



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής

Πρόγραμμα Μεταπτυχιακών Σπουδών

«Προηγμένα Συστήματα Πληροφορικής»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	ΓΡΑΦΟΘΕΩΡΗΤΙΚΕΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΗΣ GRAPHIC AND THEORETICAL METHODS FOR CREATING RECOMMENDATION SYSTEMS
Όνοματεπώνυμο Φοιτητή	Όνομα και επώνυμο: Μπαλτζή Βασιλική
Πατρώνυμο	Όνομα πατέρα: Σπυρίδων
Αριθμός Μητρώου	ΜΠΣΠ/ 11025
Επιβλέπων	Όνομα Επώνυμο, Βαθμίδα Τσιχριτζής Γεώργιος

Ημερομηνία Παράδοσης Μήνας Έτος
28 Απριλίου 2015

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Όνομα Επώνυμο
Βαθμίδα
Τσιχριτζής Γεώργιος

Όνομα Επώνυμο
Βαθμίδα
Μαρία Βίρβου

Όνομα Επώνυμο
Βαθμίδα

- Η πρώτη παράγραφος κάθε ενότητας να μην έχει εσοχή πρώτης γραμμής ενώ οι επόμενες να έχουν εσοχή πρώτης γραμμής.
- Τα Headings να είναι όλα με font **Arial Black** και όχι Bold. Το **Heading 1 να είναι 12pt**, το **Heading 2 να είναι 11pt**, το **Heading 3 να είναι 10pt**. Να μην χρησιμοποιείτε Heading 4 και πέρα.
- Να μην αφήνετε κενές γραμμές πριν ή μετά από τα headings και κάθε επίπεδο heading να απέχει 18pt before και 6pt after.
- Οι λεζάντες (captions) στα σχήματα και τους πίνακες να είναι αριστερά στοιχισμένες και να είναι **Arial bold 9pt**.
- Σε κάθε σελίδα να υπάρχει footer (Arial 8pt) με τον τίτλο της διατριβής στα αριστερά. Στο footer επίσης να υπάρχει αρίθμηση σελίδας στα δεξιά και πάλι με font Arial 8pt.
- Σε κάθε σελίδα να υπάρχει header (Arial 8pt) με το όνομα του φοιτητή στα δεξιά και το λεκτικό «Μεταπτυχιακή Διατριβή» στα αριστερά.
- Τα header και footer να απέχουν από τα άκρα του χαρτιού 2.5cm (στο Page Setup).

Πίνακας περιεχομένων

ΠΕΡΙΛΗΨΗ	7
2.1 Ορισμός	11
2.2 Διαδικασία Σύστασης Και Οι Μορφές Της	12
2.3 Διαχωρισμός Των Συστημάτων Συστάσεων	12
2.4 Συστήματα Βασισμένα Στο Περιεχόμενο (Content Based)	13
2.4.1 Μειονεκτήματα των συστημάτων που βασίζονται στο περιεχόμενο	15
2.5 Συστήματα Βασισμένα στην Συνεργασία (Collaborative Systems)	15
2.5.1 Πλεονεκτήματα Των Συστημάτων Βασισμένα στην Συνεργασία	18
2.5.2 Συστήματα Προσανατολισμένα Στην Μνήμη (Memory Based).....	18
2.5.3 Συστήματα Προσανατολισμένα Στο Μοντέλο (Model Based systems)	23
2.5.4 Βελτίωση Των Αποτελεσμάτων Των Συστημάτων Βασισμένα Στην Συνεργασία	25
2.5.5 Προβλήματα Των Συστημάτων Βασισμένα Στην Συνεργασία	26
2.6 Χρήση Δημογραφικών Δεδομένων (Demographic-Based).....	28
2.6.1 Περιορισμοί Των Δημογραφικών Προβλέψεων	28
2.7 Συστήματα Βασισμένα Στην γνώση (Knowledge- Based Systems).....	28
2.7.1 Συστήματα Βασισμένα Στην Υπόθεση (Case- Based System)	29
2.7.2 Συστήματα Βασισμένα Στον Στόχο (Goal- Based Systems)	29
2.7.3 Πλεονεκτήματα Των Συστημάτων Βασισμένα Στην Γνώση	32
2.7.3 Μειονεκτήματα Των Συστημάτων Βασισμένα Στην Γνώση	32
2.8 Αξιολόγηση Των Συστημάτων Συστάσεων	33
2.9 Υβριδικά Συστήματα (Hybrid Systems)	34
2.9.1 Παράδειγμα Υβριδικών Συστημάτων.....	35
3.1 Ταυτόχρονη Εφαρμογή Των Συστημάτων Βασισμένα Στο Περιεχόμενο Και Των Συστημάτων Βασισμένα Στην Συνεργασία	36
3.2 Χρήση Συστημάτων Με Βάση Την Συνεργασία Με Στοιχεία Των Συστημάτων Με Βάση Το Περιεχόμενο	36
3.3 Χρήση Συστημάτων Βασισμένα Στην Συνεργασία Σε Συστήματα Με Βάση Το Περιεχόμενο	37

3.4 Νέο Μοντέλο Με Χρήση Συστημάτων Με Βάση Το Περιεχόμενο Και Συστημάτων Βασισμένα Στην Συνεργασία.....	37
3.5 Υβριδικά Συστήματα Συντελεστών Βαρύτητας	37
3.6 Συστήματα Εναλλαγής Τεχνικών	38
3.7 Μικτά Συστήματα	40
3.8 Υβριδικά Συστήματα Τύπου Καταρράκτη	41
3.9 Υβριδικά Συστήματα Συνδυασμού Χαρακτηριστικών	42
3.10 Υβριδικά Συστήματα Μεταφοράς Μοντέλου.....	43
Βιβλιογραφία	55
Παράρτημα.....	

