταξύ τους ή αντικείμενα για τα οποία κάποιος παρόμοιος χρήστης εκδήλωσε ενδιαφέρον. Τα Hybrid συστήματα συνδιάζουν τα δυο παραπάνω συστήματα σε μια υβριδική προσέγγιση για να κερδίσουν καλύτερη απόδοση. Ένα τέτοιο σύστημα είναι και το σύστημα που θα αναπτύξουμε στην παρούσα διπλωματική εργασία.

4.2 Γραφική απεικόνιση Συστήματος



Σχήμα 14
Αρχιτεκτονική Μεθοδολογικού Πλαισίου Συστήματος

Φάση 1:

Σαν είσοδο και σαν Βαθμολογία ratings έχουμε την αξιολόγηση 6078 χρηστών που μας δόθηκε. Οι βαθμολογίες αυτές είναι σε μορφή txt το οποίο ενσωματώσαμε στο matlab αρχείο μας. Οι τιμές κυμαίνονται από 0 έως 1 (πραγματικοί αριθμοί-real values) και αφορούν τις αξιολογίσεις χρηστών για ταινίες σε 4 κριτήρια η κάθε μία.

Φάση 2:

Για τις ταινίες που οι χρήστες έχουν αξιολογήσει ψάχνουμε για τις διάφορες γενεές και τους χαρακτήρες από το αρχείο txt καθώς επίσης βλέπουμε την προτίμησή τους για κάποια ταινία με βάση την αξιολόγηση των 4 κριτήριων για την ταινία.Όσο μεγαλύτερος είναι ο πραγματικός αριθμός για κάποιο κριτήριο διαπιστώνουμε ότι ο εν λόγο χρήστης θεωρεί το συγκεκριμένο κριτήριο ως σημαντικό για την επιλογή του.

Στην περίπτωση των συστάσεων ταινιών τα προτεινόμενα είδη αντικειμένων είναι είναι τα είδη της ταινίας όπως τα 4 κριτήριά μας *ηθοποιία*, η *πλοκή*, η *σκηνοθεσία* και τα *οπτικά εφέ*.

Αντί να βασιζόμαστε μόνο στις αξιολογήσεις των χρηστών παίρνουμε με βάση τη τιμή του χρήστη σε κάθε κριτήριο ώστε να εξάγουμε τα βάρη στα κριτήρια για τον χρήστη. Στη συνέχεια αθροίζουμε τον συνολικό αριθμό για το συγκεκριμένο είδος προτίμησης και ανανεώνουμε το παλιό με σκοπό να βρούμε το αγαπημένο για το χρήστη π.χ για τον σκηνοθέτη. Αυτό γίνεται για την καλύτερη σύσταση μιας ταινίας στο τέλος αφού βασιζόμαστε σε δύο αντί για ένα ενδεχόμενα.

Φάση 3:

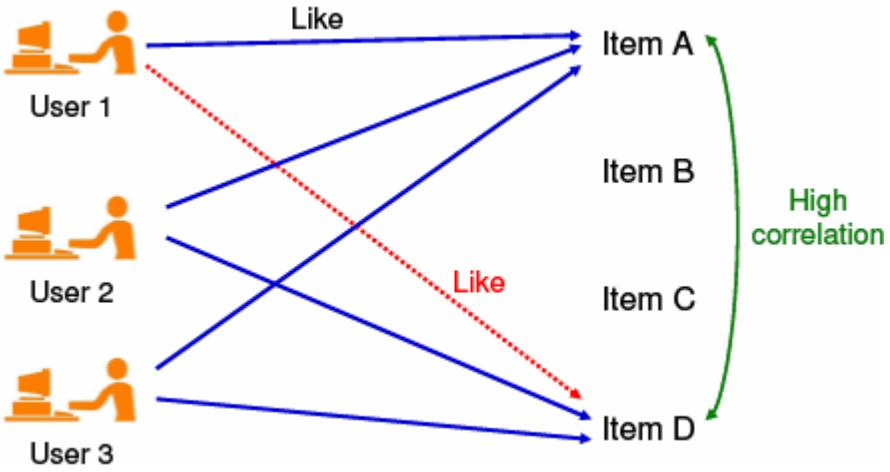
Σε αυτή τη φάση χρησιμοποιούμε τον γενετικό μας αλγόριθμο ο οποίος αναλύεται στην συνέχεια του κεφαλαίου με σκοπό να ψάξει για την κατάλληλη σύσταση (feature selection).Η επιλογή των features είναι η διαδικασία κατά την οποία το σύστημα διαλέγει ένα υποσύνολο με βάση ένα συγκεκριμένο κριτήριο. Το πρώτο βήμα του γενετικού μας αλγορίθμου είναι να προετοιμάσει τα features που μας χρειάζονται στην περίπτωσή μας και για τα 4 κριτήρια. Στην συνέχεια και στο τέλος υλοποίησης του συστήματος για την δημιουργία μιας σύστασης χρησιμοποιούμε μεθόδους γειτνίασης με βάση την πολυκριτηριακή ανάλυση λόγω των πολλών κριτηρίων μας. Οι μέθοδοι αυτοί αναλύονται εκτενέστερα στην συνέχεια του κεφαλαίου. Χρησιμοποιήσαμε τους τύπους που μας δίνονται και τους εφαρμόσαμε σε πρόγραμμα matlab. Επίσης δημιουργήσαμε στο ίδιο πρόγραμμα τον γενετικό αλγόριθμο μόνοι μας. Σαν έξοδο του συστήματός μας παίρνουμε την σύσταση μιας ταινίας για έναν συγκεκριμένο χρήστη με βάση τους γείτονες με τα ίδια ενδιαφέροντα. Μπορούμε να πούμε ότι η μέθοδος του γενετικού εφαρμόζεται γιατί είναι αξιόπιστη. Γενικά οι ευρετικοί αλγόριθμοι εξάγουν καλύτερα αποτελέσματα όσο τρέχει το πρόγραμμα και με πολυάριθμα δεδομένα εισόδου. Ειδικότερα χρησιμοποιήσαμε μόνο τους τύπους για την εύρεση γειτόνων από δημοσιεύσεις σχετικά με τα συστήματα αυτά (τους ‘μεταφράσαμε’ σε matlab) καθώς και τα δεδομένα εισόδου από την κ.Λακιωτάκη και τον κ.Ματσατσίνη. Σκοπός της εργασίας είναι η χρήση μεγάλου αριθμού δεδομένων σαν είσοδο, η αξιολόγισή τους με σκοπό μια ακριβής σύσταση και τέλος η σύγκριση με άλλες μεθόδους με βάση τα αποτελέσματα.

4.3 Ο Γενετικός Αλγόριθμος στο πρόβλημά μας

Το collaborative filtering (CF) βασίζεται στην ομοιότητα μεταξύ ενός ενεργού χρήστη και των υπόλοιπων χρηστών. Στη συνέχεια βρίσκει items τα οποία ο ενεργός χρήστης δεν έχει ξαναδεί αλλά το σύστημα υποθέτει (εκτιμά-προβλέπει) ότι θα τον ενδιαφέρουν διότι οι υπόλοιποι χρήστες που έχουν τις ίδιες προτιμήσεις με αυτόν τα προτιμούν.

Η ομοιότητα μεταξύ των προτιμήσεων των χρηστών μπορεί να μετρηθεί με το ίδιο item (item-based CF) ή με το ίδιο στυλ χρήστη (user-based CF). Αυτή η μέθοδος χρησιμοποιείται για να προβλέψει τη χρησιμότητα ενός συγκεκριμένου item για τον δεδομένο χρήστη βασισμένη σε προηγούμενα items που του αρέσουν ή στις προτιμήσεις άλλων χρηστών με τα ίδια γούστα.(Fong et al, 2011)

Στο item-based CF (Σχήμα 15) αν πολλοί χρήστες (User 2, User 3) ενδιαφέρονται για το Item A και Item D υποθέτουμε ότι τα δύο αυτά Items σχετίζονται μεταξύ τους κατά πολύ.



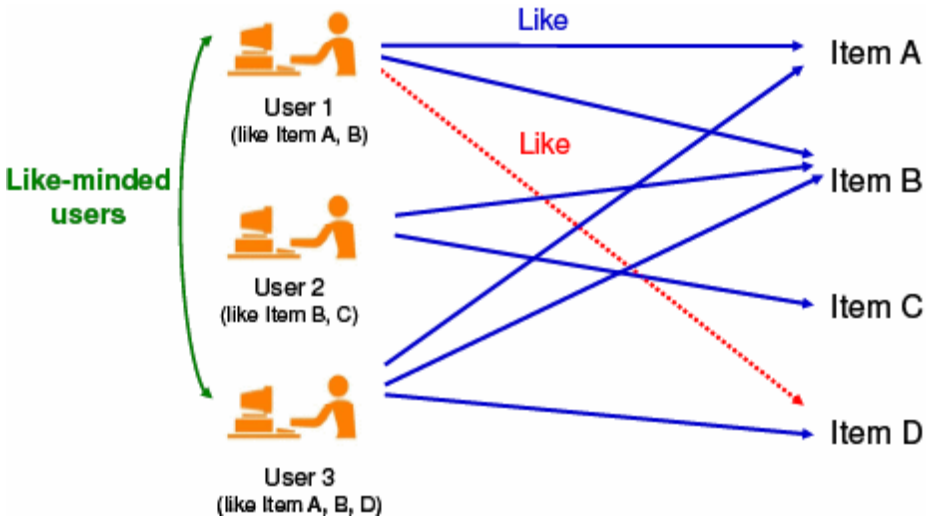
Σχήμα 15

Συσχέτιση μεταξύ χρηστών-αντικειμένων

Πηγή: S.Fong,Y.Ho,Y.Hang,2011

Οι τεχνικές item-based αναλύουν τον πίνακα user-item για να προσδιορίσουν τις σχέσεις μεταξύ των χρηστών και των αντικειμένων με σκοπό να προτίνουν συστάσεις για τους χρήστες.

Στο user-rating CF (Σχήμα 16) η προτίμηση του User3 είναι παρόμοια με αυτή του User1 διότι και οι δύο προτιμούν τα Item A και Item B από κοινού, άρα έχουν τις ίδιες σχεδόν προτιμήσεις. Ενώ ο User3 προτιμά το Item D όπως επίσης και ο User1.



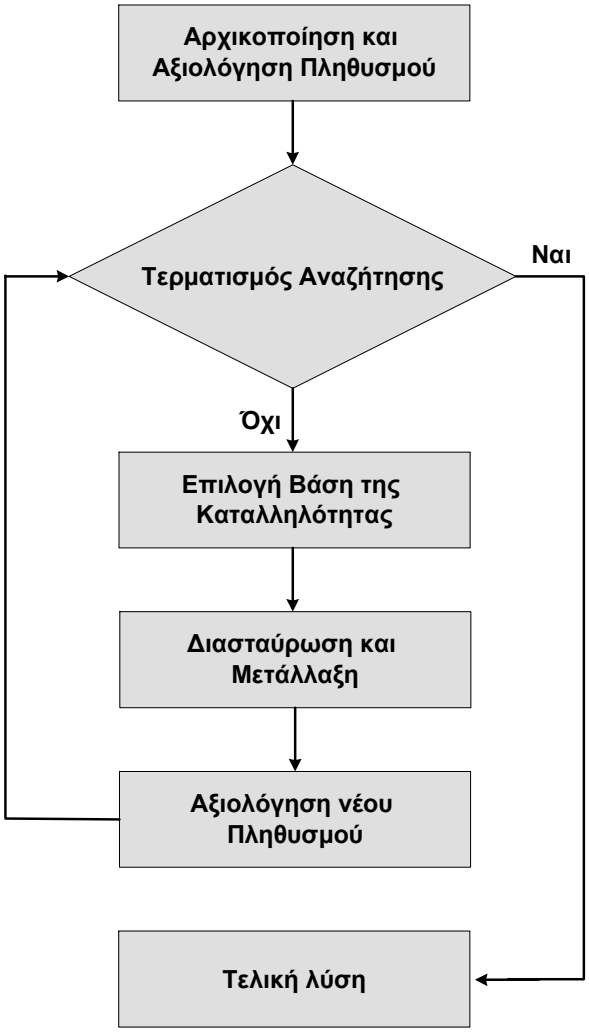
Σχήμα 16

Συσχέτιση χρηστών-αντικειμένων (παρόμοια γούστα χρηστών)

Πηγή: S.Fong,Y.Ho,Y.Hang,2011

Ένα γενικό πλάνο είναι το εξής:
Το user-based CF δίνει συστάσεις σε ομάδες ατόμων που σχετίζονται μεταξύ τους. Χρησιμοποιώντας τα ratings των χρηστών οι οποίοι σχετίζονται κατά πολύ μεταξύ τους, τα items τα οποία δεν έχουν ΄κριθεί΄ από έναν χρήστη μπορούν να προβλευθούν.(S.Fong,Y.Ho,Y.Hang,2011)

Ο μορφή του γενετικού αλγορίθμου που εφαρμόζουμε δίνεται στο Σχήμα 17:



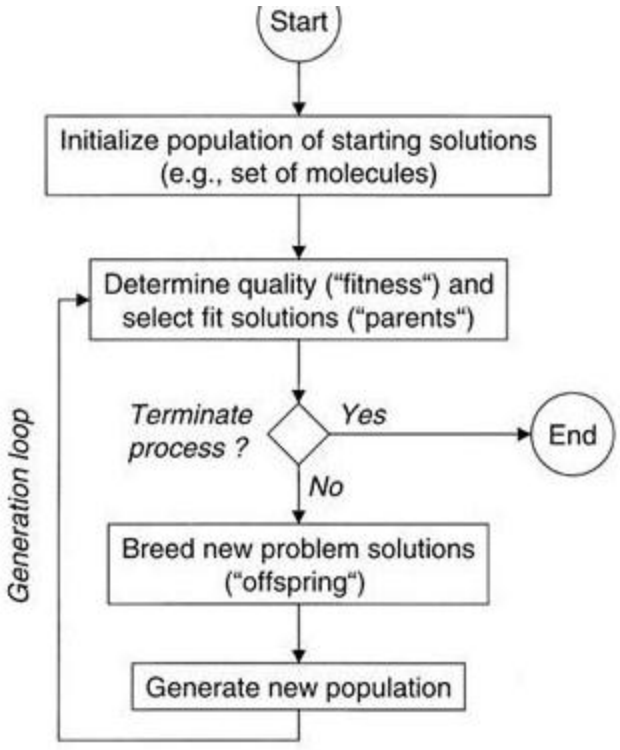
Σχήμα 17

Σε μορφή ψευδοκώδικα έχουμε:

// start with an initial time
t := 0;
// initialize a usually random population of individuals
initpopulation P (t);
// evaluate fitness of all initial individuals of population
evaluate P (t);
// test for termination criterion (time, fitness, etc.)
while termination condition not met do

// increase the time counter
t := t + 1;
// select a sub-population for offspring production
P' := selectparents P (t);
// recombine the "genes" of selected parents
recombine P' (t);
// perturb the mated population stochastically
mutate P' (t);
// evaluate its new fitness
evaluate P' (t);
// select the survivors from actual fitness
P := survive P,P' (t);
end while

και σε μορφή διαγράμματος ροής στο Σχήμα 18:



Σχήμα 18

4.3.1 Το μοντέλο Hybrid

Το μοντέλο αυτό το οποίο θα χρησιμοποιηθεί και στα πλαίσια της παρούσας διπλωματικής εργασίας συνδιάζει τις CF και content-based με σκοπό να αποφύγει τα προβλήματα που έχουν ξεχωριστά οι δύο μέθοδοι αυτοί.

Οι διαφορετικές μέθοδοι με τις οποίες δημιουργείται ένα Hybrid recommender system είναι οι εξής:

1. Εφαρμογή ξεχωριστά του collaborative και content-based μεθόδων και σύγκριση των προβλέψεών τους.
2. Ενσωμάτωση χαρακτηριστικών του content-based στην collaborative προσέγγιση.
3. Ενσωμάτωση χαρακτηριστικών του collaborative στην content-based προσέγγιση.
4. Κατασκευή ενός μοντέλου που συνδιάζει από κοινού ενός collaborative και content-based μοντέλου από κοινού.

Filtering approach	Input data
Content-based analysis	Item contents + User ratings (by a single user)
Item-based CF	Item contents + User ratings (by multiple users)
User-based CF	User demographic information + User ratings (by multiple users) +
Hybrid CF (e.g. GA-based CF)	User demographic information + Item contents + User ratings (by multiple users)

Σχήμα 19
Πηγή: S.Fong,Y.Ho,Y.Hang,2011

Στο Σχήμα 19 παρατηρούμε τι δέχεται για είσοδο το recommendation system μας. Στην περίπτωση μας χρησιμοποιείται η 4^{η} προσέγγιση χωρίς τα demographic δεδομένα.

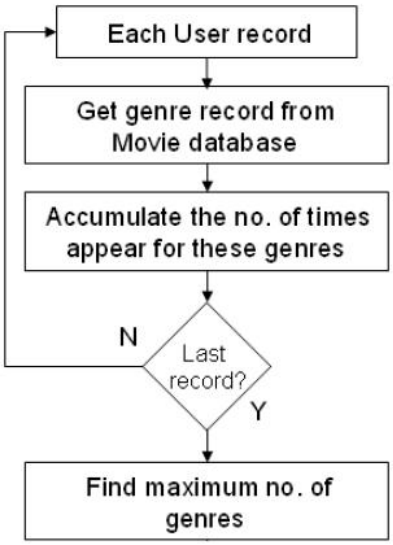
4.3.2 Fittest Τιμές

Στον αλγόριθμό μας αυτόν γίνεται η διατήρηση του πληθυσμού υπολογίζοντας μετασχηματισμούς πινάκων.

Χρησιμοποιώντας το selection, το crossover και το mutation για το γενετικό μας αλγόριθμο βρίσκουμε τις fitness τιμές επιλέγοντας αυτά που μας ταιριάζουν περισσότερο(fittest). Η συλλογή των fittest αυτών τιμών περιγράφεται από τα χρωμοσώματα τα οποία είναι σε μορφή πραγματικών τιμών.

4.3.2.1 Κωδικοποίηση Χρωμοσωμάτων

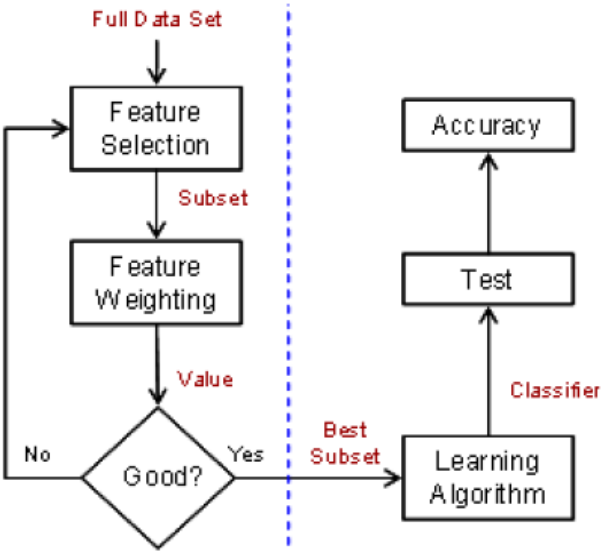
Η δομή των χρωμοσωμάτων του γενετικού αλγορίθμου σχεδιάστηκε με σκοπό να χρησιμοποιήσει όλα τα δεδομένα που μας δόθηκαν για όλους τους χρήστες για την υποστήριξη του hybrid συστήματος. Παρόλα αυτά σε μερικές περιπτώσεις θεωρήθηκε ως χρησιμότερη η μείωση των αριθμών τον χρηστών όταν ο αλγόριθμος συνέκλινε από κάποιο σημείο και μετά. Στο μοντέλο αυτό, πρώτα ο γενετικός αλγόριθμος χρησιμοποιείται για να ρυθμίσει τα βάρη των χαρακτηριστικών για κάθε ενεργό χρήστη. Αμέσως μετά το Σύστημα Συνεργασίας εφαρμόζεται για τη διαμόρφωση της γειτονιάς για τον ενεργό χρήστη. Το χρωμόσωμα στη διαδικασία του γενετικού αλγόριθμου αναπαριστάται ως ένα διάνυσμα βαρών με 4 γονίδια. Ο γενετικός αλγόριθμος ξεκινά με τυχαίους γονότυπους και αρχικό πληθυσμό 100 χρωμοσώματα. Για κάθε ενεργό χρήστη, ένα τυχαία επιλεγμένο χρωμόσωμα ανατίθεται και δοκιμάζεται από τη συνάρτηση καταλληλότητας. Η συνάρτηση καταλληλότητας μετρά την ακρίβεια πρόβλεψης των αντικειμένων π.χ ταινιών βασιζόμενη στο τρέχον χρωμόσωμα.



Σχήμα 20

Επιλογή γενεών γενετικού αλγορίθμου

Πηγή : Fong et al, 2011



Σχήμα 21

Η διαδικασία επιλογής(selection) του γενετικού αλγόριθμου

Πηγή : Fong et al,2011

[93]

Στο Σχήμα 21 δίνεται η ροή των πληροφοριών του γενετικού μας αλγορίθμου. Η επιλογή των χαρακτηριστικών είναι η διαδικασία η οποία επιλέγει ένα μικρότερο σύνολο χαρακτηριστικών με βάση ένα συγκεκριμένο από τα 4 κριτήρια του προβλήματός μας.
Για το Hybrid συστήμα αυτό και τα τέσσερα κριτήρια χρησιμοποιούνται.

4.4 Πολυκριτηριακή Ανάλυση στο πρόβλημά μας

Ο στόχος των πολυκριτηριακών recommendation systems είναι η εύρεση αντικειμένων που μεγιστοποιούν τη χρησιμότητα του κάθε χρήστη όπως και στα single-rating συστήματα. Γι' αυτό ένας από τους σημαντικότερους στόχους των συστημάτων αυτών είναι να μπορούν να προβλέψουν τη συνολική αξιολόγηση του κάθε αντικειμένου για όλους τους χρήστες με στόχο να κάνει τις καλύτερες συστάσεις.

4.4.1 AvgSimCF

Σε σχέση με την προσέγγιση του CF και λόγω της φύσης του προβλήματος της εργασίας αυτής συνιστάται η χρήση της πολυκριτηριακής προσέγγισης.
Τα πολλαπλά κριτήρια και στην περιπτωσή μας τέσσερα μας δίνουν περισσότερες πληροφορίες για τις προτιμήσεις των χρηστών απ' ότι τα παραδοσιακά recommendation systems. Έτσι εδώ το rating function αλλάζει σε Users X Items -> R_0 x R_1 x R_2 x R_3 x R_c όπου c ο αριθμός των κριτηρίων. Η χρησιμότητα των κριτηρίων είναι να μας δείξουν ποιες πτυχές του προιόντος ο χρήστης αξιολογεί ως καλύτερες.
Ο αλγόριθμος για την πολυκριτηριακή ανάλυση εκτείνεται από αυτών με την single-rating σύμφωνα με τους Adomavicius και Kwon η οποία αναφέρεται ως AvgSimCF είναι η εξής:

$$sim_{avg}(u,u') = \frac{1}{k+1}\sum_{c=0}^{k} sim_c(u,u') \quad (4.1)$$

Πηγή: Multi-Criteria Service Recommendation Based on User Criteria Preferences L.Liu,N.Mehandjiev,D.Xu (2011)

Όπου K+1 είναι τα κριτήρια .