Λίστα Εικόνων

Εικόνα 1: Πίνακας με την βαθμολογία των χρηστών για κάθε ταινία	9
Εικόνα 2: Περιγραφή Συστήματος Βασισμένο Στο Περιεχόμενο	15
Εικόνα 3: Περιγραφή Συστήματος Βασισμένο Στη Συνεργασία.....	16
Εικόνα 4: Περιγραφή Του Τρόπου Λειτουργίας Ενός Συστήματος Βασισμένο Στη Συνεργασία	16
Εικόνα 5: Περιγραφή Τρόπου Δημιουργίας Ομάδων Με Κοινά Ενδιαφέροντα	17
Εικόνα 6: Περιγραφή Του Τρόπου Λειτουργίας Των Συστημάτων Βασισμένα Στην Γνώση	29
Εικόνα 7: Το Σύστημα Entrée Από Το Chicago	31
Εικόνα 8: Σύσταση Της Ιστοσελίδας Amazon Βασισμένη Στην Συνεργασία	35
Εικόνα 9: Σύσταση Της Ιστοσελίδας Amazon Βασισμένο Στο Περιεχόμενο	35
Εικόνα 10: Σύσταση Της Ιστοσελίδας Amazon Βασισμένα Στο Ιστορικό Του Χρήστη	36
Εικόνα 11: Περιγραφή Λειτουργίας Συστήματα Συντελεστών Βαρύτητας	38
Εικόνα 12: Το Σύστημα Gixο.....	40
Εικόνα 13: Το Σύστημα PTV.....	41
Εικόνα 14: Περιγραφή λειτουργίας Συστημάτων Τύπου Καταρράκτη	42
Εικόνα 15: Περιγραφή Λειτουργίας Συστημάτων Μεταφοράς Μοντέλου	43

ΠΕΡΙΛΗΨΗ

Η εργασία αυτή δημιουργήθηκε στο πλαίσιο της πτυχιακής μου εργασίας και ασχολείται με τα συστήματα συστάσεων (Recommended Systems) και την ανάλυση αυτών. Στο πρώτο κεφάλαιο υπάρχει μια εισαγωγή η οποία περιγράφει στον αναγνώστη το θέμα και δίνει τα βασικά στοιχεία με τα οποία θα ασχοληθεί η εργασία.

Στο δεύτερο κεφάλαιο ορίζονται τα συστήματα σύστασης, περιγράφεται η διαδικασία με την οποία γίνεται η σύσταση και διαχωρίζονται οι μορφές των συστάσεων. Ακόμα περιγράφονται αναλυτικά τα συστήματα τα οποία δημιουργούν τις συστάσεις, τα πλεονεκτήματα και τα μειονεκτήματα τους. Τέλος αναφέρεται και η έννοια των υβριδικών συστημάτων τα οποία δημιουργούν συστάσεις με μεγαλύτερη αξιοπιστία από τα συμβατικά συστήματα και δίνετε ένα χαρακτηριστικό παράδειγμα ενός υβριδικού συστήματος συστάσεων.

Στο τρίτο κεφάλαιο περιγράφονται αναλυτικά κάποια υβριδικά συστήματα τα οποία χρησιμοποιούνται συχνά από πολλά συστήματα. Περιγράφεται η λειτουργία τους, τα πλεονεκτήματα και τα μειονεκτήματά τους. Υπάρχουν τέλος και αρκετά σχεδιαγράμματα και παραδείγματα τα οποία χρησιμοποιούνται για την καλύτερη κατανόηση των λειτουργιών των συστημάτων από τους αναγνώστες.

Στο τέταρτο κεφάλαιο περιγράφεται υλοποίηση ενός αλγορίθμου ο οποίος σύμφωνα με ένα δείγμα 948 χρηστών και 1682 ταινιών, δημιουργεί προβλεπόμενες τιμές σε έναν αραιό πίνακα. Τέλος παραθέτονται και οι σχετικοί πίνακες των μετρήσεων των τελικών αποτελεσμάτων.

ABSTRACT

This thesis presents the Recommended Systems. In particular, the thesis is organized as follows. Chapter 1 constitutes the introduction of the thesis which describes the objective and the basic elements of the thesis. The Recommended Systems are defined in the second chapter. Moreover, the second chapter includes the procedure that the recommendation is done and the different types of recommendation. In addition, the systems that create recommendations, as well as, their advantages and drawbacks are, also, explained in chapter 2. Finally, chapter 2 introduces the hybrid systems that create recommendations and are, generally, more reliable than the traditional systems. An example of a hybrid system is, also, presented. The operation, the advantages and the disadvantages of known hybrid systems, which are usually used by other systems, are presented in more details in chapter 3. Additionally, figures and examples of hybrid systems are illustrated in chapter 3. Chapter 4 includes the implementation of an algorithm that takes as input a sample of 948 users and 1682 movies and outputs predicted values in a sparse matrix. The corresponding tables of the measurements are, also, presented in chapter 4. Finally, Chapter 5 concludes the thesis.

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Στις μέρες μας, υπάρχει μια μεγάλη κατηγορία εφαρμογών Web που περιλαμβάνει προβλέψεις σχετικά με τις επιλογές και τις προτιμήσεις των χρηστών, που χρησιμοποιούν αξιολογήσεις σε διάφορες επιλογές τους. Μια τέτοια εγκατάσταση ονομάζεται σύστημα προτάσεων. Τα συστήματα προτάσεων χρησιμοποιούν μια πληθώρα τεχνολογιών. Μπορούμε να χωρίσουμε αυτά τα συστήματα σε τρεις μεγάλες κατηγορίες.

- συστήματα με βάση το περιεχόμενο (*content based systems*)

Τα συστήματα αυτά υλοποιούνται με αλγορίθμους όπου στηρίζονται στο περιεχόμενο του στοιχείου που μελετάται.

- συστήματα βασισμένα στην συνεργασία (*collaborative filtering systems*)
- συστήματα βασισμένα στην γνώση (*knowledge-based systems*)
- Τα υβριδικά συστήματα συνεργασίας (*Hybrid Systems*).

Τα συστήματα αυτά υλοποιούνται με αλγορίθμους όπου οι προτάσεις επηρεάζονται από τις προτιμήσεις ενός συνόλου χρηστών που χρησιμοποιούν την ίδια εφαρμογή. Βασίζονται στις αξιολογήσεις των χρηστών για κάποια συγκεκριμένα αντικείμενα, σε ένα συγκεντρωτικό πίνακα που περιλαμβάνει τις προτιμήσεις i χρηστών για j αντικείμενα ή στις γνώσεις που έχουν για τα προϊόντα ή τους χρήστες.

Ο συγκεντρωτικός πίνακας αποτελείται από δύο κλάσεις οντοτήτων οι οποίες δηλώνουν την εξάρτηση των χρηστών με τα αντικείμενα που μελετώνται. Στην δική μας μελέτη οι γραμμές εκφράζουν τους χρήστες και οι στήλες τα αντικείμενα. Αντικείμενα μπορούν να θεωρηθούν για παράδειγμα ένα σύνολο ταινιών.

Τα συστήματα προτάσεων λειτουργούν με μια κλίμακα προτίμησης δηλαδή έναν ακέραιο αριθμό που εκφράζει την αξιολόγηση του χρήστη σε ένα αντικείμενο. Συνήθως αναπαρίστανται με αστέρια, και κάθε αστέρι αθροιστικά χαρακτηρίζει-αποδίδει μια τιμή.

Καθώς ο χρήστης χρησιμοποιεί την εκάστοτε ιστοσελίδα-εφαρμογή, καλείται στο τέλος να αξιολογήσει την επιλογή του βάση αυτής της κλίμακας. Τα στοιχεία εισέρχονται στον πίνακα ως ένα διατεταγμένο σύνολο. Για παράδειγμα κάθε στοιχείο A_{ij} του πίνακα είναι η βαθμολογία του αντικειμένου, του χρήστη i ως προς το στοιχείο j .

Χαρακτηριστικό μερικών χρηστών είναι η άρνηση απόδοσης βαθμολογίας για κάποια αντικείμενα, είτε επειδή δεν είναι γνώστες αυτής της λειτουργίας είτε επειδή αγνοούν τη σημαντικότητα της. Το αποτέλεσμα (της άρνησης) αυτών είναι να δημιουργηθεί ένας πίνακας αραιός, όπου άλλοτε υπάρχει βαθμολογία και άλλοτε όχι (έλλειψη). Αυτό βέβαια καθιστά ιδιαίτερα δύσκολη την λειτουργία των συστημάτων προτάσεων που μελετάμε διότι ένας αραιός πίνακας χαρακτηρίζεται ως επιβλαβής για θέματα που απαιτούν ένα μεγάλο ποσοστό ακρίβειας για την απόδοση μιας συγκεκριμένης πρότασης.

Εφόσον λοιπόν δημιουργηθεί ο πίνακας τα συλλογικά συστήματα φιλτραρίσματος έχουν την ιδιότητα να προβλέπουν μια βαθμολογία για το στοιχείο που δεν έχει κριθεί-βαθμολογηθεί ακόμα χρησιμοποιώντας βαθμολογίες άλλων χρηστών που δηλώνουν σχεδόν ίδιες προτιμήσεις. Για παράδειγμα αν υποθέσουμε ότι ένα σύνολο χρηστών βαθμολογεί ένα σύνολο ταινιών.