4.4.2 EuclideanCF

Μια διαφορετική προσέγγιση για την πολυκριτηριακή ανάλυση είναι η ευκλείδια απόσταση μεταξύ δυο σημείων. Η απόσταση μεταξύ δυο μετρήσεων και δύο χρηστών σχετίζεται αντίστροφα. Για παράδειγμα η ευκλείδια απόσταση μεταξύ δυο χρηστών βρίσκεται από τον τύπο:

$$d_i(u,u') = \sqrt{\sum_{c=0}^{k}(r_c - r'_c)^2} \quad (4.2)$$

Πηγή: Multi-Criteria Service Recommendation Based on User Criteria Preferences L.Liu,N.Mehandjiev,D.Xu (2011)

Όπου k ο συνολικός αριθμός κριτηρίων και r_c η αξιολόγηση του χρήστη u στο κριτήριο c . Η ομοιότητα μεταξύ των χρηστών u και u’ δίνεται από τον τύπο:

$$sim(u,u') = 1/(1+\frac{1}{I}\sum_{i\in I} d_i(u,u'))$$

(4.3)

Πηγή: Multi-Criteria Service Recommendation Based on User Criteria Preferences, L.Liu,N.Mehandjiev,D.Xu (2011)

Όπου I ο αριθμός των items που αξιολόγησαν και οι δύο χρήστες ο οποίος στην περίπτωση μας είναι ίσος με ένα.

4.4.3 Πολυκριτήρια με χρήση Βαρών

Τα χαρακτηριστικά των βαρών είναι το επόμενο βήμα επιλογής. Κάθε χρήστης δίνει διαφορετική βαρύτητα στα κριτήρια ενός προιόντος. Οι χρήστες οι οποίοι δίνουν το ίδιο ή περίπου όμοιο βάρος στα συγκεκριμένα κριτήρια μεταξύ τους είναι πολύ πιο εύκολο να συγκριθούν αποτελεσματικά. Π.χ ας υποθέσουμε ότι μιλάμε για μια συγκεκριμένη ταινία και τα κριτήρια τα οποία μπορεί να επιλέξει ένας χρήστης είναι η *Σκηνοθεσία,* το *Σενάριο* και οι *Ηθοποιοί.* Κάποιος χρήστης μπορεί να θέλει να επιλέξει τη συγκεκριμένη ταινία με βάση την *Σκηνοθεσία*, ενώ ένας διαφορετικός χρήστης με βάση το *Σενάριο* της ταινίας. Οι δύο αυτοί χρήστες δίνουν διαφορετικά βάρη μεταξύ τους στα διάφορα κριτήρια. Ενώ ένας τρίτος χρήστης που εισάγεται στο σύστημα επιλέγει με βάση και αυτός την *Σκηνοθεσία* είναι πιο εύκολο και ασφαλές να συγκριθεί και ενδεχομένως να έχει τις ίδιες προτιμήσεις με τον χρήστη που επέλεξε και αυτός τον προιόν με βάση το κριτήριο *Σκηνοθεσία*. Τα κριτήρια αυτά ονομάζονται significant criteria (L.Liu,N.Mehandjiev,D.Xu (2011), Multi-Criteria Service Recommendation Based on User Criteria Preferences).

Μετά την ανάκτηση των καλύτερων βαρών μεταξύ των κριτηρίων για το χρήστη,το επόμενο βήμα είναι η διαμόρφωση της γειτονιάς με τον υπολογισμό του δείκτη ομοιότητας ο οποίος υπολογίζεται με τον τύπο:

$$\text{Similarity(k,a)} = 1 - \sqrt{\sum_{i=1}^{5} w_k^i \square \left(CPP_k^i - CPP_a^i\right)^2} \quad (4.4)$$

Πηγή: Using Genetic Algorithms for Personalized Recommendation Chein-Shung Hwang, Yi-Ching Su, and Kuo-Cheng Tseng, (2010).

όπου $w_k = w_k^1, w_k^2, w_k^3, w_k^4$ είναι τα βάρη του χρήστη k τα οποία υπολογίστηκαν στον γενετικό αλγόριθμο. Τα CPP υπολογίστηκαν με τη μέθοδο Utastar.

4.5 Μοντέλο Δημιουργίας Προφίλ (PGM)

Ο στόχος του μοντέλου αυτού είναι η δημιουργία του προφίλ προτιμήσεων για κάθε χρήστη. Το προφίλ προτιμήσεων του χρήστη (Customer Preference Profile, CPP) *k* περιγράφεται από τον μαθηματικό τύπο:

$$CPP_k = (CPP^I{}_k, CPP^2{}_k, CPP^3{}_k\ CPP^4{}_k), \quad (4.5)$$

4.6 Μοντέλο Επιλογής Γειτονιάς (NSM)

Στο μοντέλο αυτό, πρώτα ο γενετικός αλγόριθμος χρησιμοποιείται για να ρυθμίσει τα βάρη των χαρακτηριστικών για κάθε ενεργό χρήστη. Αμέσως μετά το Σύστημα Συνεργασίας εφαρμόζεται για τη διαμόρφωση της γειτονιάς για τον ενεργό χρήστη. Το χρωμόσωμα στη διαδικασία του γενετικού αλγόριθμου αναπαριστάται ως ένα διάνυσμα βαρών με 4 γονίδια. Ο γενετικός αλγόριθμος ξεκινά με τυχαίους γονότυπους και αρχικό πληθυσμό 100 χρωμοσώματα. Για κάθε ενεργό χρήστη, ένα τυχαία επιλεγμένο χρωμόσωμα ανατίθεται και δοκιμάζεται από τη συνάρτηση καταλληλότητας. Η συνάρτηση καταλληλότητας μετρά την ακρίβεια πρόβλεψης των αντικειμένων π.χ ταινιών βασιζόμενη στο τρέχον χρωμόσωμα.

$$Accuracy(k)=10 - \frac{\Sigma_1 |P(k, j) - A(k, j)|}{l} \text{ , (4.6)}$$

4.7 Μέτρηση Ομοιότητας (Similarity Measure)

4.7.1 Collaborative Filtering Approach

Στο Collaborative Filtering (CF) η εξακρίβωση των χρηστών οι οποίοι έχουν παρόμοιες προτιμήσεις με έναν ενεργό χρήστη στο παρελθόν είναι πολύ σημαντικό στοιχείο για την επιτυχής εφαρμογή του (L.Liu,N.Mehandjiev,D.Xu,2011).
Η ποιο ευρέως διαδεδομένη προσέγγιση είναι αυτή του Pearson

$$sim(u,u') = \frac{\sum_{i\in I}(r_{u,i}-\overline{r_u})(r_{u',i}-\overline{r_{u'}})}{\sqrt{\sum_{i\in I}(r_{u,i}-\overline{r_u})^2}\sqrt{\sum_{i\in I}(r_{u',i}-\overline{r_{u'}})^2}} \quad (4.7)$$

Πηγή: L.Liu,N.Mehandjiev,D.Xu (2011), Multi-Criteria Service Recommendation Based on User Criteria Preferences

Όπου το I αντιπροσωπεύει την ομάδα των κοινών items που έχουν αξιολογήσει ο χρήστης u και ο χρήστης u'. Οι τιμές της προσέγγισης αυτής ποικίλουν από -1 μέχρι +1. Όσο μεγαλύτερη είναι η τιμή τόσο πιο όμοιοι είναι δυο χρήστες ως προς τις προτιμήσεις τους, ενώ με την τιμή +1 οι χρήστες έχουν ακριβώς την ίδια προτίμηση.
Σαν πρώτη προσέγγιση θεωρούμε ότι έχουμε μόνο ένα κριτήριο κάθε φορά και θα εξετάσουμε και τα 4 κριτήρια του προβλήματός μας ξεχωριστά κάθε φορά και όχι όλα μαζί στην CF προσέγγιση.

4.8 Μοντέλο Συστάσεων(RM)

Το μοντέλο αυτό προτείνει προιόντα στον τρέχοντα χρήστη λαμβάνοντας πληροφορίες από τούς γείτονες του. Η φάση σύστασης ταινιών ολοκληρώνεται με την εγκατάσταση τον συστήματος συνεργασίας. Το σύστημα συνεργασίας υπολογίζει την απόσταση μεταξύ δύο χρηστών για την ίδια ταινία με τον τύπο:

$$\text{duu'=Similarity(u,u')} \sqrt{\sum_{n=1}^{k+1} \left(r_{un} - r_{u'n}\right)^2} \qquad (4.8)$$

Πηγή: Using Genetic Algorithms for Personalized Recommendation Chein-Shung Hwang, Yi-Ching Su, and Kuo-Cheng Tseng, (2010).

όπού r_u είναι το διάνυσμα βαθμολογίας τον χρήστη u και r_u' το διάνυσμα βαθμολογίας τον χρήστη u'. Ο δείκτης n μετρά τα 5 κριτήρια στα οποία βαθμολογήθηκαν οι ταινίες από τούς χρήστες, αλλά και τη συνολική βαθμολογία.
Ο τύπος (5.4) είναι βασικός στην λειτουργία τον μοντέλου που αναπτύχθηκε στην παρούσα εργασία. Στον υπολογισμό της απόστασης μεταξύ δύο χρηστών συνυπολογίζεται η ομοιότητά τους (η οποία υπολογίστηκε νωρίτερα με τον γενετικό αλγόριθμο),γεγονός που προσδίδει στον τύπο αυτό μεγαλύτερη πληροφορία, καθιστώντας τον πιο αξιόπιστο. Ο συμβατικός τρόπος υπολογισμού της απόστασης μεταξύ δύο χρηστών δεν λαμβάνει υπόψην τις ομοιότητες των χρηστών καθιστώντας τον ελλιπή.
Επιπλέον υπολογίζεται η συνολική απόσταση μεταξύ των δύο χρηστών η οποία δίνεται από την παρακάτω εξίσωση :

$$\text{dist}_{(u,u')} = \frac{1}{|U(u,u')|} \sum_{i \in \text{U(u, u')}} duu' \qquad (4.9)$$

Πηγή: Using Genetic Algorithms for Personalized Recommendation Chein-Shung Hwang, Yi-Ching Su, and Kuo-Cheng Tseng, (2010).

όπού U(u,u') υποδεικνύει το σύνολο των ταινιών τις οποίες είδαν και οι δύο χρήστες. Άρα, η συνολική απόσταση μεταξύ των δύο χρηστών είναι η μέση απόσταση μεταξύ των βαθμολογιών τους για όλες τις κοινές ταινίες.

Τέλος, υπολογίζεται η ομοιότητα των χρηστών η οποία είναι αντιστρόφως ανάλογη της απόστασής τούς μέσω της ακόλούθης εξίσωσης :

$$sim_{(u,u')} = \frac{1}{1 + dist(u,u')} \qquad (4.10)$$

Πηγή : Using Genetic Algorithms for Personalized Recommendation Chein-Shung Hwang, Yi-Ching Su, and Kuo-Cheng Tseng, (2010).

Αυτός ο τρόπος υπολογισμού ομοιότητας εξασφαλίζει ότι η ομοιότητα Θα πλησιάζει το μηδέν όταν η απόσταση μεταξύ δύο χρηστών μεγαλώνει και θα πλησιάζει τη μονάδα αντιστρόφως.
Μετά τον υπολογισμού τον δείκτη ομοιότητας η επόμενη εξίσωση παρέχει μία δυνητική βαθμολογία R(u,i) για κάθε ταινία που δεν έχει δει ο τρέχων χρήστης:

$$R_{(u,i)} = \left(\frac{1}{\sum_{u' \in C(u)} sim(u,u')} \sum_{u' \in C(u)} sim(u,u') R(u',i) \right) \qquad (4.11)$$

Πηγή: Using Genetic Algorithms for Personalized Recommendation Chein-Shung Hwang, Yi-Ching Su, and Kuo-Cheng Tseng, (2010).

το οποίο είναι στην πραγματικότητα ο σταθμισμένος μέσος των γνωστών βαθμολογιών, ενώ το C(υ) καθορίζει την γειτονιά του χρήστη.

Ξέρουμε επίσης ότι η στατιστική ακρίβεια μετρά πόσο κοντά είναι μία αριθμητική τιμή r που παράγεται από το σύστημα συστάσεων και αντιπροσωπεύει την αναμενόμενη βαθμολογία του χρήστη u στην ταινία ί, με την πραγματική αριθμητική τιμή r_{uZ} όπως αυτή δίνεται από τον χρήστη u για την ίδια ταινία. Το περισσότερο χρησιμοποιημένο στατιστικό μέτρο ακρίβειας είναι το μέσο απόλυτο σφάλμα MAE (Mean Absolute Error). Μετράται μόνο για τις ταινίες για τις οποίες ο χρήστης u έχει εκφράσει την άποψή τον. Υποθέτοντας ότι ο χρήστης έχει εκφράσει άποψη για η ταινίες το μέσο απόλυτο σφάλμα δίνεται στην ακόλούθη εξίσωση:

$$MAE_u = \frac{1}{n} \sum_{i=1}^{n} | \ r_{ui} \ - r'_{ui} \ | \qquad (4.12)$$

Πηγή: Using Genetic Algorithms for Personalized Recommendation Chein-Shung Hwang, Yi-Ching Su, and Kuo-Cheng Tseng, (2010).

Το μέσο MAE για ολόκληρο το σύνολο δεδομένων μπορεί να υπολογιστεί με τον μέσο όρο τον MAE για όλους τούς χρήστες δίνοντας έτσι μία συνολική εκτίμηση για την απόδοση του μοντέλου.

ΚΕΦΑΛΑΙΟ 5

ΑΝΑΛΥΣΗ ΚΑΙ ΣΧΕΔΙΑΣΗ ΣΥΣΤΗΜΑΤΟΣ

5.1 Dataset

Στην εργασία αυτή προτείνεται ένα υβριδικό σύστημα συστάσεων που χρησιμοποιεί γενετικό αλγόριθμο και πολυκριτήρια μέθοδο για στάθμιση χαρακτηριστικών για τον εντοπισμό όμοιων χρηστών που μπορεί να έχουν τα ίδια ενδιαφέροντα με τον ενεργό χρήστη και για τον προσδιορισμό των δυνητικών αναγκών του χρήστη. Πιο συγκεκριμένα, μας δόθηκε ένα σύνολο 6078 χρηστών από τον κ.Ματσατσίνη και την κ.Λακιωτάκη, K.Lakiotaki, N. F. Matsatsinis, and A.Tsoukiàs, "Multi-Criteria User Modeling in Recommender Systems", *IEEE Intelligent Systems* οι οποίοι έχουν αξιολογήσει από μια ταινία ο καθένας σε τέσσερα κριτήρια. Τα 4 κριτήρια έχουν εισαχθεί σε ένα αρχείο txt για καθέναν από τους 6078 χρήστες. Σκοπός είναι η σύσταση ταινίας για έναν νέο χρήστη ο οποίο εισάγεται στο σύστημα. Τα δεδομένα αυτά ανακτήθηκαν από το Yahoo!Movies όπου οι χρήστες έδωσαν τις προτιμήσεις τους σε ταινίες βασιζόμενες σε 4 κριτήρια. Τα κριτήρια αυτά είναι η ηθοποιία, η πλοκή, η σκηνοθεσία και τα οπτικά εφέ. Ο κάθε χρήστης κατά μέσω όρο αξιολόγησε δέκα ταινίες.

5.2 Ανάλυση του Προβλήματος- Collaborative Filtering Προσέγγιση

Στις επόμενες παραγράφους θα περιγράψουμε τη δημιουργία ενός memory-based συστήματος. Στην συνέχεια θα εντάξουμε στο πρόβλημά μας και την multi-criteria προσέγγιση. Οι πλειονότητα των memory-based συστημάτων ανήκουν στην οικογένεια του Collaborative Filtering (CF) και στην ουσία χρησιμοποιούν γειτονικούς χρήστες καθώς και ομοιότητες μεταξύ των χρηστών για να κάνουν τις συστάσεις τους. Στην περίπτωσή μας αρχικά χρησιμοποιούμε τις γραμμές του πίνακα (αρχείου txt) ας τον ονομάσουμε R.

Τα κύρια μέρη του αλγόριθμου μας είναι :

- Η αναπαράσταση
- Η επιλογή γειτνίασης
- Οι γενιές των συστάσεων

Στην αναπαράσταση τα δεδομένα εισόδου καθορίζονται σαν μια συλλογή αξιολογήσεων (ratings), πραγματικών τιμών 6078 χρηστών για μια ταινία. Οι χρήστες δεν αξιλογούν όλες τις ταινίες αλλά μια και αυτό μπορεί να δημιουργήσει μη προσιτά αποτελέσματα για τον αλγόριθμό μας.

Στην επιλογή γειτνίασης η οποία είναι και το πιο σημαντικό βήμα για τη διαδικασία της σύστασης υπολογίζεται η ομοιότητα μεταξύ των χρηστών στον πίνακα R. Οι χρήστες οι οποίοι είναι όμοιοι με τον χρήστη α θα εφαρμόσουν μια επιλογή γειτνίασης με αυτόν. (αναφέρθηκε στο προηγούμενο κεφάλαιο). Ο σχηματισμός γειτνίασης (Neighnorhood Formation) γίνεται με δύο τρόπους :

Έχοντας τον πίνακα R με τους χρήστες και τις ταινίες η ομοιότητα ενός χρήστη α και ενός β μπορεί να υπολογιστεί με τις μεθόδους **Pearson Correlation Similarity** και **Cosine/Vector Similarity** οι οποίες είναι οι κύριες μεθόδοι για μέτρηση της ομοιότητας στα recommendation systems και αναφέρθηκαν εκτενώς στο προηγούμενο κεφάλαιο.

Στις γενιές των συστάσεων ξεχωρίζουμε των ενεργό χρήστη και τον χρήστη από τον οποίο θέλουμε να πάρουμε την πρόβλεψη και τον επιλέγουμε ως γείτονα. Αυτό γίνεται στην περίπτωσή μας με δύο τρόπους:

- Με την μέθοδο **Aggregate Neighborhood** scheme η οποία δεν βρίσκει τον κοντινότερο γείτονα αλλά συλλέγοντας τους χρήστες στο κέντρο της τρέχουσας ομάδας γειτόνων. Ειδικότερα η μέθοδος αυτή δημιουργεί μια γειτονία μήκους λ. Αρχικά διαλέγει τον χρήστη ο οποίος είναι πιο κοντά στον ενεργό μας χρήστη. Οι δύο αυτοί χρήστες αποτελούν την γειτονιά και η επιλογή του καινούριου γείτονα βασίζεται στην γειτονιά αυτήν.
- Με την μέθοδο **Center-based scheme** στην οποία δημιουργείται μια γειτονιά από τον ενεργό χρήστη. Εδώ

επιλέγονται από τον πίνακα similarity και ειδικότερα από τις γραμμές του οι χρήστες οι οποίοι έχουν την μεγαλύτερη ομοιότητα με τον χρήστη μας.

Το τελικό μέρος είναι η δημιουργία μιας πρόβλεψης η οποία θα αποτελείται από έναν πραγματικό αριθμό ή μια σύσταση η οποία θα είναι μια λίστα των top-N αντικειμένων για τα οποία ο χρήστης θα ενδιαφέρεται. Σε κάθε περίπτωση το αποτέλεσμα θα βασίζεται στην επιλογή γειτνίασης των χρηστών.

Όσον αφορά την πρόβλεψη είναι μια αριθμητική τιμή η οποία θα αναπαριστά την προβλεπόμενη γνώμη ενός ενεργού χρήστη για μια συγκεκριμένη ταινία. Η τιμή αυτή θα πρέπει να βρίσκεται στα αριθμητικά όρια του πινάκα που είχαμε σαν είσοδο.

5.3 Επέκταση αλγορίθμου για Πολυκριτηριακή ανάλυση

Ο σκοπός της πολυκριτηριακής ανάλυσης στο recommendation system της διπλωματικής αυτής εργασίας είναι να βρίσκει ταινίες οι οποίες μεγιστοποιούν την χρησιμότητα του κάθε χρήστη όπως και στα single-rating συστήματα παραπάνω. Γι’ αυτό είναι χρήσιμος ο υπολογισμός ενός συνολικού rating της ταινίας για κάθε χρήστη διότι το σύστημα βασίζεται σε αυτές τις αξιολογήσεις για να προτείνει τις κατάλληλες ταινίες στους χρήστες. Η διαφορά εδώ είναι ότι έχουμε πολύ παραπάνω πληροφορία όσον αφορά τους χρήστες και τις ταινίες λόγω των πολλαπλών κριτηρίων, γεγονός το οποίο βοηθά το σύστημά ας για καλύτερες συστάσεις.

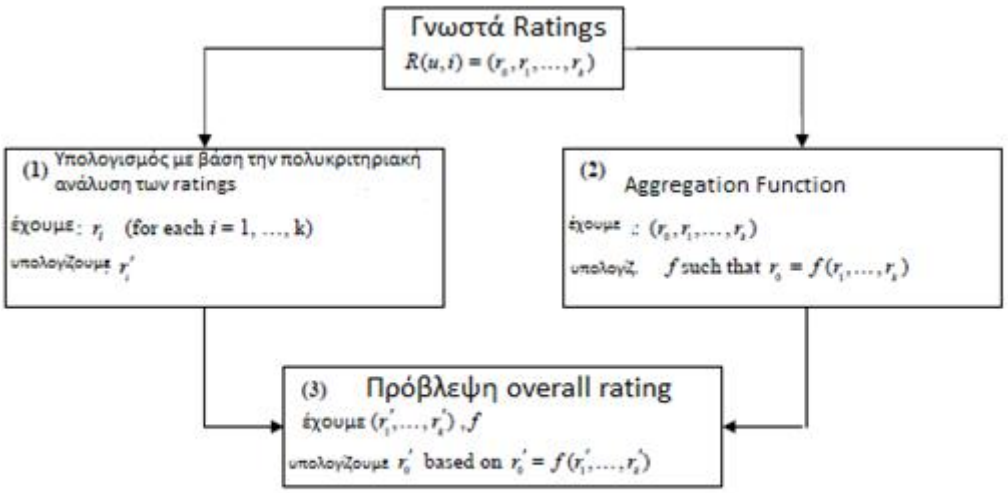
Ο πίνακας R είναι ίδιος με την προηγούμενη περίπτωση μόνο που εδώ προσθέτονται όλα τα κριτήρια καθώς επίσης και το overall rating. Στην περίπτωσή μας έχουμε 6078 χρήστες οι οποίοι αξιολογούν μια ταινία με τέσσερα κριτήρια. Ας υποθέσουμε ότι ο χρήστης 6079 δεν έχει αξιολογήσει την ταινία. Για να προβλέψουμε την αξιολόγηση του τελευταίου χρήστη μας το σύστημά μας θα πρέπει να βρει τους χρήστες τους οποίους είναι πιο κοντά με τον χρήστη μας και έχουν δει τη συγκεκριμένη ταινία. Υπάρχει η πιθανότητα μερικοί χρήστες που το overall rating τους ταιριάζει στην περίπτωσή μας να έχουν διαφορετικά γούστα από τον χρήστη που θέλουμε να κάνουμε την σύσταση. Αυτή

είναι μια περίπτωση που εξετάζει το multi-criteria system σε αντίθεση με το single-criteria το οποίο θα την υποσκέλιζε.