Εικόνα 1: Πίνακας με την βαθμολογία των χρηστών για κάθε ταινία

	u_1	u_2	u_3	u_4	u_5	u_6
A	4			5	1	
B	5	5	4			
C				2	4	5
D		3				3

Στο παράδειγμά μας στις γραμμές αναπαρίστανται οι χρήστες και στις στήλες οι ταινίες. Αναφέρουμε ως HP1,HP2,HP3 την τριλογία του Χάρι Πότερ ,TW το twilight και SW1,SW2,SW3 το Starwars1, Starwars2, Starwars3. Τα στοιχεία όπου είναι κενά είναι αυτά τα οποία δεν βαθμολογήθηκαν ποτέ από το χρήστη. Σκοπός λοιπόν των συλλογικών συστημάτων φιλτραρίσματος είναι να προβλέψει τα κενά σύμφωνα με τα στοιχεία του πίνακα. Για παράδειγμα παρατηρούμε ότι ο χρήστης A έχει βαθμολογήσει την ταινία του HP1 με τέσσερα, ενώ δεν έχει βαθμολογήσει την HP2 και HP3.Ο χρήστης B έχει αξιολογήσει την ίδια ταινία με πέντε και τις υπόλοιπες με πέντε και τέσσερα αντίστοιχα. Σύμφωνα με την προτίμηση του B, εάν ο A παρακολούθησε την πρώτη ταινία του Χάρι Πότερ και του άρεσε θα μπορούσαμε να κάνουμε μια γρήγορη εκτίμηση και για τα δύο επόμενα μέρη σύμφωνα με την εκτίμηση του γείτονά του. Αυτό κατά συνέπεια σημαίνει ότι θα έβλεπε με ευχαρίστηση και τα δύο επόμενα μέρη. Άρα η πρόταση που θα δημιουργούταν από το σύστημα θα ήταν σαφέστατα για τον A, το δεύτερο μέρος του Χάρι Πότερ καθώς και το τρίτο.

ΚΕΦΑΛΑΙΟ 2: ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ (Recommended systems)

2.1 Ορισμός

Με την ανάπτυξη του διαδικτύου το μέγεθος των πληροφοριών που είναι διαθέσιμο για το κοινό είναι τόσο μεγάλο που οι άνθρωποι έχουν πρόβλημα να τις συλλέξουν και να τις χρησιμοποιήσουν για τις ανάγκες τους. Για τον λόγο αυτό δημιουργήθηκαν διάφορες μέθοδοι για να γίνετε ευκολότερα ο διαχωρισμός των περιττών από των χρήσιμων πληροφοριών. Οι μέθοδοι αυτοί έχουν την δυνατότητα με την χρήση κάποιων αυτοματοποιημένων τεχνικών να δίνουν στον χρήστη ακριβώς αυτό που επιθυμεί και όχι άχρηστες ή κακόβουλες πληροφορίες. Τα συστήματα που υπάρχουν για να φιλτράρουν τις πληροφορίες μπορούν να διακριθούν σε δύο κατηγορίες (Malone, Grant, Turbak, Brobst, & Cohen, 1987):

- Τα συστήματα Cognitive filtering, τα οποία μπορούν να καταλάβουν τι ακριβώς αναζητάει ο χρήστης και να χρησιμοποιήσουν ειδικά φίλτρα για να αντιστοιχίσουν το περιεχόμενο της αναζήτησης με το περιεχόμενο των πληροφοριών. Αυτό το φιλτράρισμα γίνεται με την χρήση συγκεκριμένων λέξεων ως κλειδιά αναζήτησης.
- Τα συστήματα Sociological filtering τα οποία αναζητούν πληροφορίες ανάλογα με τις προσωπικές σχέσεις τις οποίες έχουν οι χρήστες.

Τα συστήματα συστάσεων (Recommender Systems) αποτελούν την εξιδανίκευση των συστημάτων φιλτραρίσματος της πληροφορίας και στόχος τους είναι να παρουσιάσουν στον χρήστη τις πληροφορίες που τον ενδιαφέρουν προσωπικά. Τα συστήματα συστάσεων δημιουργήθηκαν για να μπορέσουν να επιλύσουν τα βασικά προβλήματα των συστημάτων τα οποία είναι βασισμένα σε κείμενο. Τα προβλήματα αυτά πηγάζουν από τον μεγάλο όγκο δεδομένων που υπάρχουν διαθέσιμα για αναζήτηση.

Ως συστήματα συστάσεων ορίζονται τα συστήματα όπου οι άνθρωποι παρέχουν προτάσεις ως εισόδους οι οποίες έπειτα συγκεντρώνονται και κατευθύνονται σε συγκεκριμένους παραλήπτες (Resnick & Varian, 1997).