Στην παρούσα διπλωματική εργασία υλοποιήσαμε τον αλγόριθμο του collaborative συστήματος σε matlab που δημιουργήσαμε μόνοι μας ώστε να προσαρμόζεται στα multi-criteria ratings που είχαμε. Όπως αναφέραμε και πιο πάνω εφαρμόσαμε αρκετούς τρόπους εύρεσης όμοιων χρηστών. Αυτό αναφέρεται στην μεταβλητή similarity στον κώδικά μας. Στην συνέχεια του κεφαλαίου αυτού θα δείξουμε τι γίνεται όταν εντάσσουμε βάρη στα κριτήριά μας στην πολυκριτηριακή ανάλυση.

5.4 Aggregation Function

Στην ενότητα αυτή του κεφαλαίου μας παρουσιάζουμε μια διαφορετική προσέγγιση. Αρχικά μετατρέπουμε τον πίνακα μας με τις αξιολογήσεις των χρηστών σε 4, όσα και τα κριτήριά μας, single-rating προβλήματα και χρησιμοποιούμε την τεχνική του απλού collaborative filtering για να προσεγγίσουμε τις εκτιμήσεις για κάθε κριτήριο ξεχωριστά. Στην συνέχεια βρίσκουμε την Aggregation Function μέσω στατιστικών από το ήδη γνωστά ratings. Τέλος σε συνδιασμό με τα πολυκριτηριακά δεδομένα και την συνάρτηση που υπολογίσαμε υπολογίζουμε το overall rating. Σε μορφή διαγράμματος η μέθοδος αυτή δίνεται στο Σχήμα 22.



Σχήμα 22
προσέγγιση Aggregation Function

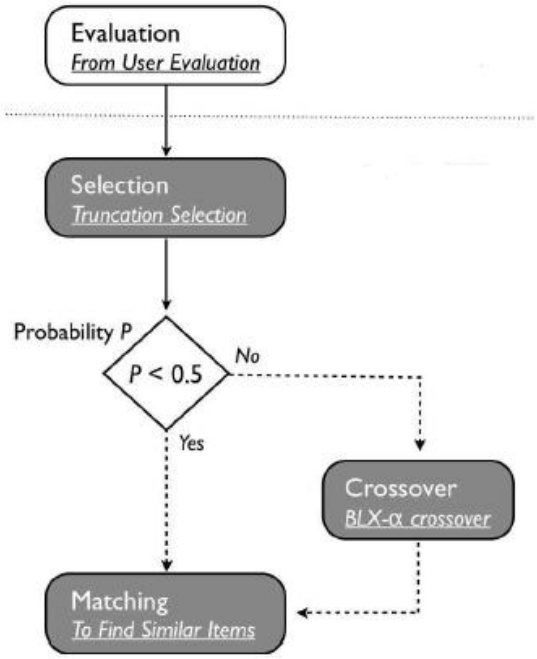
5.5 Σύνοψη Συστήματος

Το σύστημα συστάσεων που περιγράφεται στην παρούσα διπλωματική εργασία βασίζεται στους γενετικούς αλγορίθμους. Η τεχνική content-based filtering που εφαρμόσαμε χρησιμεύει στην παραγωγή του αρχικολυ πληθυσμού του γενετικού μας αλγορίθμου. Υπάρχει η fitness function για κάθε ταινία του συστήματος. Για κάθε χρήστη έχουμε τα ratings για την ταινία σύμφωνα με τις προτιμήσεις του για κάθε κριτήριο της ταινίας. Το recommendation system εξελίσει ουσιαστικά έναν αρχικό πληθυσμό βασιζόμενα στα δεδομένα των χρηστών μέχρι να εκπληρώσει ένα κριτήριο τερματισμού.

Η επιλογή (selection) γίνεται με τον εξής απλό τρόπο: αν μια ταινία έχει χαμηλή αξιολόγηση δεν θα μπορεί να μπορεί να προταθεί ώστε να κάνει κάποια σύσταση το σύστημα. Μόνο ένας αριθμός των καλύτερων περιπτώσεων θα προταθεί οι οποίες έχουν υψηλό fitness value.

Μετά το selection τα μισά από τα αντικείμενα θα περάσουν στο crossover. Εδώ βρίσκουμε όμοια αντικείμενα από τα εξαγόμενα των προηγούμενων μεθόδων και τα εναπομείναντα στο σύστημα. Η εύρεση της ομοιότητας έχει αναλυθεί εκτενώς στα προηγούμενα κεφάλαια. Η

επόμενη φάση του αλγορίθμου είναι το mutation. Σχηματικά είναι το εξής:



Σχήμα 23

Εισάγουμε τώρα στον γενετικό μας αλγόριθμο με προτεραιότητα αρχικά να επιλέξουμε (selection) τις αντίστοιχες βέλτιστες λύσεις. Δίνουμε στην μεταβλητή wij random τιμές που κινούνται ανάμεσα στον αριθμό μηδέν και ένα. Στη συνέχεια έχοντας δημιουργήσει ζεύγη εναλλακτικών επιλογών μεγιστοποιούμε κάθε φορά τα wij ενός από τα τέσσερα κριτήριά μας κάθε φορά δίνοντας τους τιμή ένα και κρατάμε την διαφορά που δημιουργήθηκε μεταξύ των εναλλακτικών που δημιουργήθηκε μεταξύ των εναλλακτικών που συγκρίναμε για το αντίστοιχο κριτήριο και τον αντίστοιχο χρήστη δημιουργώντας τιμές καταλληλότητας (fitness values). Παράλληλα κρατάμε πληροφορία για τους χρήστες τις εναλλακτικές επιλογές και τα κριτήρια αξιολόγησης που τώρα σαν τιμές έχουν κρατήσει τα νέα wij. Η fitness function αξιολογείται για κάθε ομάδα παρέχοντας τα βέλτιστα wij που στην συνέχεια κανονικοποιούνται. Κανονικοποίηση (normalization) σημαίνει η

διαίρεση των τιμών καταλληλότητας του κάθε wij με το άθροισμα όλων των τιμών καταλληλότητας έτσι ώστε το άθροισμα αυτών που προκύπτει να είναι ίσο με 1. Στην υλοποίηση της εργασίας αυτής κρατάμε έναν πίνακα Fs στον οποίο αποθηκεύουμε κάθε φορά τις τιμές των αντικειμενικών συναρτήσεων για όλες τις λύσεις. Θέτοντας λοιπόν με 1 τις τιμές των αντίστοιχων wij θα μειωθούν απ' τις τιμές των διαφορών των εναλλακτικών επιλογών και θα μας διαμορφώσουν το διάνυσμα νέων wij λύσεων παιδιών που θα αποτελέσει με την σειρά του το διάνυσμα λύσεων γονέων.

Κάθε ενεργός χρήστης k επιλέγει τυχαία 1 προιόντα που, στην περίπτωσή μας, l=1. 0 συντελεστής P(k,j) αναπαριστά τις προβλεπόμενες βαθμολογίες του χρήστη *k* για το προιόν j, ενώ ο συντελεστής A(k,j) αναπαριστά τις πραγματικές βαθμολογίες που έχει δώσει ο τρέχων χρήστης. Οι προβλεπόμενες βαθμολογίες υπολογίστηκαν μέσω του γενετικού αλγόριθμού.

Στους γενετικούς αλγορίθμους είναι διαδεδομένοι οι τελεστές της μετάλλαξης (mutation) και της διασταύρωσης (crossover). Στο μεγαλύτερο κομμάτι της υλοποίησης κάναμε χρήση της διασταύρωσης ενός σημείου (one-point crossover) και διασταύρωσης δύο σημείων (two point crossover). Αυτό σημαίνει στην πρώτη περίπτωση ένα ενιαίο σημείο διαλέγεται από τις σειρές των λύσεων των γονέων και από κει και πέρα ανταλλάσσονται οι λύσεις με το προκύπτων διάνυσμα να αποτελεί τις λύσεις των παιδιών. Στην $2^{η}$ περίπτωση εκτός από το σημείο έναρξης της διασταύρωσης θέτουμε και σημείο τερματισμού της διασταύρωσης με όλα τα ενδιάμεσα στοιχεία να εναλλάσσονται και να μας παράγουν το νέο διάνυσμα που αποτελεί τις λύσεις των παιδιών.

Σχήμα 24

Crossover 1 point
Πηγή : google

Σχήμα 25

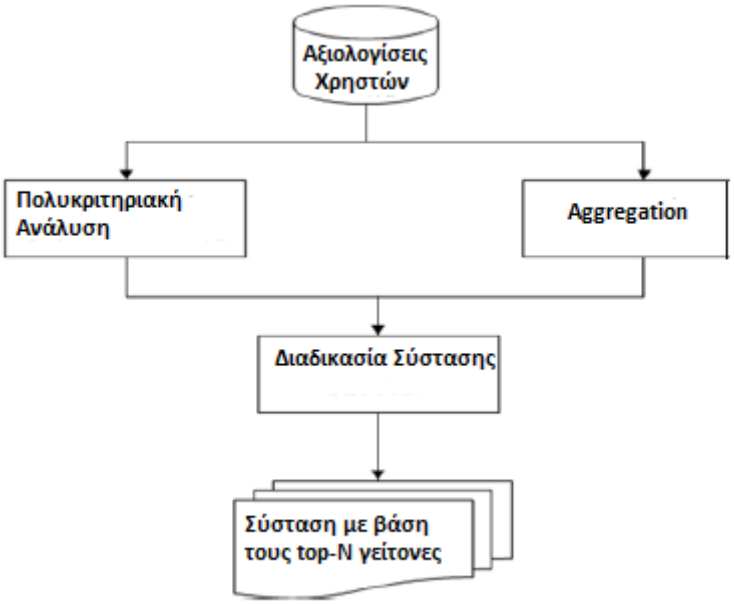
Crossover 2 point
Πηγή : goggle

Σε κάθε εκτέλεση του αλγορίθμου και καθώς περνάμε από γενιά σε γενιά έχουμε μια κατανομή των λύσεων σε συγκεκριμένα σημεία. Μας ενδιέφερε να βρίσκουμε την απόσταση των βέλτιστων λύσεων από εκείνες που είναι συνωστισμένες σε αυτά τα σημεία.
Αν οι τιμές είναι μικρότερες από ένα κατώφλι το οποίο έχουμε ορίσει εμείς τότε η λύση αποφασίζεται σαν μερικώς βέλτιστη και μεγιστοποιεί την πιθανότητα να επιλεγεί σαν επόμενος γονέας για την επόμενη γενιά λύσεων.
Σε αυτό το σημείο κάνουμε χρήση τελεστών αντικατάστασης (replacement) και μετάλλαξης (mutation) με πολύ μικρότερη πιθανότητα.
Οι τελευταίοι τελεστές αντικατάστασης εφαρμόζονται σε λύσεις οι

οποίες είχαν απορριφθεί στην προηγούμενη φάση. Με αυτόν τον τρόπο δίνουμε μια πολύ μικρή πιθανότητα κάποιες απ' τις προηγούμενες λύσεις να έχουν απορριφθεί λόγω σφαλμάτων.

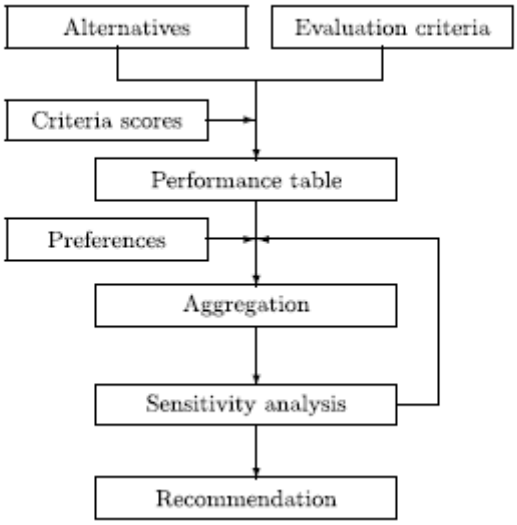
Αυτή η διαδικασία συνεχίζεται μέχρι να παρατηρήσουμε σύγκλιση των τιμών σε συγκεκριμένες τιμές των wij που αργότερα θα χαρακτηριστούν ως ολικά βέλτιστες. Σύγκλιση τιμών διαπιστώνεται μετά από περίπου 150 γενεές.

Ο αλγόριθμος συνεχίζει να εξελίσσεται μέχρι την ικανοποίηση κάποιου κριτηρίου τερματισμού. Στην περίπτωσή μας το κριτήριο τερματισμού ορίστηκε ως 150 γενεές. Για κάθε εξέλιξη γενιάς, τα χρωμοσώματα για την επόμενη γενιά επιλέγονται με χρήση της μεθόδού roulette wheel sel ection έτσι ώστε να εξασφαλιστεί αναλογικά τυχαία επιλογή. 'Ολα τα χρωμοσώματα μετά συνδυάζονται χρησιμοποιώντας διασταύρωση μονού σημείού με πιθανότητα 0.95. Μετά τη διασταύρωση, ακολουθεί η διαδικασία της μετάλλαξης κατά την οποία τα γονίδια τον κάθε χρωμοσώματος προσαρμόζονται έτσι ώστε κάθε χρωμόσωμα να αθροίζει στη μονάδα, αντιπροσωπεύοντας τα βάρη που δίνει ο χρήστης στα τέσσερα κριτήρια με πιθανότητα.



Σχήμα 27

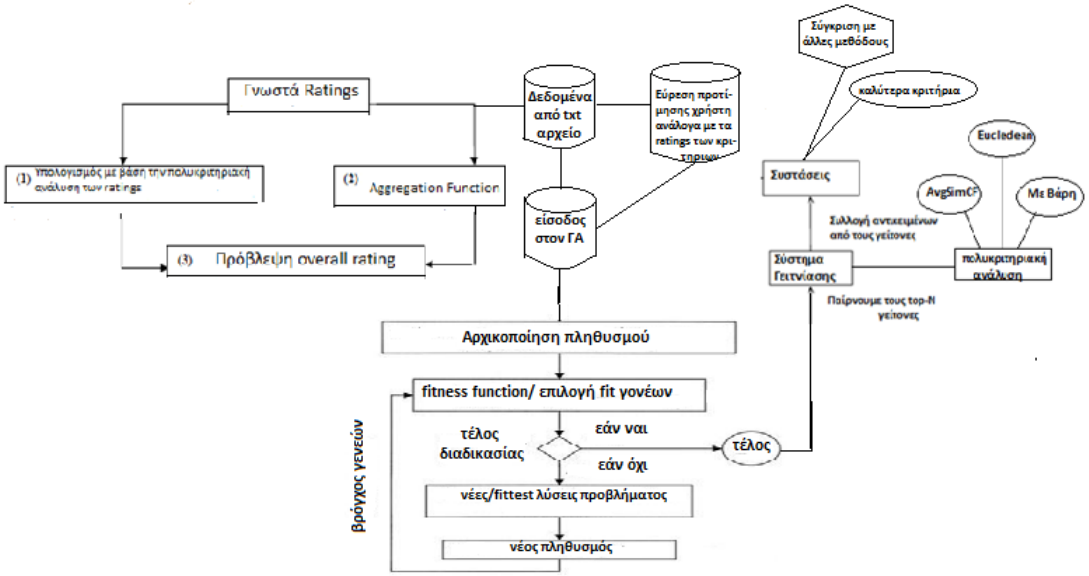
Γενικό πλάνο



Σχήμα 28

Ειδικότερο πλάνο

5.6 Σχήμα Αρχιτεκτονικής Συστήματος



Σχήμα 29

Αρχιτεκτονική Συστήματος

ΚΕΦΑΛΑΙΟ 6

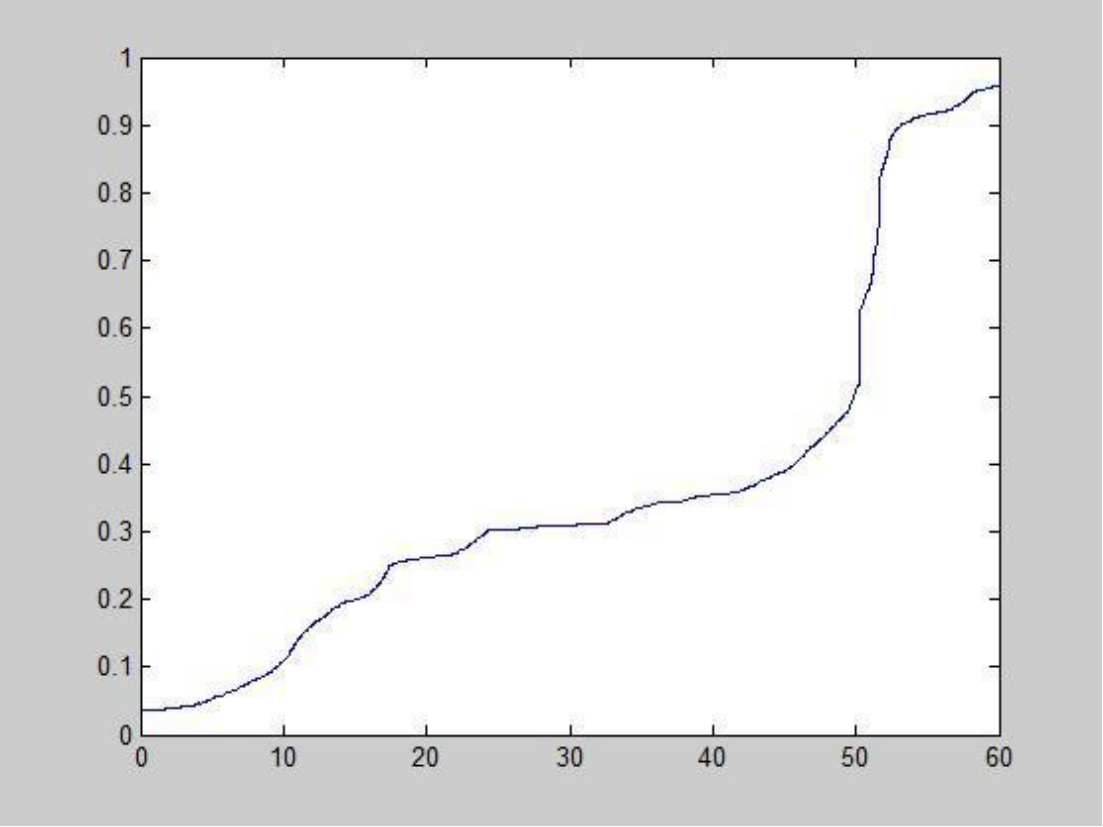
ΠΑΡΟΥΣΙΑΣΗ ΣΥΣΤΗΜΑΤΟΣ ΜΕΣΩ ΕΦΑΡΜΟΓΗΣ ΣΕ ΠΡΑΓΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ

Αρχικά δουλεύουμε για κάθε ένα χρήστη ξεχωριστά. Για τον κάθε χρήστη αυτόν θα γίνουν μια σειρά υπολογισμών και στο τέλος θα γίνει μια γενικευμένη εκτίμηση για το σύνολο των χρηστών.
Σαν πρώτο βήμα πρέπει να επιλέξουμε μέσω της διαδικασίας επιλογής (selection) τις αντίστοιχες βέλτιστες λύσεις. Δίνουμε στην μεταβλητή wij random τιμές που κινούνται ανάμεσα στον αριθμό μηδέν και ένα. Στη συνέχεια έχοντας δημιουργήσει ζεύγη εναλλακτικών επιλογών μεγιστοποιούμε κάθε φορά τα wij ενός από τα τέσσερα κριτήριά μας κάθε φορά δίνοντας τους τιμή ένα και κρατάμε την διαφορά που δημιουργήθηκε μεταξύ των εναλλακτικών που δημιουργήθηκε μεταξύ των εναλλακτικών που συγκρίναμε για το αντίστοιχο κριτήριο και τον αντίστοιχο χρήστη δημιουργώντας τιμές καταλληλότητας (fitness values). Παράλληλα κρατάμε πληροφορία για τους χρήστες τις εναλλακτικές επιλογές και τα κριτήρια αξιολόγησης που τώρα σαν τιμές έχουν κρατήσει τα νέα wij. Η fitness function αξιολογείται για κάθε ομάδα παρέχοντας τα βέλτιστα wij που στην συνέχεια κανονικοποιούνται. Κανονικοποίηση (normalization) σημαίνει η διαίρεση των τιμών καταλληλότητας του κάθε wij με το άθροισμα όλων των τιμών καταλληλότητας έτσι ώστε το άθροισμα αυτών που προκύπτει να είναι ίσο με 1. Στην υλοποίηση της εργασίας αυτής κρατάμε έναν πίνακα Fs στον οποίο αποθηκεύουμε κάθε φορά τις τιμές των αντικειμενικών συναρτήσεων για όλες τις λύσεις. Θέτοντας λοιπόν με 1 τις τιμές των αντίστοιχων wij θα μειωθούν απ' τις τιμές των διαφορών των εναλλακτικών επιλογών και θα μας διαμορφώσουν το διάνυσμα νέων wij λύσεων παιδιών που θα αποτελέσει με την σειρά του το διάνυσμα λύσεων γονέων.

Σε κάθε εκτέλεση του αλγορίθμου και καθώς περνάμε από γενιά σε γενιά έχουμε μια κατανομή των λύσεων σε συγκεκριμένα σημεία. Μας

ενδιέφερε να βρίσκουμε την απόσταση των βέλτιστων λύσεων από εκείνες που είναι συνωστισμένες σε αυτά τα σημεία.
Αν οι τιμές είναι μικρότερες από ένα κατώφλι το οποίο έχουμε ορίσει εμείς τότε η λύση αποφασίζεται σαν μερικώς βέλτιστη και μεγιστοποιεί την πιθανότητα να επιλεγεί σαν επόμενος γονέας για την επόμενη γενιά λύσεων.
Σε αυτό το σημείο κάνουμε χρήση τελεστών αντικατάστασης (replacement) και μετάλλαξης (mutation) με πολύ μικρότερη πιθανότητα.
Οι τελευταίοι τελεστές αντικατάστασης εφαρμόζονται σε λύσεις οι οποίες είχαν απορριφθεί στην προηγούμενη φάση. Με αυτόν τον τρόπο δίνουμε μια πολύ μικρή πιθανότητα κάποιες απ᾽ τις προηγούμενες λύσεις να έχουν απορριφθεί λόγω σφαλμάτων.

Αυτή η διαδικασία συνεχίζεται μέχρι να παρατηρήσουμε σύγκλιση των τιμών σε συγκεκριμένες τιμές των wij που αργότερα θα χαρακτηριστούν ως ολικά βέλτιστες. Σύγκλιση τιμών διαπιστώνεται μετά από περίπου 150 γενεές.



Σχήμα 30
Κατανομή βέλτιστων w
[115]

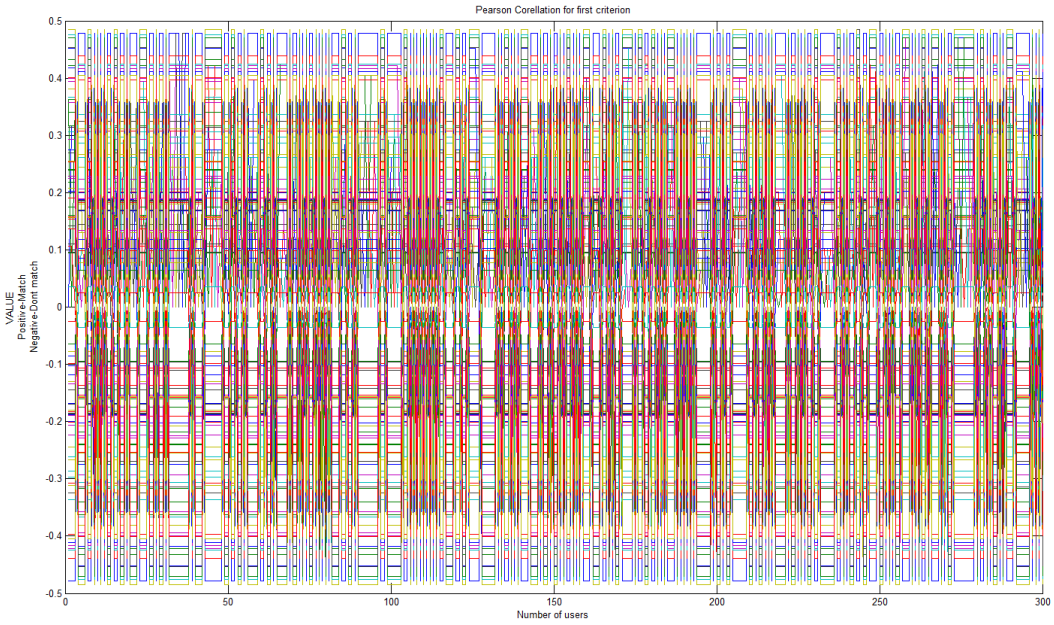
Ο αλγόριθμος συνεχίζει να εξελίσσεται μέχρι την ικανοποίηση κάποιου κριτηρίου τερματισμού. Στην περίπτωσή μας το κριτήριο τερματισμού ορίστηκε μέχρι 150 γενεές. Για κάθε εξέλιξη γενιάς, τα χρωμοσώματα για την επόμενη γενιά επιλέγονται με χρήση της μεθόδού roulette wheel sel ection έτσι ώστε να εξασφαλιστεί αναλογικά τυχαία επιλογή. 'Ολα τα χρωμοσώματα μετά συνδυάζονται χρησιμοποιώντας διασταύρωση μονού σημείού με πιθανότητα 0.95. Μετά τη διασταύρωση, ακολουθεί η διαδικασία της μετάλλαξης κατά την οποία τα γονίδια τον κάθε χρωμοσώματος προσαρμόζονται έτσι ώστε κάθε χρωμόσωμα να αθροίζει στη μονάδα, αντιπροσωπεύοντας τα βάρη που δίνει ο χρήστης στα τέσσερα κριτήρια με πιθανότητα.

6.1 Collaborative Filtering χωρίς πολυκριτηριακή ανάλυση

Σαν πρώτο βήμα του γενετικού μας αλγορίθμου στα παρακάτω σχήματα θα παρουσιάσουμε τα αποτελέσματα θεωρώντας ότι ο πίνακας εισόδου δεδομένων μας είναι ένας πίνακας με 6078 χρήστες που έχουν αξιολογήσει έναν αριθμό με βάση κάθε φορά ένα κριτήριο αντί για τέσσερα που ορίζει το πρόβλημά μας. Η προσπάθειά μας αυτή χρησιμεύει στην σύγκριση των αποτελεσμάτων μεταξύ της πολυκριτηριακής ανάλυσης στο πρόβλημά μας και στην απλή προσέγγιση σύστασης ενός recommendation system. Εφαρμόζοντας τον γενετικό μας αλγόριθμο θεωρώντας ότι υπάρχει μόνο ένα κριτήριο (πρώτη στήλη του αρχείου txt εισόδου των δεδομένων μας) έχουμε τις παρακάτω γραφικές παραστάσεις για τους πρώτους τριακόσιους χρήστες του προβλήματός μας. Θεωρούμε ως είσοδο τριακόσιους χρήστες διότι με τον συγκεκριμένο αριθμό χρηστών μπορούμε να έχουμε τα ίδια σχεδόν αποτελέσματα με το σύνολο των 6078 χρηστών λόγο σύγκλισης του αλγορίθμου.

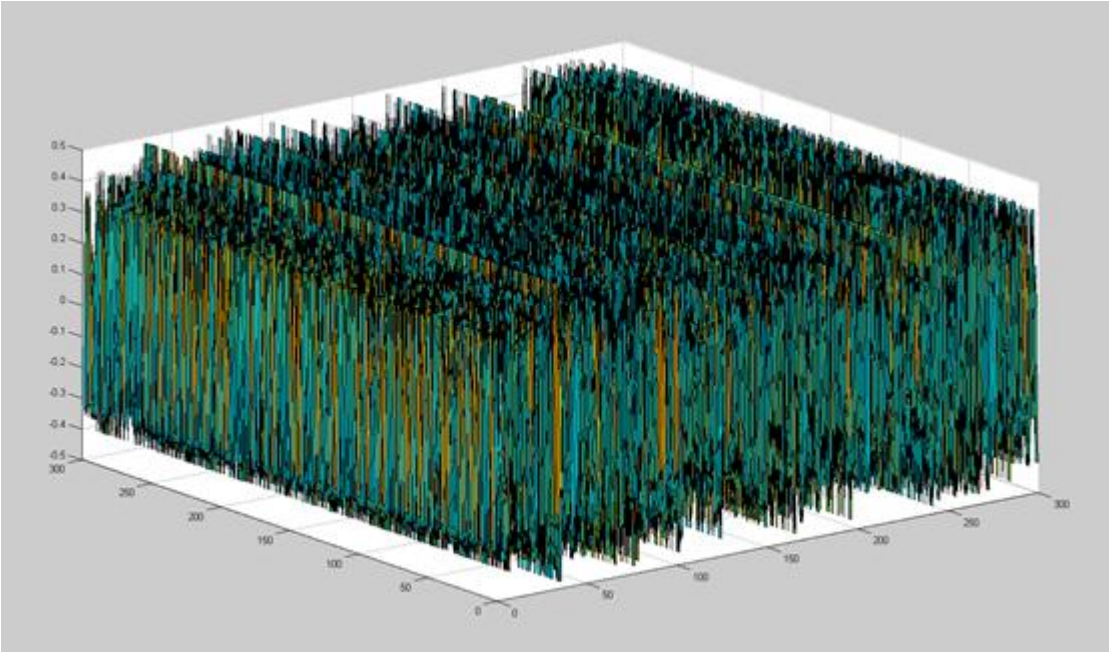
Στην περίπτωση της μη πολυκριτηριακής ανάλυσης και θεωρώντας σαν μοναδικό κριτήριο μόνο το πρώτο σαν το μοναδικό.

Για το πρώτο κριτήριο έχουμε:



Σχήμα 31

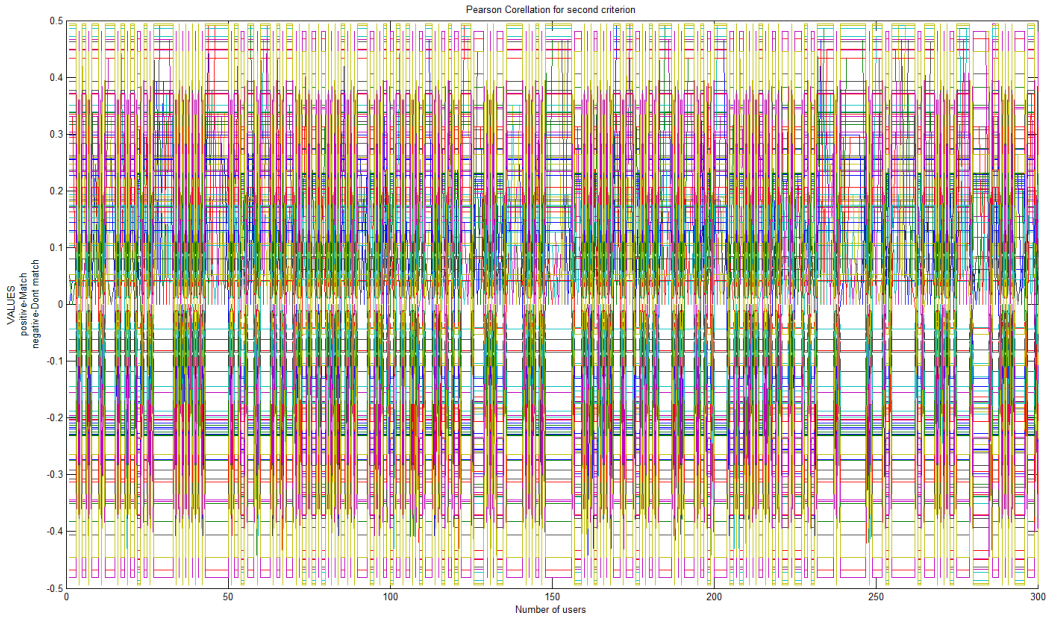
Και σε μορφή τρισδιάστατου πίνακα:



Σχήμα 32

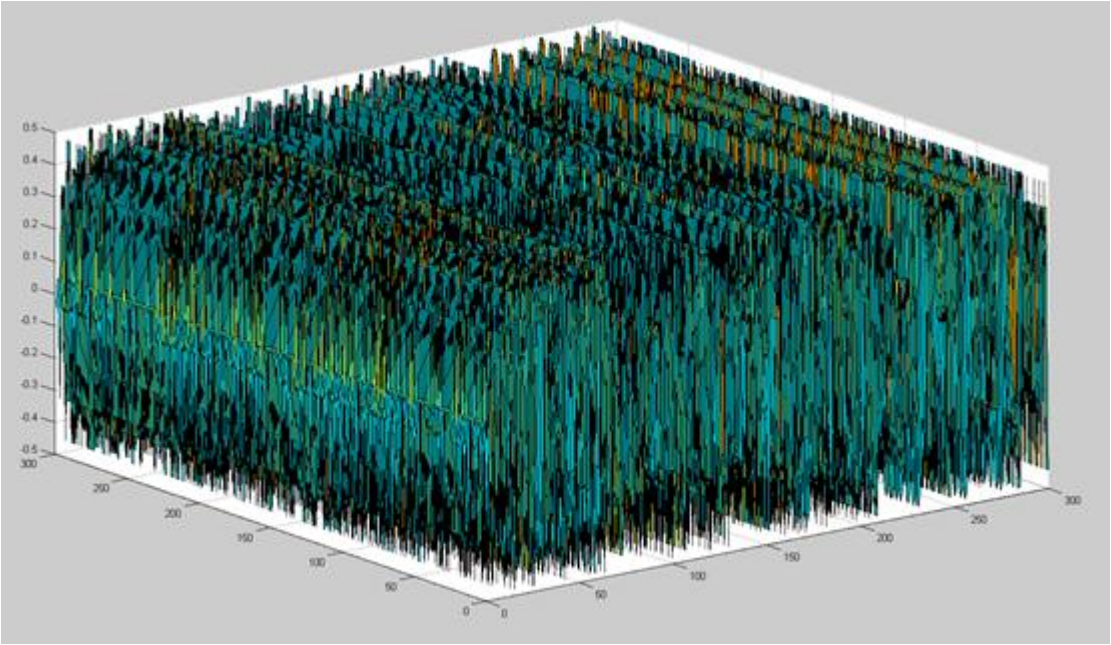
Στα παραπάνω δύο σχήματα βλέπουμε την ομοιότητα των χρηστών μεταξύ τους με βάση τις προτιμήσεις τους με βάση μόνο το πρώτο κριτήριο.Στην συνέχεια θα αναλύσουμε τη διαφορά της πολυκριτηριακής ανάλυσης και το πόσο επηρεάζει την ακρίβεια μιας σύστασης. Όπως είναι λογικό αν κάποιος χρήστης αξιολογεί με βάση 4 κριτήρια αντί για ένα, η παραπάνω πληροφορία μπορεί να δώσει πιο ακριβή αποτελέσματα.

Για το δεύτερο κριτήριο έχουμε:



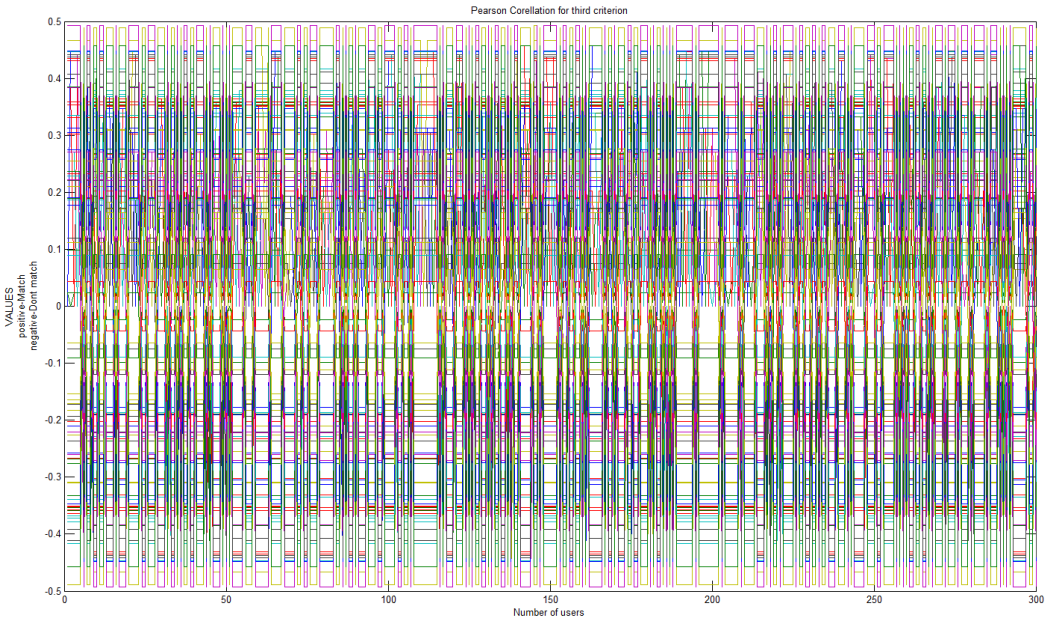
Σχήμα 33

Και σε μορφή τρισδιάστατου πίνακα:



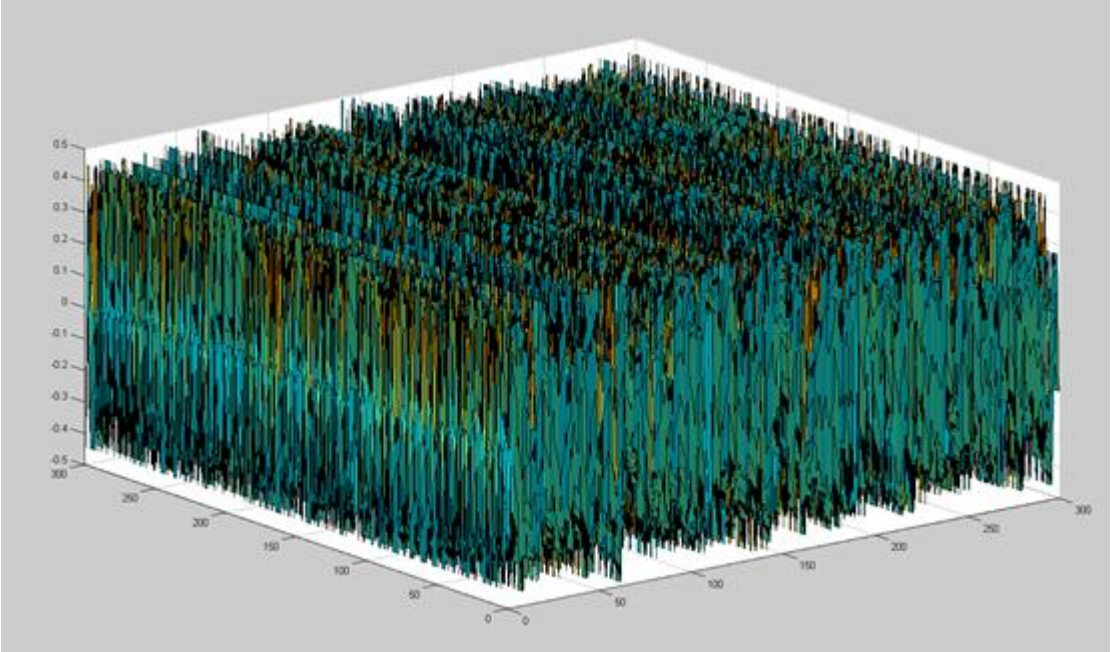
Σχήμα 34

Για το τρίτο κριτήριο έχουμε:



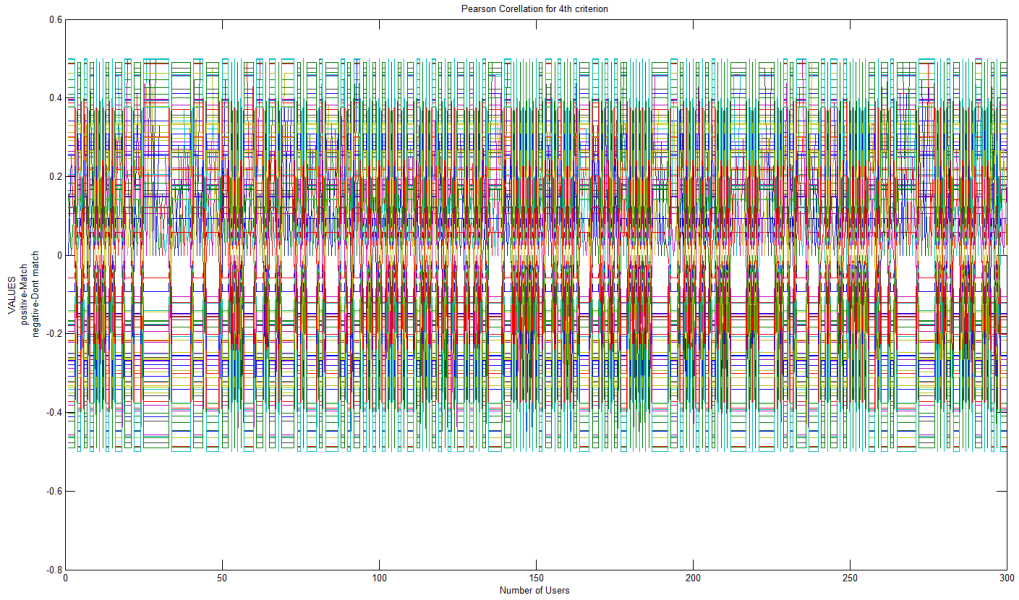
Σχήμα 35

Και σε μορφή τρισδιάστατου πίνακα:



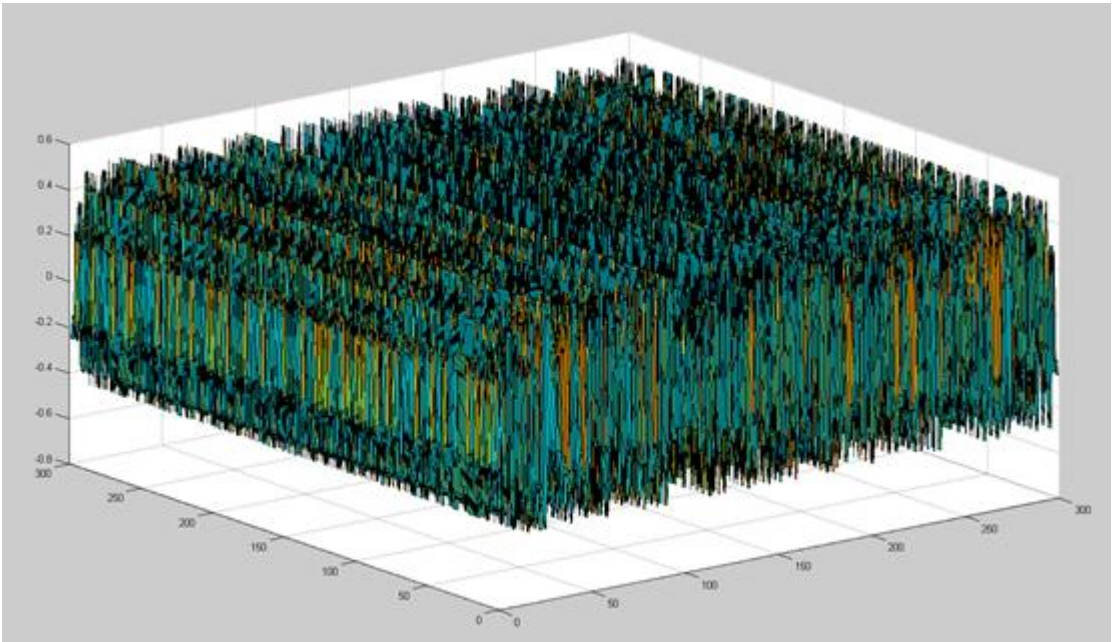
Σχήμα 36

Τέλος για το 4^{ο} κριτήριο έχουμε:



Σχήμα 37

Και σε μορφή τρισδιάστατου πίνακα:



Σχήμα 38

Με βάση το 4^{ο} κριτήριο βλέπουμε στο Σχήμα 38 ότι ο τρισδιάστατος πίνακας είναι περισσότερο συμπυκνωμένος από ότι στους

προηγούμενους στους οποίους εξετάζαμε τα πρώτα τρία κριτήρια και κυμαίνεται σε τιμές λίγο μεγαλύτερες απ΄ότι τις 0,6 -0,4 που βλέπουμε στα άλλα κριτήρια.. Αυτό μας παρέχει την πληροφορία ότι με βάση το 4^{ο} κριτήριο μπορούμε να κάνουμε μια καλύτερη πρόβλεψη, δηλαδή μας δίνει περισσότερη πληροφορία από ότι τα προηγούμενα 3.