Οι τομείς οι οποίοι ώθησαν την δημιουργία των συστημάτων συστάσεων είναι (Papagelis, 2005):

- Ανάκτηση πληροφοριών: Τα συστήματα ανάκτησης πληροφοριών είναι τα συστήματα τα οποία έχουν την δυνατότητα να αναζητήσουν, να ταξινομήσουν και να ανακτήσουν πολλά δεδομένα τα οποία εκτός από κείμενο μπορούν να είναι και σε άλλες μορφές όπως εικόνα ή ήχος. Επιπλέον τα συστήματα αυτά δίνουν την δυνατότητα στον χρήστη να βελτιστοποιήσει την λίστα των αποτελεσμάτων.
- Η εξατομίκευση: Η εξατομίκευση είναι μία μορφή μάρκετινγκ η οποία προσπαθεί να δημιουργήσει προϊόντα τα οποία είναι ιδανικά για τον κάθε χρήστη ξεχωριστά. Αυτό συμβαίνει έπειτα από αλληλεπίδραση με τον κάθε χρήστη και από τον έλεγχο του ιστορικού των ενεργειών του χρήστη.
- Διαχείριση εμπιστοσύνης: Στο διαδίκτυο υπάρχουν πολλές πληροφορίες οι οποίες προέρχονται από άτομα τα οποία μεροληπτούν πάνω σε ορισμένα θέματα. Έτσι είναι σημαντικό να είναι γνωστή η πηγή της πληροφορίας έτσι ώστε να μπορέσει να κρίνει ο ενδιαφερόμενος αν οι πληροφορίες είναι έγκυρες ή όχι. Ο όγκος των δεδομένων στο διαδίκτυο είναι τόσο μεγάλος που είναι πολύ δύσκολο να αξιολογηθούν όλες οι πηγές που υπάρχουν. Έτσι μια τεχνική που χρησιμοποιείται για να εξακριβωθεί αν μια πηγή είναι ορθή είναι η εμπιστοσύνη σε άτομα τα οποία είναι γνωστά.

Αυτοί και αρκετοί ακόμα τομείς ώθησαν στην δημιουργία του συστήματος συστάσεων το οποίο εξαπλώθηκε με μεγάλη ταχύτητα και εφαρμόστηκε σε πολλές μορφές στο διαδίκτυο όπως το ηλεκτρονικό εμπόριο, συστήματα αξιολόγησης ταινιών ή μουσικής.

2.2 Διαδικασία Σύστασης Και Οι Μορφές Της

Κάθε σύστημα συστάσεων δέχεται μια είσοδο και στην συνέχεια ακολουθεί μια διαδικασία και παράγει κάποια αποτελέσματα. Τα στοιχεία που μπορεί ένα σύστημα συστάσεων να πάρει ως είσοδο προέρχονται είτε από τα δεδομένα του χρήστη, είτε από τα δεδομένα του αντικειμένου είτε από κάποιες αλληλεπιδράσεις μεταξύ του χρήστη με το αντικείμενο.

Κάθε χρήστης δημιουργεί ένα προφίλ στο οποίο έχει μέσα στοιχεία που θα τον βοηθήσουν να έχει σωστά αποτελέσματα από τα συστήματα συστάσεων. Όμοια ένα αντικείμενο έχει κάποια χαρακτηριστικά τα οποία συμπληρώνουν ένα προφίλ σχετικά με το αντικείμενο, και χρησιμοποιούνται για να μπορούν να δίνουν την αντιστοίχιση από τα συστήματα συστάσεων.

Τα δεδομένα στο προφίλ του χρήστη μπορούν να εισαχθούν είτε άμεσα είτε έμμεσα. Άμεση είναι η διαδικασία κατά την οποία ο χρήστης βάζει στο σύστημα τα προσωπικά του δεδομένα και πληροφορίες για τα ενδιαφέροντά του. Έμμεση είναι η διαδικασία κατά την οποία το σύστημα αντλεί πληροφορίες για τον χρήστη από τις προσωπικές του σελίδες, είτε από τις πληροφορίες που παρέχει το προσωπικό του κινητό τηλέφωνο είτε από το ιστορικό των αγορών που έχει κάνει.

Ο βαθμός αλληλεπίδρασης του χρήστη και του αντικειμένου μπορεί να γίνει με πολλούς τρόπους. Μπορεί για παράδειγμα να ζητηθεί από το σύστημα στον χρήστη να βαθμολογήσει τον βαθμό της χρησιμότητας ενός αντικειμένου για αυτόν προσωπικά. Ο τρόπος που μπορεί να γίνει η βαθμολόγηση μπορεί να διαφέρει επίσης. Μπορεί για παράδειγμα ο χρήστης είτε να απαντήσει με ένα ναι ή με ένα όχι, είτε να βαθμολογήσει το αντικείμενο από μία κλίμακα τιμών, είτε να γράψει με απλό κείμενο την άποψή του για το συγκεκριμένο αντικείμενο. Πιο συχνά χρησιμοποιούνται τα συστήματα που χρησιμοποιούν μορφές αξιολόγησης που δεν απαιτούν πολύ χρόνο γιατί οι χρήστες συνήθως δεν ενδιαφέρονται να αξιολογήσουν τα αντικείμενα αν αυτό απαιτεί αρκετό χρόνο.