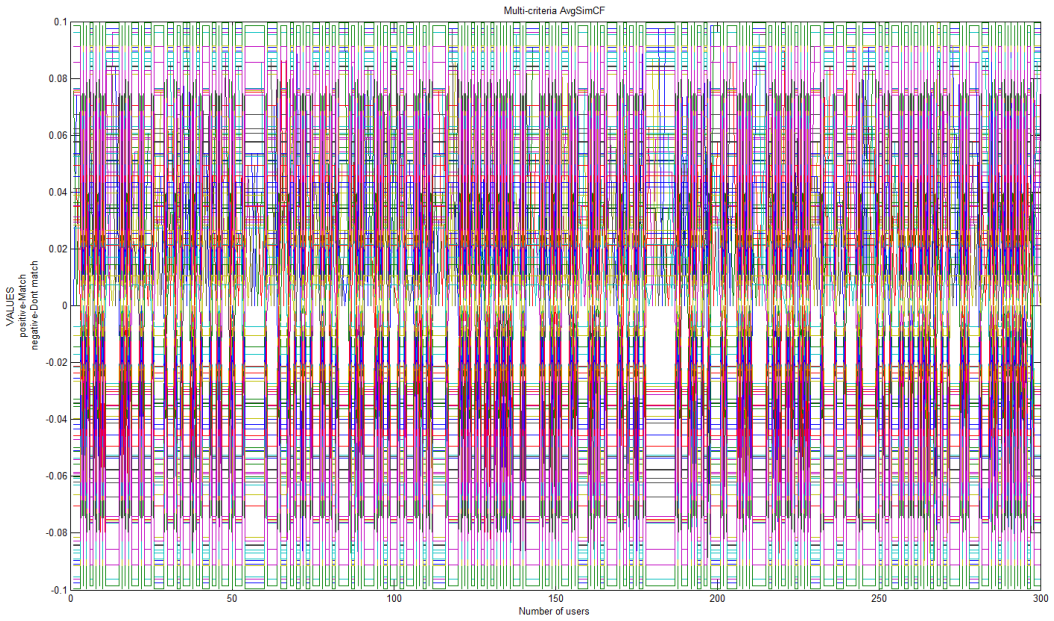
Και σε κώδικα matlab έχουμε για την παραπάνω προσέγγιση:

```
%%%pearson with one critirion
initialno1critirion=table(:,1);
no1critirion=W(:,1);
for i=1:NU
    for k=1:NU

        average(i,1)=average(i,1)+table(i,j)/4;
        if(i~=k)
        averge_pearson(i,1)= (W(i,4)-0.average)^2;
        averge_pearson2(k,1)= (W(k,4)-0.average)^2;

    Similarities4(i,k)= ((W(i,1)-average)*(W(k,1)-
average))/(sqrt(average_pearson(i,1)*sqrt(average_pearson2(k,1)))) ;
        end
    end
    end
```

6.2 Multi-criteria Approach

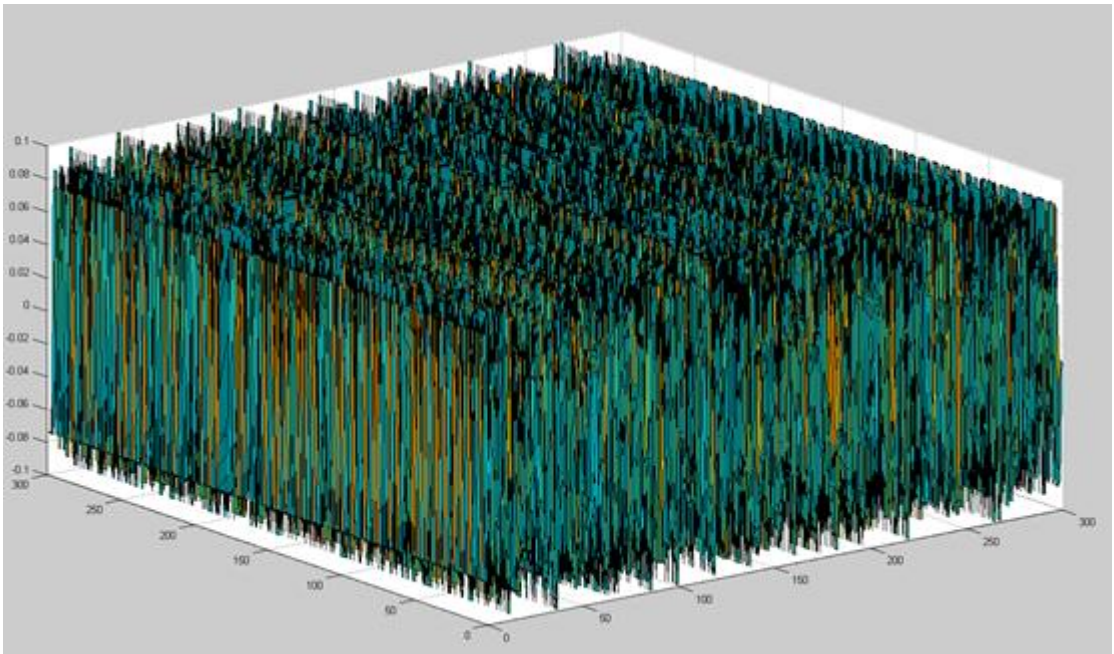
Στην περίπτωση μας με πολυκριτηριακή ανάλυση δεδομένων, εξετάζουμε στον γενετικό μας αλγόριθμο όχι μόνο 1 αλλά και τα 4 κριτήρια του προβλήματός μας ταυτόχρονα. Με βάση την πολυκριτηριακή ανάλυση δεδομένων συγκρίνουμε τρεις μεθόδους για της οποίες παραθέτουμε τις γραφικές παραστάσεις που προκύπτουν από το εργαλείο matlab στα παρακάτω σχήματα.

Οι γραφικές παραστάσεις που προκύπτουν είναι οι εξής:



Σχήμα 39

Και σε μορφή τρισδιάστατου πίνακα:



Σχήμα 40

Σε αυτές τις γραφικές παραστάσεις έχουμε να εξετάσουμε μεταξύ των

χρηστών αντί για ένα, τέσσερα κριτήρια. Είναι εύκολο να παρατηρήσουμε την απόκλιση για την πολυκριτηριακή ανάλυση στα παραπάνω δύο σχήματα. Μεταξύ των χρηστών οι διαφορές είναι πολύ μικρότερες και κυμαίνονται μεταξύ των τιμών 0,1 και -0,08. Αυτό μας δίνει την πληροφορία ότι ο γενετικός μας αλγόριθμος ο οποίος έχει επεκταθεί ώστε να καλύπτει τα πολλαπλά κριτήρια του προβλήματός μας βρίσκει τους χρήστες μέσω της λειτουργίας τους με πιο όμοιες προτιμήσεις από ότι μόνο με ένα κριτήριο. Έτσι είναι ευκολότερο να προτείνει το σύστημα μια ταινία σε έναν χρήστη με βάση τις προτιμήσεις χρηστών με τα ίδια γούστα.

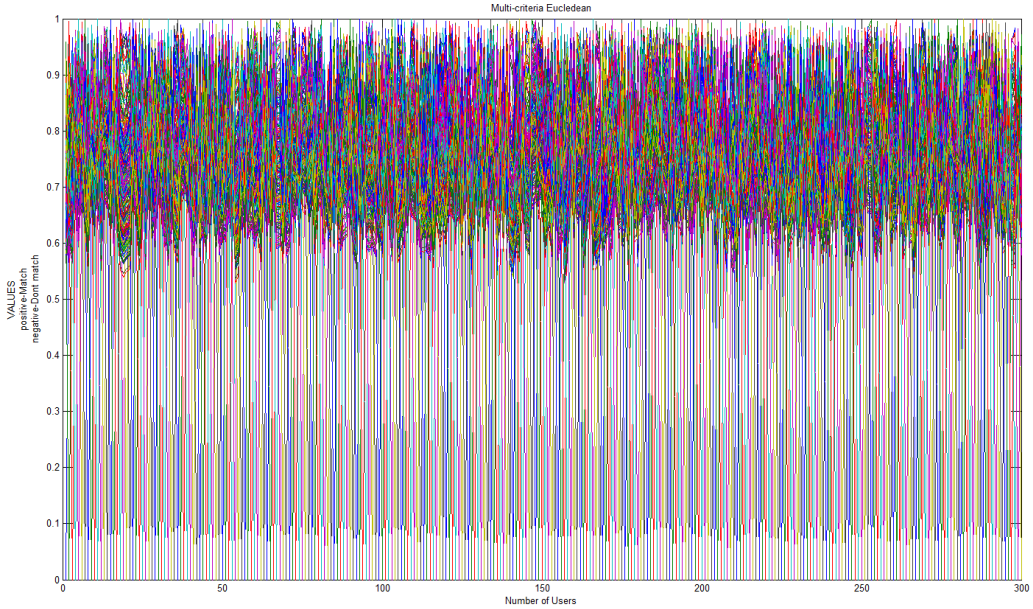
Και σε κώδικα matlab έχουμε για την παραπάνω προσέγγιση:

```
NU=6078;
sum=zeros(NU);
riza=zeros(NU); % The factor under the root
%Similarities=zeros(NU); % Similarity index
for i=1:NU
for j=1:NU
if i==j
sum(i,j)=1;
else
for k=1:4
sum(i,j)=sum(i,j)+((W(i,k)-W(j,k))^2);
end
riza(i,j)=sqrt(sum(i,j));
Similarities1(i,j)=1-riza(i,j);
end
end
end
```

6.3 EuclideanCF

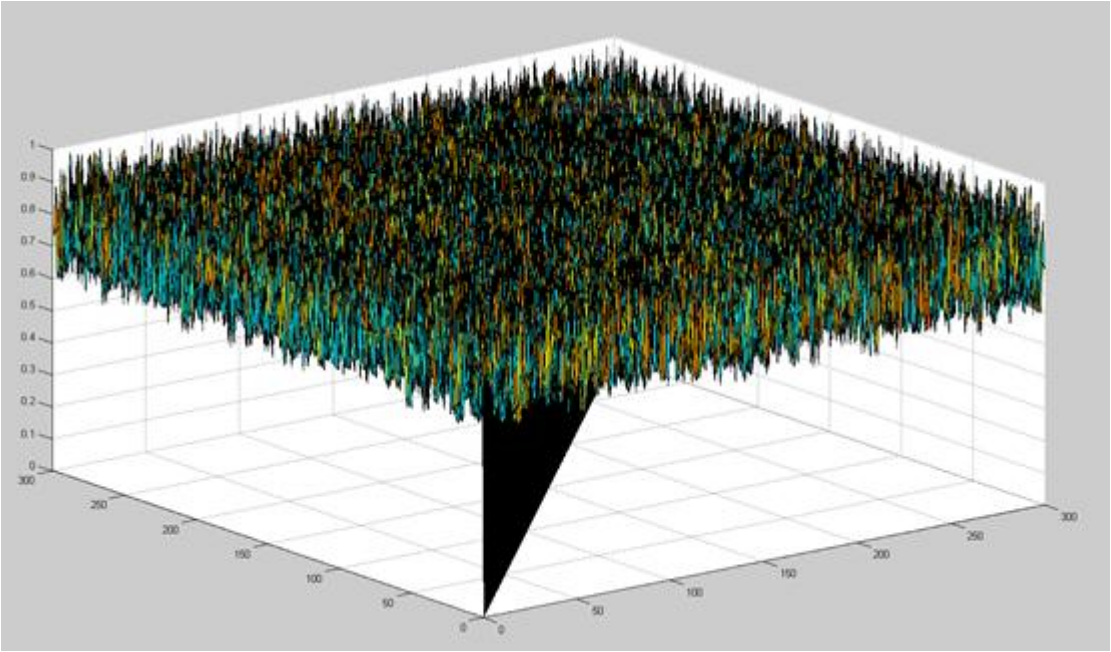
Στην EucledeanCF προσέγγιση έχουμε επίσης πολλαπλά κριτήρια. Η προσέγγιση μας παρέχει μια διαφορετική αντιμετώπιση όσον αφορά την εύρεση γειτόνων, δηλαδή χρηστών με τις ίδιες προτιμήσεις με τον

χρήστη για τον οποίο θέλουμε να κάνουμε την σύσταση μιας ταινίας μέσω του γενετικού μας αλγορίθμου. Στις παρακάτω δύο γραφικές παραστάσεις παρατηρούμε τα αποτελέσματα τις προσέγγισης αυτής.



Σχήμα 41

Και σε μορφή τρισδιάστατου πίνακα:



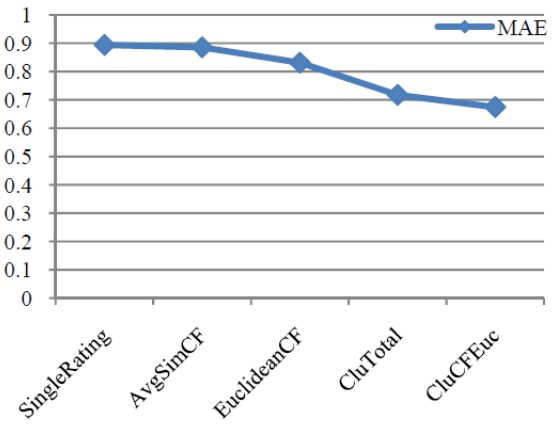
Σχήμα 42

Στις γραφικές αυτές παραστάσεις βλέπουμε μόνο θετικές τιμές πράγμα που σημαίνει ότι η προσέγγιση EucledeanCF βρίσκει χρήστες με ακριβώς ίδια γούστα απ' ότι οι προηγούμενες, έτσι μπορεί να παρέχει ακριβείς προβλέψεις κατά πολύ περισσότερο.

Και σε κώδικα matlab:

```
%%eucledia
NU=6078;
sum=zeros(NU);
riza=zeros(NU); % The factor under the root
%Similarities=zeros(NU); % Similarity index
for i=1:NU
for j=1:NU
if i==j
sum(i,j)=1;
else
for k=1:4
sum(i,j)=sum(i,j)+((W(i,k)-W(j,k))^2);
end
riza(i,j)=sqrt(sum(i,j));
Similarities3(i,j)=1/(1+riza(i,j));
end
end
end
```

Οι τελευταίες δυο προσεγγίσεις μας δίνουν καλύτερα αποτελέσματα από τα single-rated systems. Βλέπουμε καθαρά στο παραπάνω σχήμα ότι οι τιμές έχουν μικρότερη απόκλιση και είναι θετικές ως επί το πλείστον.
Στο παρακάτω σχήμα υπάρχει η κατανομή του Mean Absolute Error για τις προσεγγίσεις που παρουσιάσαμε.

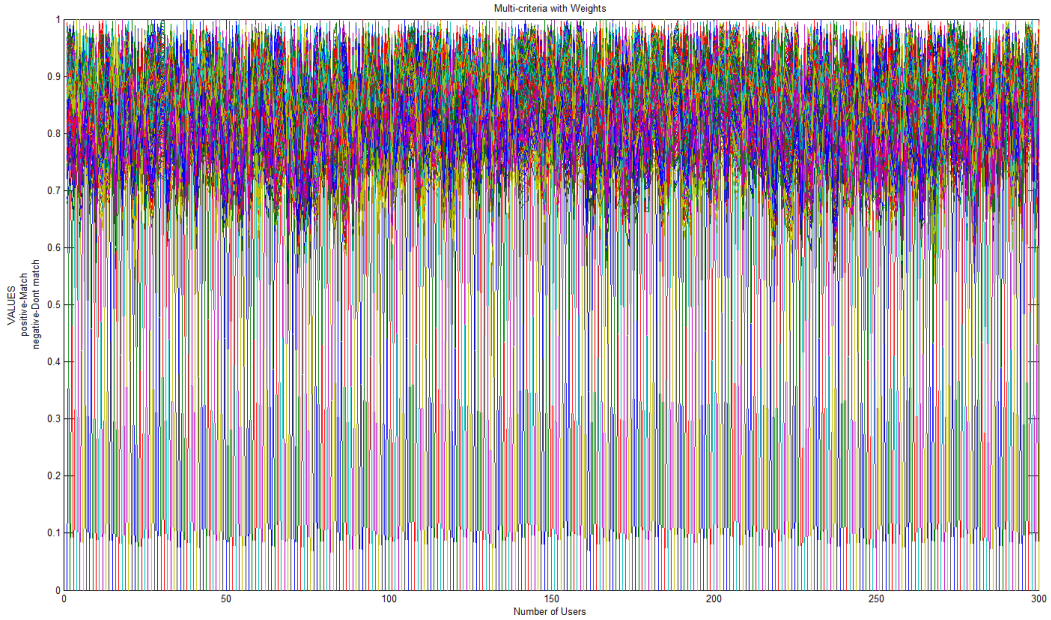


Σχήμα 43

6.4 Πολυκριτήρια με χρήση Βαρών

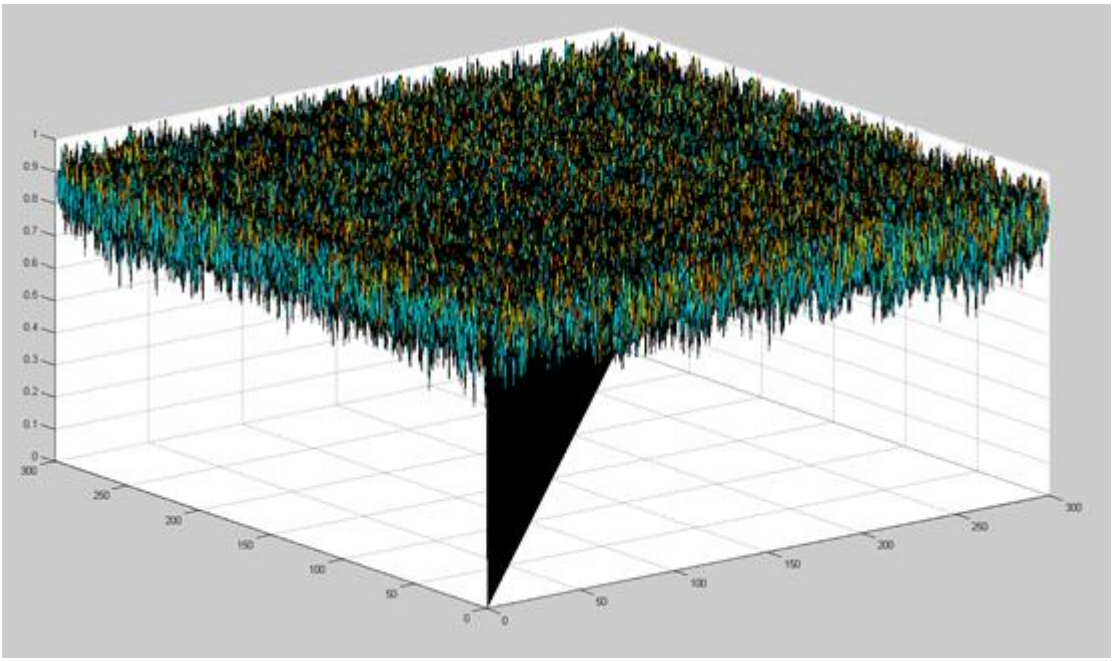
Ερχόμαστε τώρα στην περίπτωση στην οποία το κάθε κριτήριο κρίνεται διαφορετικά με βάση τον κάθε χρήστη.Ας θέσουμε ένα παράδειγμα. Για την σύσταση μιας ταινίας ένας συγκεκριμένος χρήστης μπορεί να θεωρεί ως σημαντικότερο την σκηνοθεσία από ότι τα ειδικά εφέ. Ένας άλλος χρήστης επίσης μπορεί να θεωρεί ως σημαντικότερο κριτήριο για να δει μια ταινία την ηθοποιία από την σκηνοθεσία. Αυτό μας δίνει να καταλάβουμε ότι δεν γίνεται όλοι οι χρήστες να αξιλογηθούν το ίδιο. Πρέπει για τον καθένα να γίνονται συστάσεις με βάση την σπουδαιότητα των κριτηρίων στα οποία έχουν δώσει προτεραιότητα. Έτσι στην μέθοδο αυτή εξάγουμε βάρη για κάθε κριτήριο, δηλαδή έναν αριθμό για κάθε κριτήριο που ορίζει την σπουδαιότητα του με βάση τις προκαθορισμένες αξιολογήσεις των χρηστών. Στις επόμενες γραφικές παραστάσεις

παρατηρούμε τα αποτελέσματα τις μεθόδου αυτής.



Σχήμα 44

και σε μορφή τρισδιάστατου πίνακα



Σχήμα 45

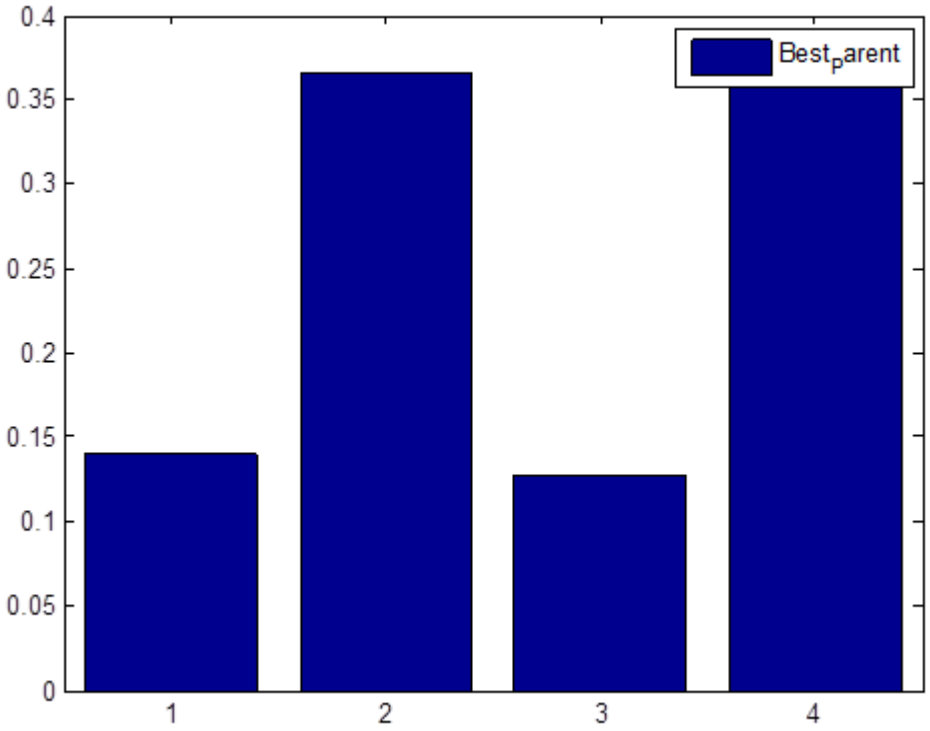
Είναι εύκολο να παρατηρήσει κανείς εδώ ότι οι γραφικές παραστάσεις μας επαληθεύουν την προηγούμενή μας υπόθεση. Η ομοιότητα των χρηστών μεταξύ τους είναι πάρα πολύ μεγάλη. Αυτό προκύπτει διότι έχουμε τιμές μεταξύ του +0,7 και 1 το οποίο 1 μας δείχνει ότι δυο χρήστες έχουν ακριβώς τα ίδια ενδιαφέροντα. Αυτό είναι λογικό να προκύψει διότι εξετάζονται οι χρήστες που έχουν όμοια ενδιαφέροντα μεταξύ τους με την χρήση βαρών για κάθε κριτήριο.