Όταν τα στοιχεία του κάθε χρήστη εισαχθούν στο σύστημα τότε αρχίζει να επεξεργάζεται τα δεδομένα και να τα παρουσιάζει είτε σε όλους τους χρήστες είτε σε αυτούς που ζήτησαν να μάθουν τα αποτελέσματα. Η διαδικασία σύστασης μπορεί να γίνει σε διάφορες μορφές, μπορεί να γίνει είτε ένας προς έναν το οποίο γίνεται όταν ένας χρήστης απευθύνεται σε έναν άλλον χρήστη. Αυτή η μορφή σύστασης γίνεται όταν ένας χρήστης έχει χρησιμοποιήσει ένα προϊόν και το προτείνει σε έναν άλλο χρήστη όπως ακριβώς γίνεται και με την συνομιλία δύο φίλων. Στις ηλεκτρονικές μορφές σύστασης ένας χρήστης μπορεί να κάνει μια σύσταση στον εαυτό του ώστε να μην ξεχάσει να αγοράσει ή να δοκιμάσει ένα αντικείμενο (Reischach, Michahelles, & Schmidt).

Μια ακόμα μορφή σύστασης είναι όταν συγκεντρώνονται πολλοί χρήστες σε έναν τόπο και αξιολογούν μαζικά ένα προϊόν και αυτή η αξιολόγηση είναι διαθέσιμη σε όλους τους ενδιαφερόμενους χρήστες. Η μορφή αυτή όπως είναι φανερό είναι από πολλούς χρήστες και απευθύνετε σε πολλούς χρήστες γι' αυτόν τον λόγο ο ενδιαφερόμενος θα πρέπει να αποφασίσει από μόνος του αν θα του είναι χρήσιμο ένα προϊόν.

Επιπλέον υπάρχουν περιπτώσεις όπου αξιολογήσεις από πολλούς χρήστες συγκεντρώνονται και αθροίζονται μόνο για έναν χρήστη. Επειδή όμως κάθε άνθρωπος έχει διαφορετικά ενδιαφέροντα είναι πολύ πιθανό να μην συμφωνεί με την πλειοψηφία σχετικά με ένα συγκεκριμένο προϊόν. Γι' αυτόν τον λόγο αναπτύχθηκαν συγκεκριμένοι αλγόριθμοι οι οποίοι έχουν την ικανότητα να συγκεντρώνουν τα στοιχεία του κάθε χρήστη και να προτείνει στον έναν ενδιαφερόμενο το προϊόν με τις κριτικές από άτομα που έχουν τα ίδια ενδιαφέροντα σύμφωνα με το προφίλ τους. Η διαδικασία αυτή ονομάζεται συνεργατικό φίλτράρισμα (Reischach, Michahelles, & Schmidt).

Η τελευταία μορφή της διαδικασίας σύστασης που μπορεί να αναφερθεί είναι όταν κάποιος ειδικός σε ένα θέμα κάνει μια κριτική η οποία επηρεάζει πολύ κόσμο. Για παράδειγμα όταν ένας κριτικός θέατρου παρακολουθήσει μια παράσταση και μοιραστεί τις απόψεις του σε ένα μέσο κοινωνικής δικτύωσης όπου τον ακολουθούν χιλιάδες άτομα. Αυτή η μορφή ανταλλαγής απόψεων για ένα θέμα έχει γίνει πολύ δημοφιλής τα τελευταία χρόνια και πολλοί άνθρωποι σπεύδουν να συμβουλευτούν από ειδικούς από κριτικές που υπάρχουν στις σελίδες κοινωνικής δικτύωσης.

2.3 Διαχωρισμός Των Συστημάτων Συστάσεων

Τα συστήματα συστάσεων χωρίζονται σε διάφορες κατηγορίες οι οποίες διαφέρουν στο τρόπο με τον οποίο γίνονται οι συστάσεις μεταξύ των χρηστών (Balabanovic & Shoham, <https://www.ischool.utexas.edu/>, 1997):

- Τα συστήματα που είναι βασισμένα στο περιεχόμενο (Content Based). Σε αυτό το σύστημα οι πληροφορίες σχετικά με το κείμενο και τα χαρακτηριστικά του συγκεντρώνονται σε ένα προφίλ το οποίο μπορούν να χρησιμοποιήσουν τα συστήματα συστάσεων για να προβλέψουν αν ο χρήστης τα θεωρεί χρήσιμα ή όχι.
- Τα συστήματα τα οποία είναι βασισμένα στην συνεργασία (Collaborative Systems). Τα συστήματα αυτά κρίνουν αν ένας χρήστης θα τα θεωρεί χρήσιμες κάποιες πληροφορίες ανάλογα με το ιστορικό που υπάρχει. Εξετάζουν δηλαδή αν οι χρήστες που έμαχθαν παλαιότερα παρόμοιες πληροφορίες τις βρήκαν χρήσιμες και ανάλογα αποφασίζουν αν είναι χρήσιμες ή όχι.
- Τα συστήματα βασισμένα στην γνώση (Knowledge- Based Systems). Σε αυτήν την κατηγορία τα συστήματα στηρίζονται πάνω στην γνώση τους για τα αντικείμενα, για τους χρήστες και για τον τρόπο κάλυψης των αναγκών και κάνουν τις ανάλογες συστάσεις προς τους χρήστες.
- Τα υβριδικά συστήματα συνεργασίας (Hybrid Systems). Τα συστήματα αυτά χρησιμοποιούν ένα συνδυασμό των δύο παραπάνω κατηγοριών για να παρουσιάσουν όσο τον δυνατόν καλύτερα αποτελέσματα.

2.4 Συστήματα Βασισμένα Στο Περιεχόμενο (Content Based)

Στην συγκεκριμένη τεχνική προτείνονται στο χρήστη αντικείμενα όμοια με αυτά που είχε προτιμήσει στο παρελθόν (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005). Οι αξιολογήσεις εκφράζονται από ένα χρήστη και δεν έχει καμία εξάρτηση με όλους τους υπόλοιπους. Το κύριο σημείο εδώ είναι το ίδιο το αντικείμενο. Δηλαδή τα χαρακτηριστικά που το εκφράζουν και το πώς το ένα αντικείμενο διαφέρει από τα υπόλοιπα. Κάθε αντικείμενο είναι διαφορετικό έχοντας διαφορετικά γνωρίσματα που το περιγράφουν, για παράδειγμα ας σκεφτούμε ότι έχουμε ένα διδιάστατο πίνακα A με χρήστες και αντικείμενα. Ως αντικείμενα εκφράζουμε τις ταινίες που μπορεί να επιλέξει ο χρήστης, ενώ ως χαρακτηριστικά θα μπορούσαμε να εκφράσουμε για παράδειγμα συγκεκριμένους ηθοποιούς, σκηνοθέτες, είδη, το θέμα κ.α

Η χρησιμότητα $u(c,s)$ του αντικειμένου S ως προς τον χρήστη c βασίζεται στη χρησιμότητα $u(c,s_i)$ για κάθε $s_i \in S$ όπου είναι παρόμοιο με το αντικείμενο S (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005). Σε προσπάθεια πρότασης μιας ταινίας σε ένα χρήστη U , τα συστήματα φιλτραρίσματος με βάση το περιεχόμενο θα προσπαθήσουν να κατανοήσουν τις ομοιότητες μεταξύ των ταινιών, όπου ο χρήστης U έχει βαθμολογήσει υψηλά (συγκεκριμένους ηθοποιούς, σκηνοθέτες, είδη, το θέμα κ.α), και μόνο όσες έχουν υψηλό βαθμό ομοιότητας σε σχέση με τις προτιμήσεις του χρήστη θα προταθούν. Η ανάλυση του περιεχομένου των αντικειμένων είναι μια από τις βασικές λειτουργίες της συγκεκριμένης τεχνικής. Η παραπάνω μέθοδος βασίζεται στην ανάκτηση πληροφοριών καθώς και την αναζήτηση πληροφορίας μέσω φιλτραρίσματος.

Σύγχρονες μέθοδοι φιλτραρίσματος με βάση το περιεχόμενο λειτουργούν ώστε να προτείνουν αντικείμενα που περιέχουν πληροφορίες όπως κείμενα, ιστοσελίδες, ή ειδήσεις. Εδώ συναντάμε μια βελτιωμένη μέθοδο ανάκτησης της πληροφορίας η οποία βασίζεται στο προφίλ των χρηστών. Το προφίλ των χρηστών περιέχει προτιμήσεις, πληροφορίες, και τις ανάγκες των χρηστών.

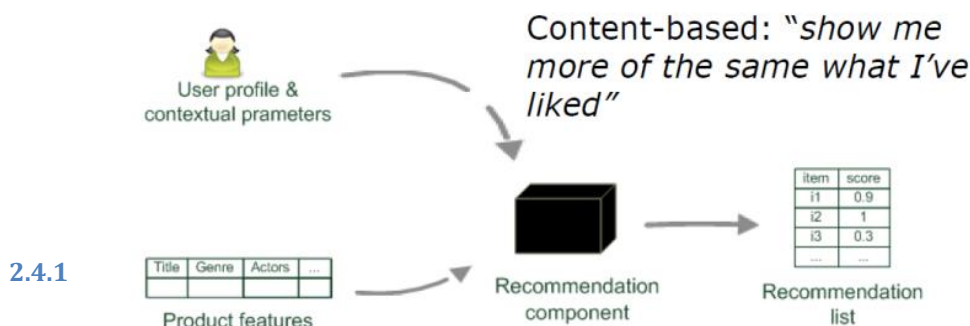
Η πράξη και η θεωρία αποδεικνύει ότι οι πληροφορίες του προφίλ των χρηστών πηγάζουν ρητά από τον ίδιο το χρήστη, μέσω ερωτηματολογίων που υποβλήθηκαν ή από την συναλλακτική συμπεριφορά τους στην πάροδο του χρόνου (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005). Άλλωστε, όπως αναφέρθηκε παραπάνω τα συστήματα αυτά λειτουργούν με βάση το περιεχόμενο. Το περιεχόμενο σε αυτές τις περιπτώσεις αξιολογείται με λέξεις κλειδιά. Η σημαντικότητα της λέξης k_j σε ένα έγγραφο d_j ορίζεται με βάση το βάρος της $w(i,j)$ και υπολογίζεται με διάφορους τρόπους. Η παραπάνω διαδικασία είναι μια μέθοδος δημιουργίας του προφίλ των χρηστών. Σε αυτήν την περίπτωση, χρησιμοποιούνται πολλοί αλγόριθμοι μηχανικής μάθησης (machine-learning algorithms) για να μάθει το σύστημα το προφίλ του χρήστη.