Και σε κώδικα matlab:

```
NU=6078;
sum=zeros(NU);
riza=zeros(NU); % The factor under the root
%Similarities=zeros(NU); % Similarity index
for i=1:NU
for j=1:NU
if i==j
sum(i,j)=1;
else
for k=1:4
sum(i,j)=sum(i,j)+WEIGHTS(1,k)*((W(i,k)-W(j,k))^2);
end
riza(i,j)=sqrt(sum(i,j));
Similarities5(i,j)=1-riza(i,j);
end
end
end
```

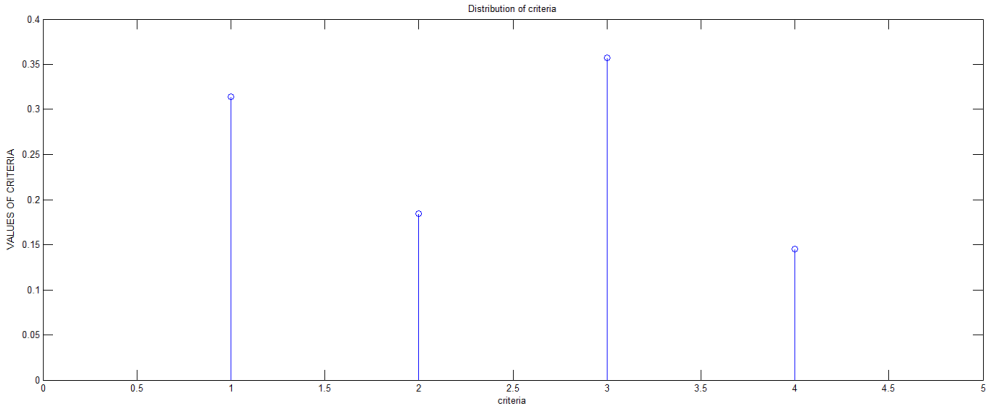
Συμπερασματικά βλέπουμε ότι συκρίνοντας τα αποτελέσματα εφαρμογής των μεθόδων μεταξύ τους βλέπουμε ότι τα καλύτερα αποτελέσματα τα παίρνουμε με την διαδικασία τις πολυκριτηριακής ανάλυσης με βάρη. Οι τιμές κατανέμονται καλύτερα και έχουμε αρκετά περισσότερες θετικές τιμές. Στις δύο παρακάτω γραφικές παραστάσεις παρατηρούμε την κατανομή των βαρών για την μέθοδο αυτή.

Τα 4 καλύτερα κριτήρια για τον γονέα είναι τα εξής:

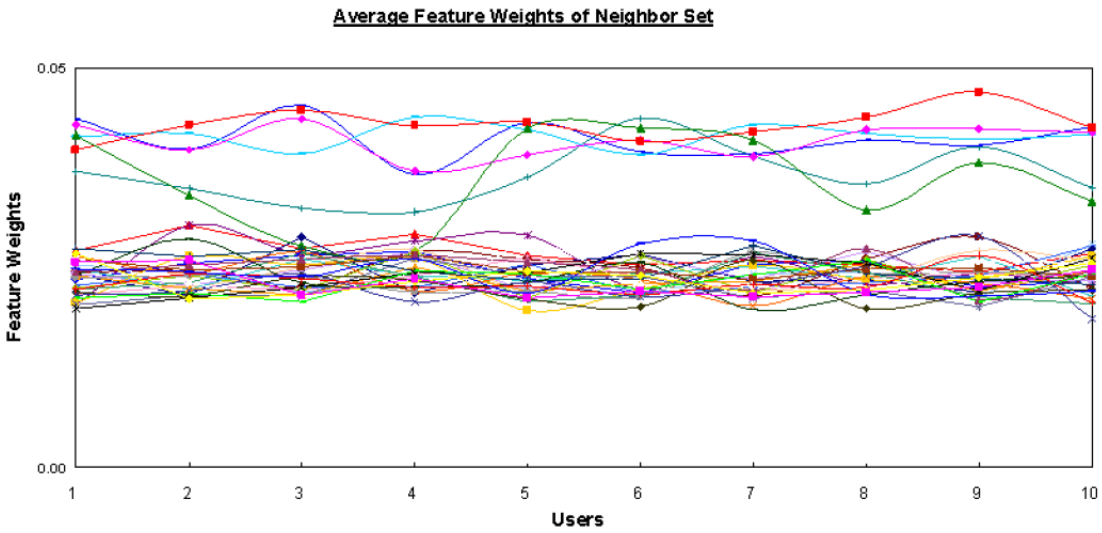


Σχήμα 46

Ενώ τα τέσσερα καλύτερα κανονικοποιημένα βάρη των κριτηρίων μετά το τέλος του γενετικού μας αλγορίθμου είναι:



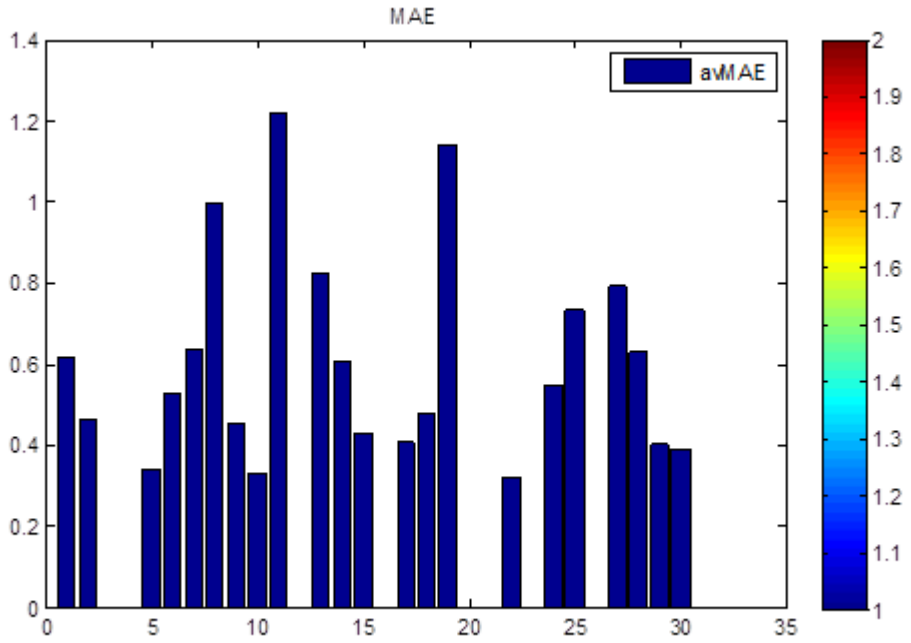
Σχήμα 47



Σχήμα 48

6.5 Υπολογισμός και Απεικονίσεις MAE

Στην παρακάτω γραφική παράσταση παρατηρούμε την κατανομή του Mean Absolute Error.



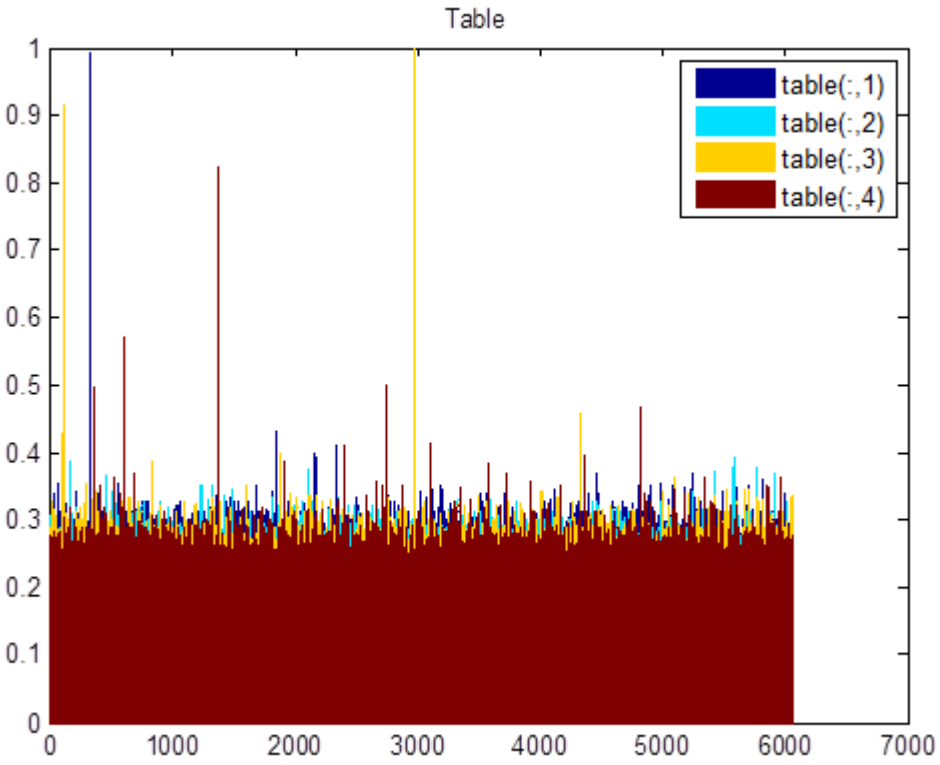
Σχήμα 49

[131]

6.6 Γραφικές Απεικονίσεις Προβλήματος

Στην υποενότητα αυτή του κεφαλαίου μας παρουσιάζουμε γενικά διάφορες γραφικές παραστάσεις οι οποίες μας δείχνουν την πορεία της εργασίας μας και τα αποτελέσματα της δουλειάς μας.

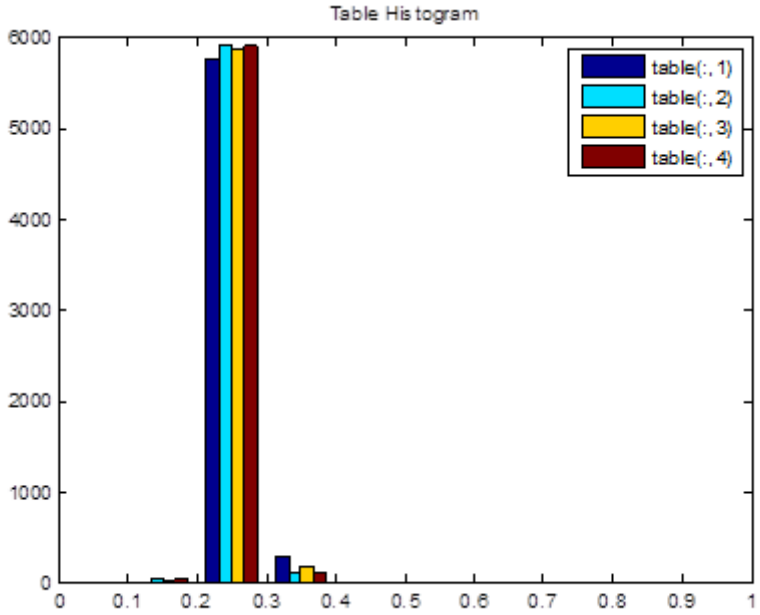
Συγκεκριμένα έχουμε τον πίνακα εισόδου αξιολόγησης με τα δεδομένα που μας δόθηκαν τα οποία κυμαίνονται από 0 μέχρι 1 με ακρίβεια τεσσάρων δεκαδικών ψηφίων είναι της μορφής:



Σχήμα 50
Ο αρχικός πίνακας εισόδου δεδεμένων

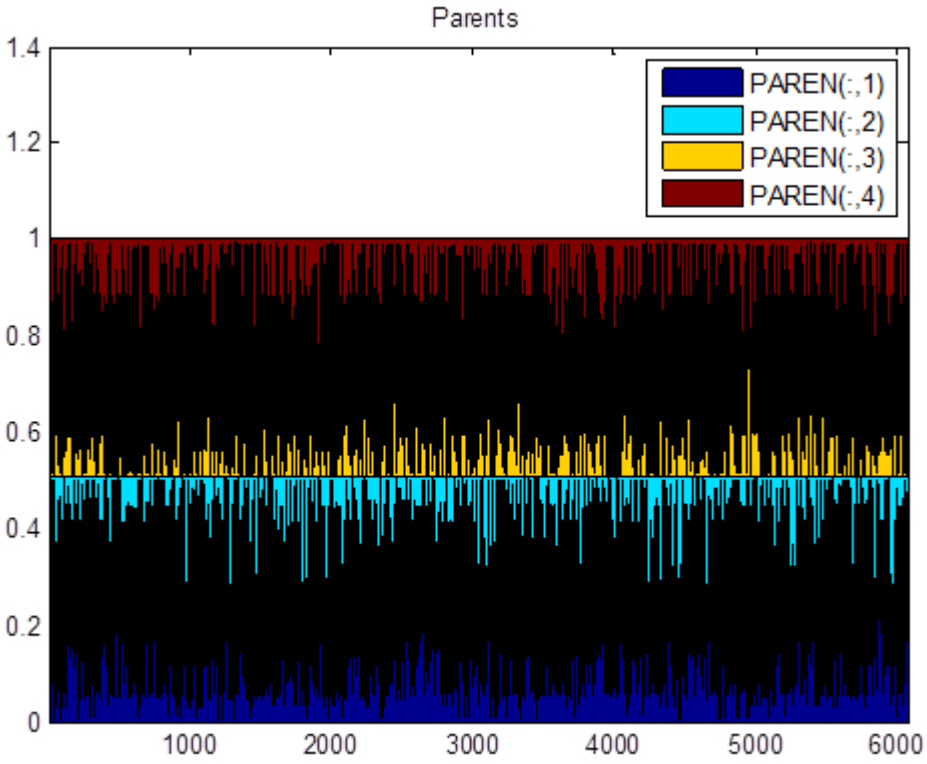
Όπου με διαφορετικό χρώμα είναι οι τιμές του κάθε χρήστη για καθένα από τα τέσσερα κριτήρια.

Και σε μορφή ιστογράμματος:



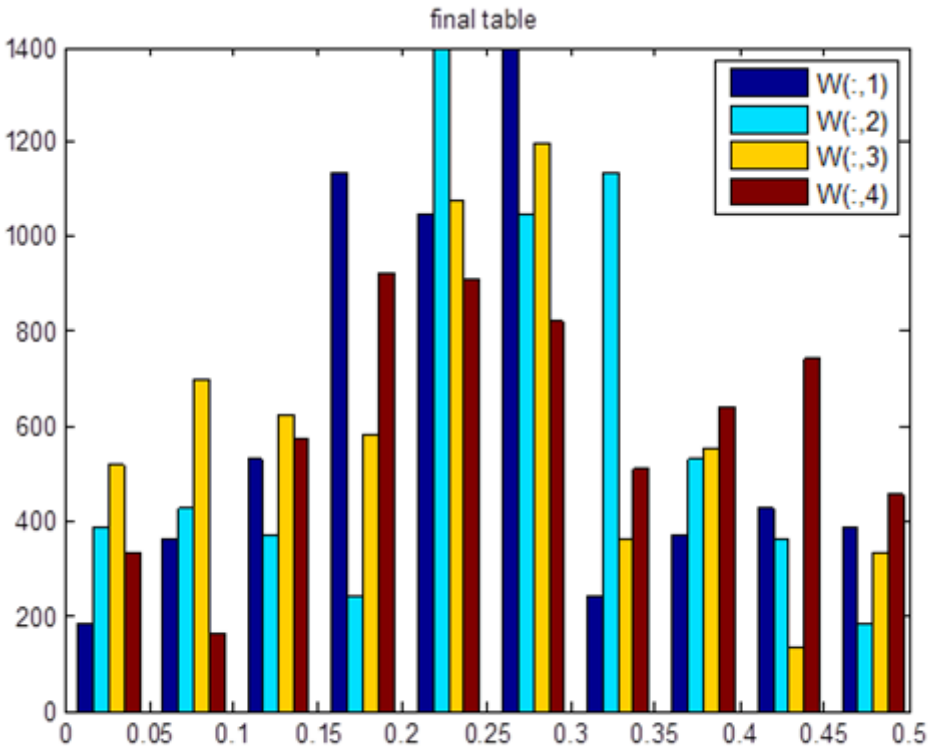
Σχήμα 51

Με τη δημιουργία του γενετικού μας αλγορίθμου δημιουργούμε απογόνους οι οποίοι με τη σειρά τους προσαρμόζονται στα αρχικά μας δεδομένα ανάλογα με τον αριθμό των γενεών του αλγορίθμου μας. Ο πίνακας των γονέων μετά το πέρας των επαναλήψεων είναι :



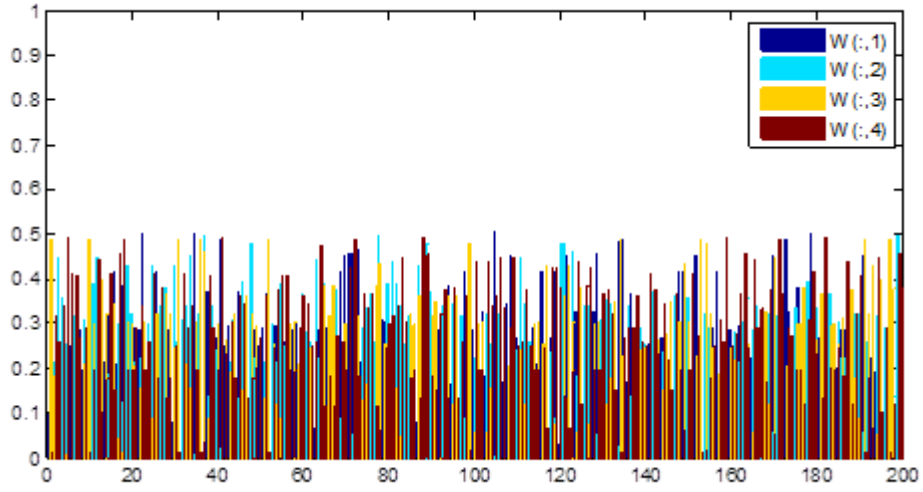
Σχήμα 52

Ο τελικός πίνακας ο οποίος μας δίνει τα μοντελοποιημένα μας αποτελέσματα είναι της μορφής :



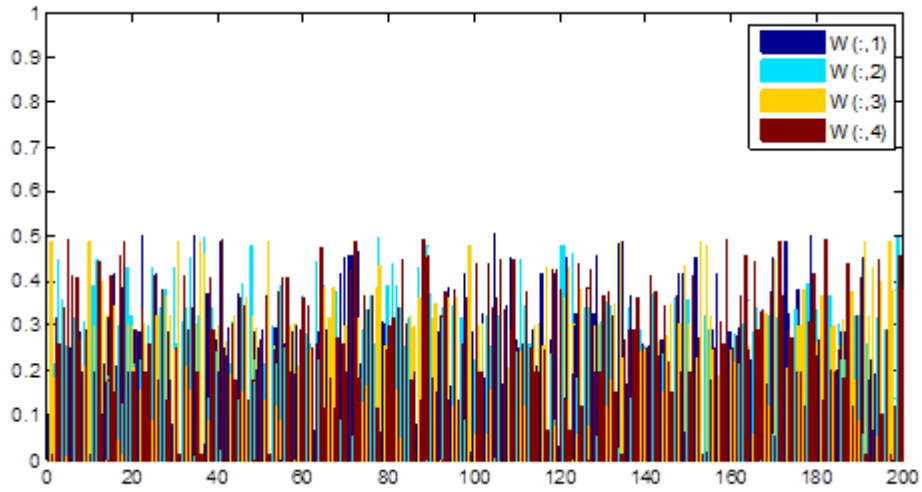
Σχήμα 53

ενώ κάνοντας zoom για μια συγκεκριμένη περιοχή για τους χρήστες με αριθμό από 1 έως 200 έχουμε



Σχήμα 54

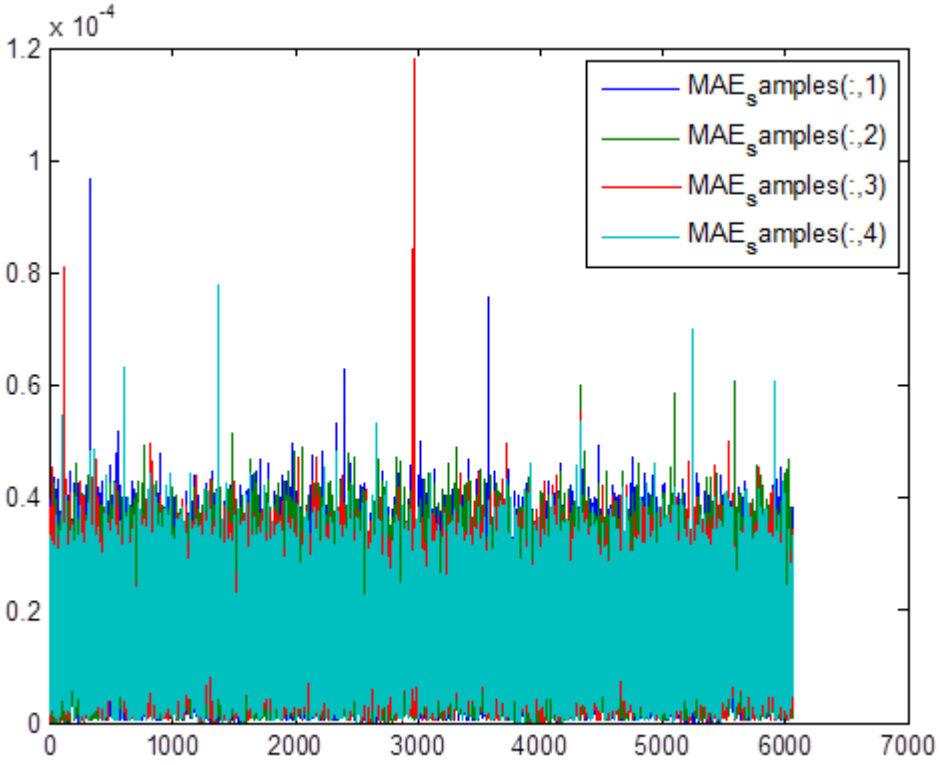
final table



Σχήμα 55

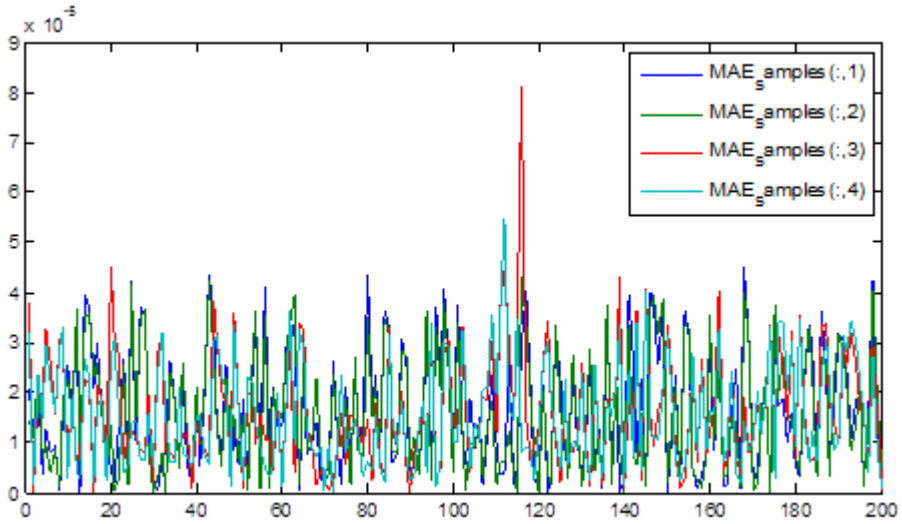
Υπολογίζουμε επίσης το συνολικό MAE το οποίο μας δίνει την τιμή MAE= 0,4044 κανονικοποιημένο.

Το MAE για κάθε στοιχείο του πίνακά μας παρουσιάζει την εξής γραφική παράσταση



Σχήμα 56

Και ειδικότερα στην περιοχή μεταξύ 200 χρηστών σε ένα τυχαίο δείγμα βλέπουμε ότι κατανέμονται ισότιμα

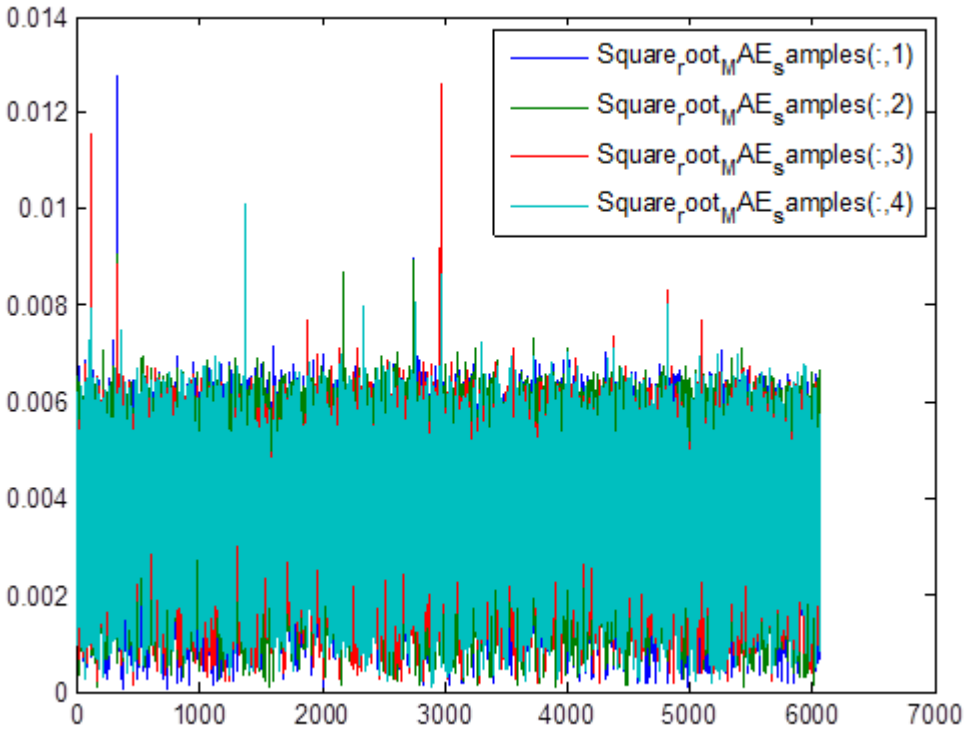


Σχήμα 57

Στην συνέχεια παραθέτουμε το τετραγωνικό μέσο απόλυτο σφάλμα όπως και στις παραπάνω γραφικές

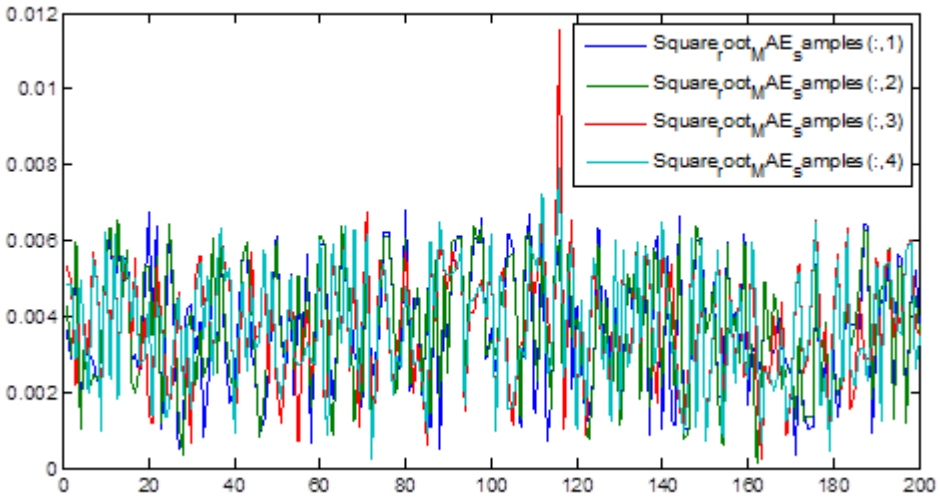
Υπολογίζουμε το συνολικό τετραγωνικό MAE το οποίο μας δίνει την τιμή square root MAE= 0,6392 κανονικοποιημένο.

Το square root MAE για κάθε στοιχείο του πίνακά μας παρουσιάζει την εξής γραφική παράσταση



Σχήμα 58

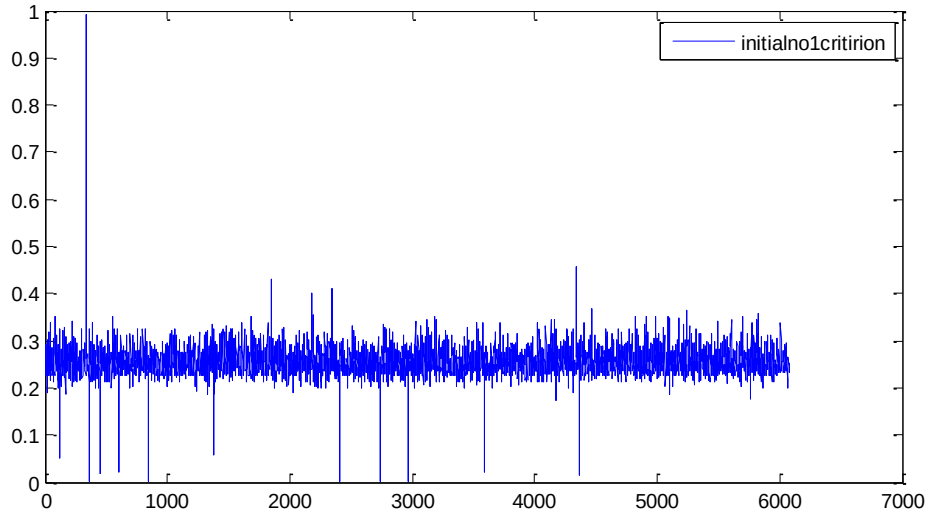
Και ειδικότερα στην περιοχή μεταξύ 200 χρηστών έχουμε



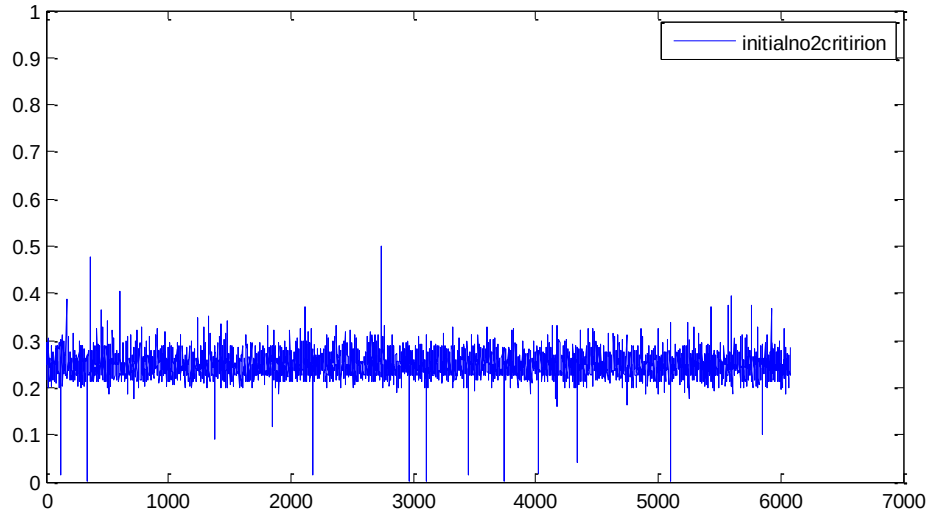
Σχήμα 59

Στη συνέχεια στα πλαίσια της διπλωματικής αυτής εργασίας θα παρουσιάσουμε τις κατανομές των βαρών ανά κριτήριο

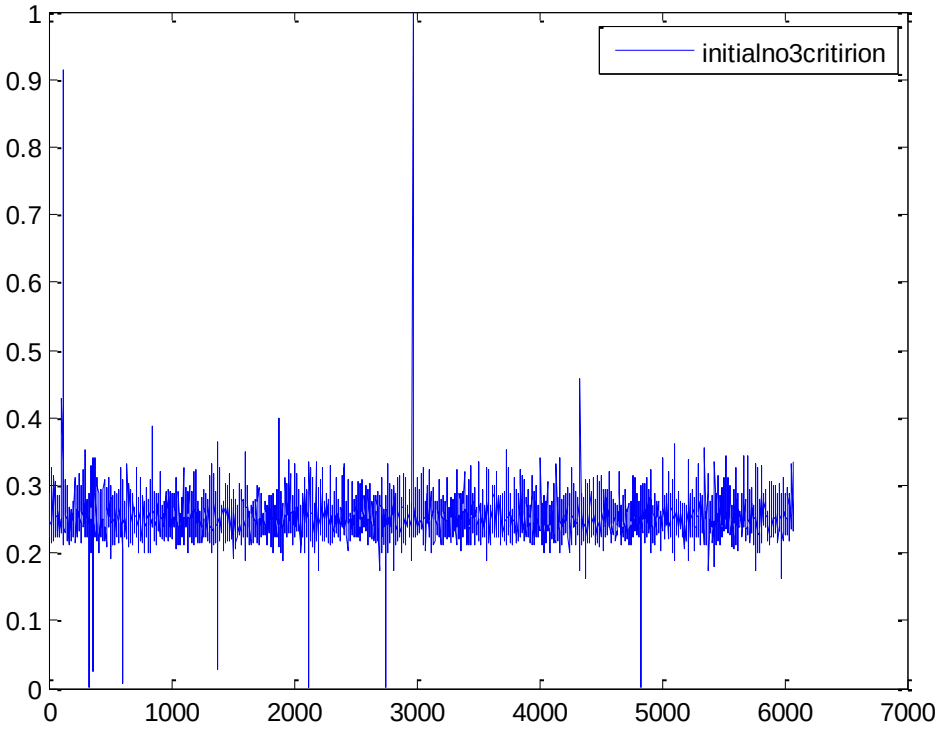
Τα αρχικά μας δεδομένα για τα 4 κριτήρια



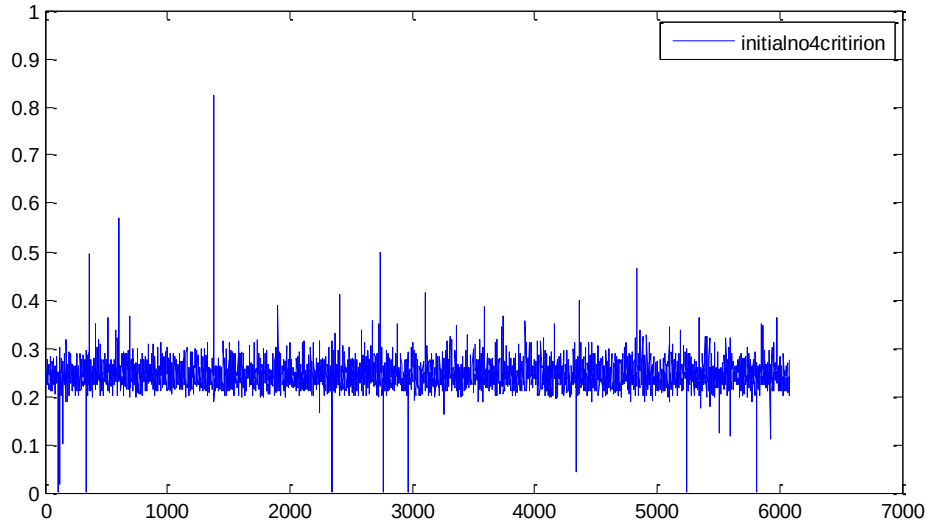
Σχήμα 60



Σχήμα 61

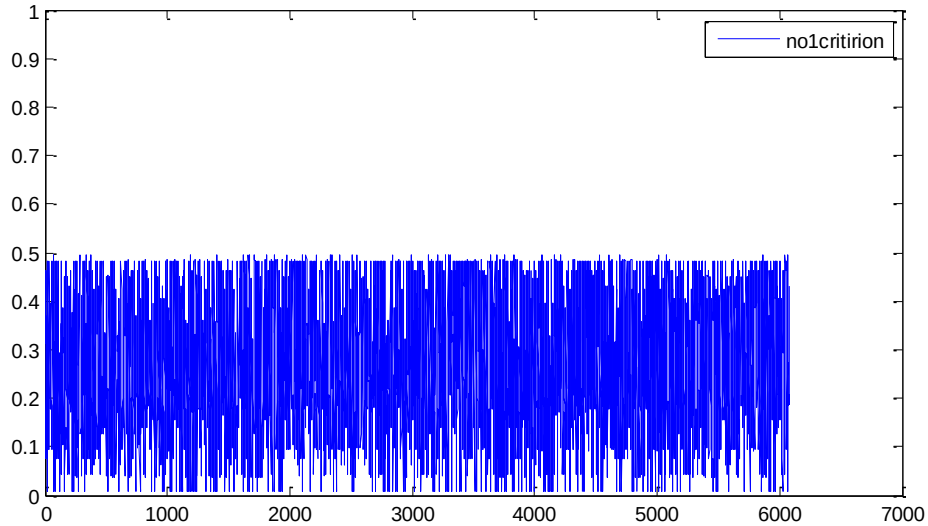


Σχήμα 62

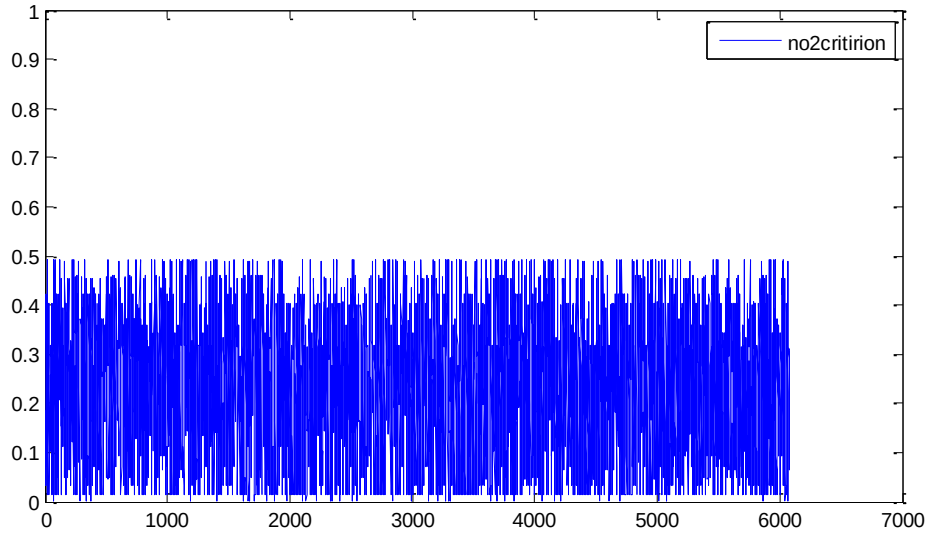


Σχήμα 63

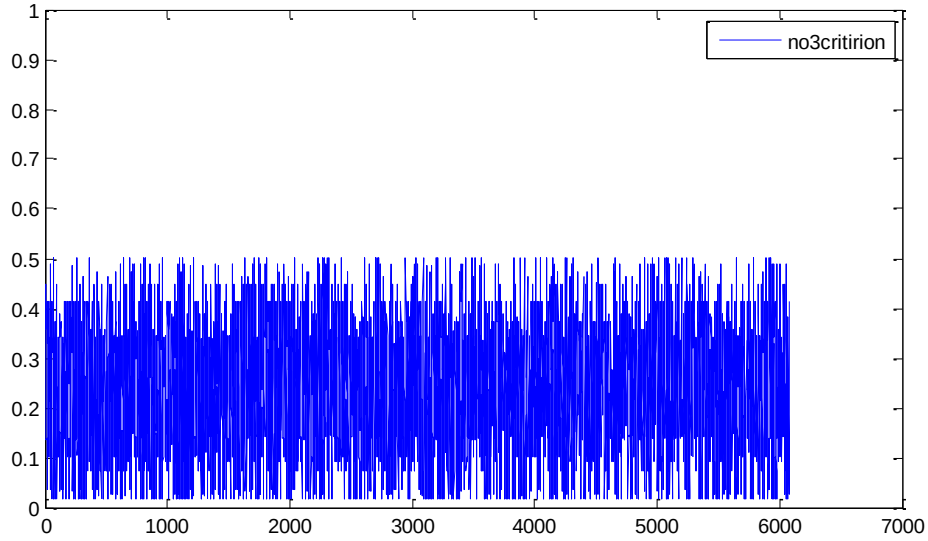
Τα κανονικοποιημένα κριτήρια



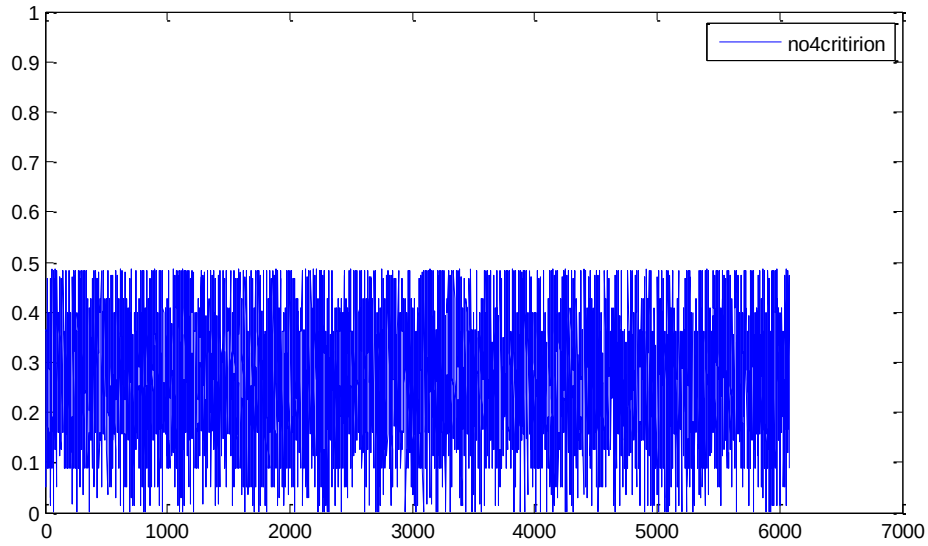
Σχήμα 64



Σχήμα 65



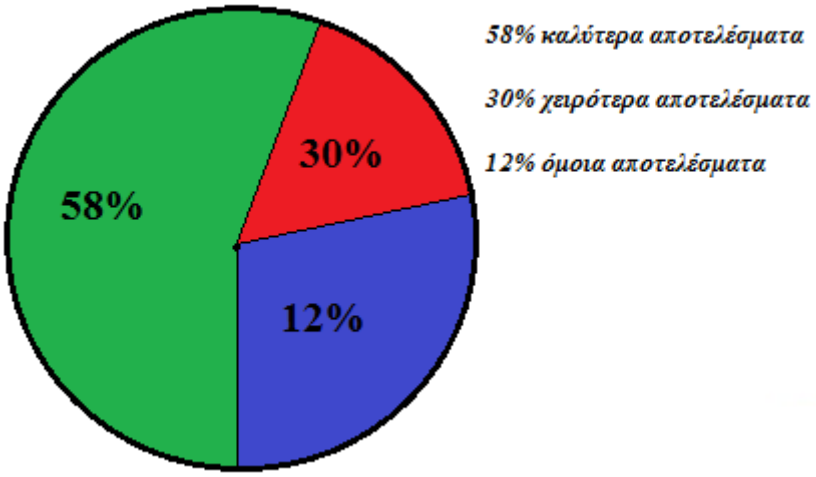
Σχήμα 66



Σχήμα 67

6.7 Αξιολόγηση Συστημάτων Συστάσεων

Με την ανάκτηση των μέσων απολύτων σφαλμάτων είναι δυνατή η σύγκριση με άλλα μοντέλα συστάσεων που χρησιμοποιούν διαφορετικές μεθοδολογίες.
Μια μεθοδολογία με την οποία τα αποτελέσματα της παρούσας διπλωματικής συγκρίνονται είναι με τα αποτελέσματα που παρουσιάστηκαν από K.Lakiotaki, N. F. Matsatsinis, and A.Tsoukiàs, "Multi-Criteria User Modeling in Recommender Systems", *IEEE Intelligent Systems.*.
Παρακάτω παραθέτουμε τις γραφικές παραστάσεις για σύγκριση των αποτελεσμάτων.

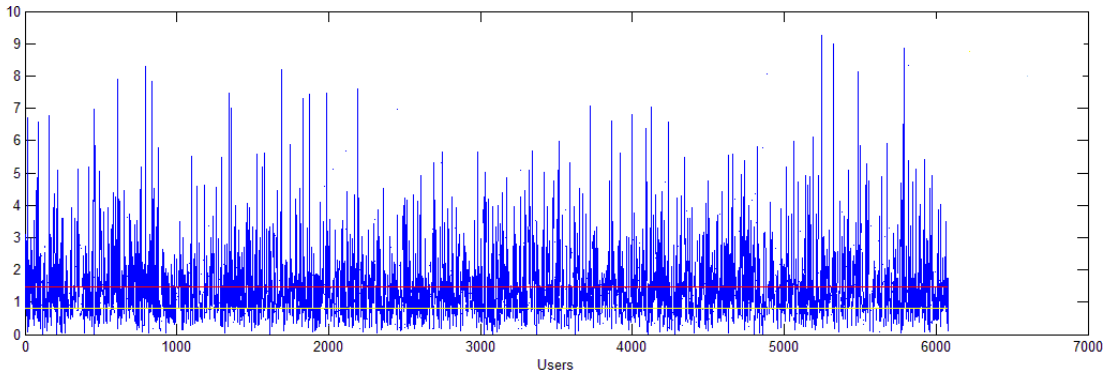


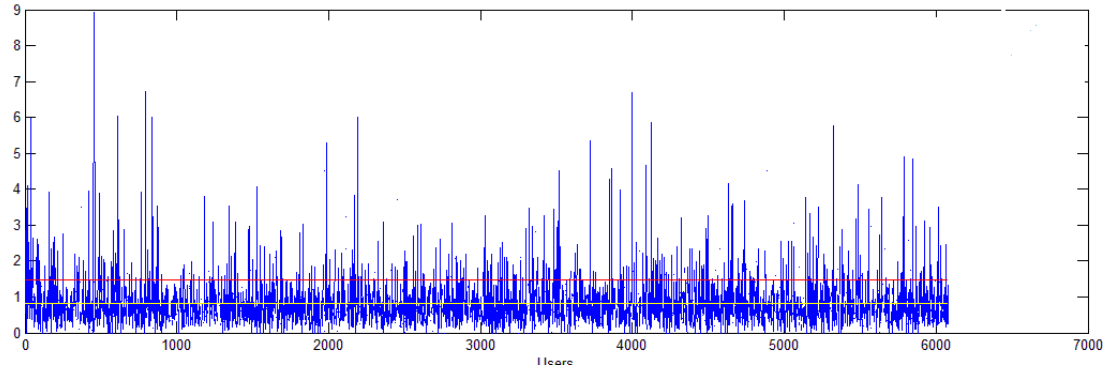
Σχήμα 68

Διαφοροποιήσεις Γενετικού Αλγορίθμου

Στο διάγραμμα αυτό φαίνεται πως το μοντέλο σύστασης ταινιών της εργασίας μας που χρησιμοποιεί γενετικό αλγόριθμο βελτιώνει εν γένει τα αποτελέσματα για τους περισσότερους χρήστες.Προσεγγιστικά βελτιώνει περιπού 3500 αποτελέσματα χρηστών. Δηλαδή ενώ το μοντέλο MRCF εμπεριείχε σφάλμα στην σωστή πρόβλεψη των ταινιών προς σύσταση στους χρήστες το μοντέλο της εργασίας μειώνει το σφάλμα αυτό. Υπήρξαν περίπου 600 χρήστες για τους οποίους το σφάλμα φαίνεται να αυξάνεται ενώ για περίπου 30% των χρηστών το σφάλμα παραμένει σχεδόν το ίδιο. Εδώ βλέπουμε ότι ουσιαστικά βελτιώνει το σφάλμα για την σωστή πρόβλεψη του αντικειμένου προς τους χρήστες σε σχέση με το προηγούμενο μοντέλο.

Συγκρίνοντας τα απόλυτα σφάλματα ξανά διαπιστώνουμε το εύρος που προκύπτει είναι αρκετά μικρότερα και σε τιμές μεγαλύτερες στην περίπτωσή μας.

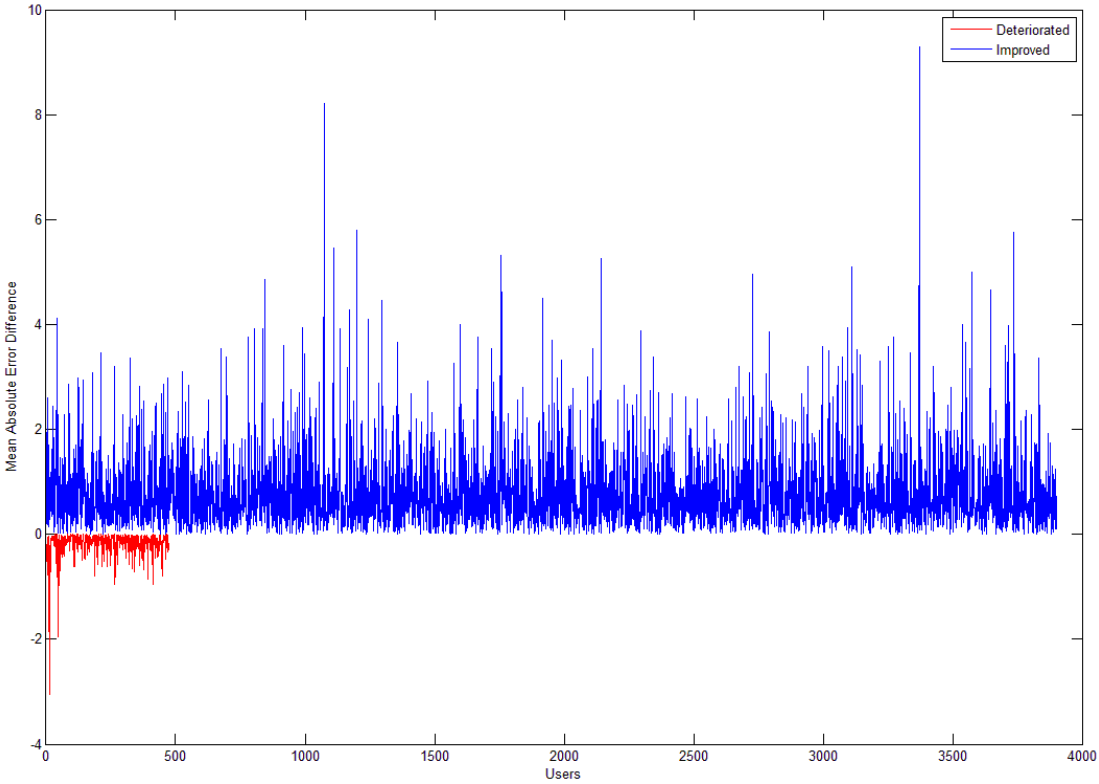




Σχήμα 69

Ουσιαστικά στην υποενότητα αυτή του κεφαλαίου μας εξετάζουμε τις διαφορές μεταξύ της δημιουργίας ενός recommendation system με βάση έναν γενετικό αλγόριθμο και της ήδη υπάρχουσας μεθοδολογίας ενός recommendation system με clustering.

Στο Σχήμα 69 φαίνονται τα μέσα απόλυτα σφάλματα της δικής μας μεθόδου και της μεθόδου με clustering. Έτσι δίνεται μια πρώτη εικόνα για το μέγεθος των σφαλμάτων από την οποία φαίνεται ότι τα σφάλματα της μεθόδου συστάσεων MRCF έχουν μεγαλύτερο εύρος, είναι δηλαδή μεγαλύτερα.



Σχήμα 70

Στο παραπάνω σχήμα απεικονίζεται το μέγεθος των βελτιώσεων που παρατηρήθηκαν πιο συγκεκριμένα μέσω των MAE.

Τα αποτελέσματα μας δείχνουν ότι με χρήση της πολυκριτηριακής ανάλυσης τα αποτελέσματα είναι καλύτερα διότι το MAE είναι μικρότερο.

Methods	Mean MAE	Best Difference Value
MRCF	1.4825	9.298
GAs+CF	0.8118	3.053

Σχήμα 71

ΚΕΦΑΛΑΙΟ 7

ΣΥΜΠΕΡΑΣΜΑΤΑ-ΜΕΛΛΟΝΤΙΚΕΣ ΕΡΓΑΣΙΕΣ

Η παρούσα διπλωματική εργασία παρουσιάζει την ανάπτυξη ενός recommendation system βασιζόμενου στη μοντελοποίηση του προφίλ των χρηστών με στόχο την προσωποποιημένη αναζήτηση ανάλογα με τα ειδικά χαρακτηριστικά του χρήστη και του προς αναζήτηση αντικειμένου (π.χ μουσική, video, βιβλίο, άρθρο κλπ). Η μοντελοποίηση βασίζεται στην πολυκριτηριακή θεωρία αποφάσεων καθώς και στη χρήση μεθόδων και τεχνικών από το χώρο της τεχνητής νοημοσύνης όπως οι ευρετικοί και μεθευρετικοί αλγόριθμοι. Η μεθοδολογία προβλέπει την δυναμική μεταβολή των χαρακτηριστικών αναλόγως με τη συμπεριφορά των χρηστών (μεταβολή των χαρακτηριστικών ανάλογα με τα χαρακτηριστικά, τη συχνότητα και τα αποτελέσματα των αναζητήσεων). Τα αποτελέσματα των αναζητήσεων κατατάσσονται ανάλογα με τη προσδοκώμενη χρησιμότητα της ανακτωμένης πληροφορίας.

Η ανάγκη για προσωποποιημένη αναζήτηση αποτελεί κυρίαρχο ζητούμενο αφού καθημερινά όλο και περισσότερες πληροφορίες γίνονται διαθέσιμες μέσω ηλεκτρονικών μέσων.

Επίσης υπάρχει μια τρομακτική αύξηση των διαθέσιμων πληροφοριών, των πηγών και της πληροφορίας πράγμα που οδηγεί ταυτόχρονα στην δυσκολία με την οποία αυτές οι πληροφορίες μπορούν να αποκτηθούν από τον χρήστη.

Αρχικά, παρουσιάστηκαν οι υφιστάμενη κατάσταση για τα recommendation systems και την προσωποποιημένη αναζήτηση με βάση του προφίλ των χρηστών. Στην συνέχεια αναπτύχθηκαν οι βασικές αρχές των συστημάτων που χρησιμοποιούν διαδικασίες Information

Filtering για την αναζήτηση πληροφοριών. Ακολούθησε μια γενική αναφορά στις τεχνικές μοντελοποίησης, στις μεθοδολογίες απόκτησης και ανανέωσης του μοντέλου του χρήστη καθώς και στους τρόπους μάθησης των συστημάτων μοντελοποίησης. Στο επόμενο κεφάλαιο, παρουσιάστηκαν συστήματα μοντελοποίησης χρήστη δίνοντας έτσι μια σαφή εικόνα για τις τάσεις που επικρατούν στον σχεδιασμό και την υλοποίηση τέτοιων συστημάτων. Στο 4^{ο} κεφάλαιο παρουσιάστηκε το θεωρητικό υπόβαθρο για την πολυκριτηριακή ανάλυση καθώς και τον τρόπο λειτουργίας και υλοποίησης των ευρετικών και μεθευρετικών αλγορίθμων και την χρήση τους στην παρούσα εργασία. Στο 5^{ο} κεφάλαιο έγινε η ανάλυση και η σχεδίαση του συστήματός μας.

Στην παρούσα εργασία προτάθηκε ένα υβριδικό σύστημα συστάσεων που βασίζεται σε γενετικούς αλγορίθμους και στην τεχνική του συστήματος συνεργασίας. Στο σύστημα αυτό ενσωματώνονται δεδομένα για την προτίμηση σε τέσσερα κριτήρια για κάθε αντικείμενο του εκάστοτε χρήστη για την κατασκευή του προφίλ προτίμησης των χρηστών του συστήματος. Ο γενετικός αλγόριθμος εφαρμόζεται για την βελτιστοποίηση του διανύσματος βαρών των χαρακτηριστικών των αντικειμένων τα οποία χρησιμοποιούνται στον υπολογισμό της ομοιότητας μεταξύ των χρηστών. Στην έρευνα αυτή επιβεβαιώθηκε ότι το σύστημα συνεργασίας έχει υψηλή απόδοση όταν ενσωματώνονται σε αυτό πληροφορίες για τα βάρη στάθμισης των χαρακτηριστικών των αντικειμένων.

Η προτεινόμενη μεθοδολογία ανάπτυξης του συστήματος έχει την δυνατότητα εξέλιξης και ενσωμάτωσης νέων αντιλήψεων σε όλα τα επιμέρους τμήματά της. Μια από τις πιθανές προεκτάσεις της συγκεκριμένης μεθοδολογίας αφορά την υιοθέτηση της ELECTRE TRI (Yu,1992) για την κατάταξη των ανακτώμενων άρθρων.

Παράλληλα, θα μπορούσε να υιοθετηθεί η μεθοδολογία δημιουργίας κοινοτήτων χρηστών - communities (Paliouras et al, 2002) βασισμένων στο είδος των αναζητήσεων που ο κάθε χρήστης πραγματοποιεί.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. A.D. Lovie P. Lovie, The Flat Maximum Effect and Linear Scoring Models for Prediction

2. A.Felfernig, G. Friedrich, and L. Schmidt-Thieme, "Guest Editors' Introduction:Recommender Systems," IEEE Intelligent Systems, 2007

3. A.G.Moukas User Modeling in a MultiAgent Evolving System

4. A.Kobsa (2000), Generic User Modeling Systems, User Modeling and User-Adapted Interaction ch 11: pp 49-63

5. A.Pretschner, S.Gauch,(1999), Personalization on the Web, Technical Report ITTC-FY2000-TR-13591-01, Information and Telecommunication Technology Center Department of Electrical and Computer Science, University of Kansas

6. Adomavicius, and A. Tuzhilin, (2005), Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions.

7. Alfred Kobsa (1993) User Modeling - Recent Work, Prospects and Hazards

8. Architecture of the World Wide Web, Volume One. W3C Recommendation, http://www.w3.org/TR/webarch/, 15 December 2004. pp. 5

9. Badrul S, George K, Joseph K, & John R (2001), Item-Based Collaborative filtering recommendation algorithms

10. Balabanovic, M., (1998), Exploring versus Exploiting when Learning User Models for Text Recommendation. User Modeling and User-Adapted Interaction 8(1-2), 71-102.

11. Beuthe & Scannella (2001), Multiple Critiria Decision Analysis: state of the art surveys

12. Billsus D,. and Pazzani M. ,(1998), Learning collaborative information filters. In Proceedings of the Fifteenth International Conference on Machine Learning, pp 43-52.

13. Billsus D, Pazzani (1999), A hybrid user model for news classification

14. Billsus D, Pazzani, Chen(2000), A learning agent for wireless news access, Department of Information and Computer Science University of California, Irvine

15. Breese J S, Heckerman D & Kadie C (1998), Empirical Analysis of Predictive Algorithms for Collaborative Filtering

16. Burke R (2002), Hybrid Recommender Systems: Survey and Experiments

17. C.Basu, H.Hirsh, W.Cohen (1998), Recommendation as Classification: Using Social and Content-Based Information in Recommendation, AAAI-98 Proceedings, AAAI (www.aaai.org)

18. Cho, Y.H., J. K. Kim, S.H. Kim (2002), A personalized recommender system based on web usage mining and decision tree induction, Expert Systems with Applications, vol. 23,no. 3, pp. 329–342.

19. D. W. Connolly and L. Masinter. *RFC 2854:* The 'text/html' Media Type. (2000) pp 6

20. D.Kelly, N.J. Belkin, Modeling Characteristics of the User's Problematic Situation with Information Search & Use Behaviors, School of Communication, Information & Library Studies, Rutgers University

21. Despotis(1990), A multi-criteria analysis of stated preferences among freight transport alternatives (2003)

22. Edwards (1977), Von Winterfeldt & Edwards (1986) , Edwards & Barron (1994), MultiAttribute Rating Technique (SMART)

23. Francesco Ricci, Lior Rokack, Bracha Shapira,Introduction to Recommender Systems Handbook

24. G. Linden, B. Smith, and J. York. Amazon.com,(2003), Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing, http://dx.doi.org/10.1109/MIC.2003.1167344

25. G.Adomavicius, A.Tuzhilin (2005) Toward the next generation of recommender systems - a survey of the state-of-the-art and possible extensions, IEEE Transactions on Knowledge and Data Engineering, vol 17, no 6.

26. G.Haubl, K.B Murray, V.Trifts (2002), Personalized Product Presentation: The Influence of ElectronicRecommendation Agents on Consumer Choice

27. Glance, N., Arregui, D. and Dardenne, M. (1999), Making Recommender Systems Work for Organizations

28. Goldberg, D., Nichols, D., Oki, B.M., Terry, D. Using collaborative filtering to weave an information Tapestry

29. Hanani, U., Shapira B., Shoval P., (2000), Information Filtering: Overview of Issues, Research and Systems, Israel.

30. Hwang & Yoon (1981), The Technique for Order Preference by Similarity to Ideal Solution

31. Jameson, A., (1995). Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues. User Modeling and User-Adapted Interaction 5(3-4), pp 193-251.

32. Jameson, A., GroÞmann-Hutter, B.March, L. and Rummer, R. (2000), Creating an Empirical Basis for Adaptation Decisions. In: IUI2000 Proceedings of the International Conference on Intelligent User Interfaces. New Orleans, Louisiana, pp.149-156

33. K.Lakiotaki, N. F. Matsatsinis, and A.Tsoukiàs, "Multi-Criteria User Modeling in Recommender Systems", *IEEE Intelligent Systems.* (accepted)

34. K. Lakiotaki, P. Delias, V. Sakkalis and N. F. Matsatsinis, "User Profiling based on Multi-criteria Analysis: The role of Utility Functions", *Operational Research: An International Journal*, 9(1),3-16, (2009)

35. K. Lakiotaki, S. Tsafarakis, and N. Matsatsinis. Uta-rec: a recommender system based on multiple criteria analysis. In *Proc. of the 2008 ACM conference on Recommender systems*, pages 219–226. ACM New York, NY, USA, 2008.

36. K.Mock, V.R.Vemuri, Adaptive User Models For Intelligent Information Filtering

37. Kass, R., and Finin,T., (1989), The role of user modeling in cooperative interactive systems, International Journal of Intelligent Systems

38. Kay J., (2001), Learner Control, User Modeling & User Adapted Interaction, vol 11:p. 111-127

39. Keeney,Raiffa (1976), Decisions with multiple objectives: Preferences and value tradeoffs

40. Kobsa, A., (1993),User Modeling: Recent Work, Prospects and Hazards.

41. Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J. (1997), Applying collaborative filtering to usenet news. Communications of the ACM

42. L.Liu,N.Mehandjiev,D.Xu (2011), Multi-Criteria Service Recommendation Based on User Criteria Preferences, in: Proceedings of the fifth ACM conference on Recommender systems (RecSys '11).

43. M. Deshpande and G. Karypis. (2004), Item-based top-N recommendation algorithms. ACM Trans. Inf. Syst., http://doi.acm.org/10.1145/963770.963776

44. M.Dastani, N.Jacobs, C.M. Jonker, J.Treur, Modeling User Preferences and Mediating Agents in Electronic Commerce, Department of Artificial Intelligence, Vrije University Amsterdam

45. Malone, Grant, Turbak, Brost, Cohen (1987), Intelligent information-sharing systems

46. Middleton S.E ,Alani H, Shadbolt N & De Roure D.C (2002), Exploiting Synergy Between Ontologies and Recommender Systems. The Semantic Web Workshop, World Wide Web

47. Montaner, Lopez, Rosa (2003), A taxonomy of Recommender Agents on the Internet

48. Morita, M. and Shinoda, Y., (1994). Information filtering based on user behavior analysis and best match retrieval, Proceedings of the 17th Annual Intl. ACM SIGIR

49. Moukas A., Maes P., (1998), Amalthea: An Evolving Multi – Agent Information Filtering and Discovery System for the WWW, Autonomous Agents and Multi – Agents Systems, vol. l, Kluwer Academic Publishers, pp 59-88

50. Nasraoui, O., Krishnapuram, R. (2000), Extracting Web User Profiles Using Relational Competitive Fuzzy Clustering. In International Journal on Artificial Intelligence Tools, Vol. 9(4), pp. 509-526,

51. Nasraoui, O., Frigui, H., Joshi, A., Krishnapuram, R.(1999), Mining Web Access Logs Using Relational Competitive Fuzzy Clustering. 8th Intl. Fuzzy Systems Association World Congress - IFSA 99,

52. Nasraoui, O., Petenes, C.(2003),Combining Web Usage Mining and Fuzzy Inference for Website Personalization". In WEBKDD 2003: Web Mining as Premise to Effective Web Applications, pp. 37-46.

53. Newell, S. C., (1997), User models and filtering agents for improved internet information retrieval. Journal of User Modeling and User- Adapted Interaction.

54. Perkowitz, M. and Etzioni, O., (2000),Towards Adaptive Web Sites: Conceptual Framework and Case Study. Artificial Intelligence 118(1-2), pp 245-275.

55. Perlack R.D. (1995), Determination of the Potential Market Size and Opportunities for Biomass-to-Electricity Projects in China, from the Proceedings, Second Biomass Conference of the Americas: Energy, Environment, Agriculture, and Industry pp. 49-54. Meeting held August 21-24, Portland, Oregon; published by National Renewable Energy Laboratory, Golden, Colorado.

56. Resnick & Varian (1997), Recommender systems. Communications of the ACM

57. Resnick P., Iacovou N., Suchak M., Bergstorm P., and Riedl J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, pages 175- 86, Chapel Hill, North Carolina, 1994. ACM.

58. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J. (1994). An open architecture for collaborative filtering of netnews. In: Proceedings of CM Conference on Computer Supported Cooperative World

59. Rich, E.A (1979), User Modeling via Stereotypes. Cognitive Science 3, pp 329-354.

60. Roy, B. (1996), *Multicriteria Methodology for Decision Aiding*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

61. S. N. Schiaffino, A.Amandi, User profiling with Case-Based Reasoning and Bayesian Networks

62. S.E Middleton, N.R Shadbolt, D.C De Roure, Ontological User Profiling in Recommender Systems, Intelligence, Agents, Multimedia Group, University of Southampton

63. S.Fong,Y.Ho,Y.Hang Using Genetic Algorithm for Hybrid Modes of Collaborative Filterig in Online Recommenders http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4626625

64. S.M. Brown, E.Santos Jr, S. B. Banks, Utility Theory-Based User Models for Intelligent Interface Agents

65. Schafer J.B, Konstan J.A, Riedl J (2000), E-commerce Recommendation Applications

66. Shardanand U & Maes P. (1995), Social information filtering: algorithms for automating word of mouth

67. Srinivas, M., Patnaik, L.M. (1994), Genetic algorithms: A survey.

68. Stathacopoulou, R., Grigoriadou, M., Magoulas, G.D.(2003),A Neuro-fuzzy Approach in Student Modeling. Proceedings of the 9th Int. Conf. on User Modeling, UM2003, LNAI 2702, pp. 337-342.

69. Stefani , A. & Strapparava, C., (1999),Exploiting NLP techniques to build user Model for Web sites: The use of WorldNet in SiteIF Project, Second Workshop on Adaptive Systems and User Modeling on the World Wide Web, 8th International World Wide Web Conference, Toronto Canada, May 11-14-1999.

70. T. Berners-Lee, L. Masinter, and M. McChahill. *RFC 1738*: Univorm resource locators *(URL)*. (1994).

71. T. Berners-Lee, R. T. Fielding, and H. Frystyk. *RFC (1945):* Hypertext Transfer Protocol -*HTTP/1.0*. 1996. pp 6

72. Using Genetic Algorithms for Personalized Recommendation Chein-Shung Hwang, Yi-Ching Su, and Kuo-Cheng Tseng, (2010).

73. Van Nostrand Reinhold, Davis, L.(1991). Handbook of Genetic Algorithms.

74. Wei J, Moreau L & Jennings N.R (2005), A market-based approach to recommender systems,

75. Wei-Po Lee, Chih-Hung Liu, Cheng-Che Lu, Intelligent agent-based systems for personalized recommendations in Internet commerce ,Expert Systems with Applications 22, 2002, pp 275-284

76. Zukerman I., Albrecht D. W, Predictive Statistical Models for User Modeling, User Modeling & User-Adapted Interaction vol. 11, 2001, pp 5-18

77. Ματσατσίνης Ν.(2010), Συστήματα υποστήριξης Αποφάσεων, Νέες Τεχνολογίες

78. Παλαιολόγος Κ.(2009), Ταξινόμηση με χρήση αλγορίθμων Data Mining και ασαφούς λογικής: Μια εφαρμογή στο μετρό του Παρισιού, Πολυτεχνείο Κρήτης