Πυρήνας της παραπάνω προσέγγισης είναι η επεξεργασία του περιεχομένου που περιγράφει τα αντικείμενα που πρόκειται να προταθούν. Τα αντικείμενα μπορεί να είναι τόσο διαφορετικά μεταξύ τους αφού

εξαρτώνται από το πλήθος και το είδος διαφόρων χαρακτηριστικών που τα περιγράφουν. Παρόλο που κάθε αντικείμενο μπορεί να περιγράφεται από τον ίδιο μικρό αριθμό χαρακτηριστικών δεν συμβαίνει το ίδιο και σε αντικείμενα όπως ιστοσελίδες, άρθρα νέων ή έγγραφα. Σε αυτή τη περίπτωση δεν υπάρχουν χαρακτηριστικά με καθορισμένες τιμές και έτσι η χρήση μιας τεχνικής μοντελοποίησης εγγράφων για ομόριζα θεωρείται αναγκαία. Μια μέθοδος παρουσίασης αδόμητων δεδομένων είναι η VSM (Vector Space Model). Η VSM είναι μια αναπαράσταση εγγράφων κειμένου την οποία θα παρουσιάσουμε αναλυτικότερα σε επόμενο κεφάλαιο – ενότητα.

Η τεχνική του συνεργατικού φιλτραρίσματος εξυπηρετεί το χρήστη που αναφέρει «δείξτε μου κι άλλα περισσότερα, όπως αυτά που μου είχαν αρέσει στο παρελθόν» (Zanker, 2010).

Εικόνα 2: Περιγραφή Συστήματος Βασισμένο Στο Περιεχόμενο



Μειονεκτήματα των συστημάτων που βασίζονται στο περιεχόμενο

Τα συστήματα που βασίζονται στο περιεχόμενο έχουν κάποιους περιορισμούς όπως (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005):

- Περιορισμοί οι οποίοι υπάρχουν από την ανάλυση του περιεχομένου. Τα συστήματα τα οποία λειτουργούν με βάση το περιεχόμενό τους δυσκολεύονται να αξιολογήσουν το περιεχόμενο ενός βίντεο ή μιας φωτογραφίας. Η αξιολόγηση τους γίνεται με την χρήση αυτοματοποιημένων αλγορίθμων και το αποτέλεσμα τους δεν είναι σίγουρο πως μπορεί να είναι πάντα σωστό. Επιπλέον σε κάποια κείμενα τα οποία χαρακτηρίζονται με το ίδιο προφίλ, τα συστήματα συστάσεων με βάση το περιεχόμενο δεν μπορούν να ξεχωρίσουν πιο κείμενο είναι αυτό που χρειάζεται ο χρήστης και πιο όχι.
- Επιπλέον τα συστήματα αυτά επιλέγουν πληροφορίες οι οποίες έχουν υψηλό βαθμό συσχέτισης με το προφίλ του χρήστη. Έτσι αν ένας χρήστης αναζητήσει κάτι στο οποίο δεν ταιριάζει με το προφίλ που είχε δημιουργήσει τότε τα αποτελέσματα θα ήταν πολύ περιορισμένα σε σύγκριση με άλλους χρήστες. Για την αντιμετώπιση αυτού του προβλήματος έχουν δημιουργηθεί αλγόριθμοι οι οποίοι προσθέτουν τυχαιότητα στις προτάσεις του συστήματος.
- Τέλος δημιουργούνται πολλά προβλήματα με τους νέους χρήστες οι οποίοι δεν έχουν συγκεντρώσει πολλά στοιχεία στο προφίλ τους και έτσι τα αποτελέσματα που έχουν από τα συστήματα συστάσεων δεν μπορούν να είναι αξιόπιστα.

2.5 Συστήματα Βασισμένα στην Συνεργασία (Collaborative Systems)

Η τεχνική του φιλτραρίσματος βασισμένα στην συνεργασία στηρίζεται στις αξιολογήσεις των χρηστών που μοιράζονται παρόμοια ενδιαφέροντα μεταξύ τους. Είναι η δημοφιλέστερη και πλέον η πιο κυρίαρχη μέθοδος συστάσεων. Η κεντρική ιδέα των συστημάτων συνεργατικού φιλτραρίσματος είναι ότι αν ένα σύνολο ατόμων αρέσκονται στα ίδια αντικείμενα στο παρελθόν, τότε πιθανότατα θα μπορούν να μοιράζονται και τις ίδιες προτιμήσεις. Για αυτό το λόγο προσπαθούν να προβλέψουν την προτίμηση ενός συγκεκριμένου χρήστη βασιζόμενα σε αντικείμενα που προηγουμένως είχαν βαθμολογηθεί από άλλους χρήστες.

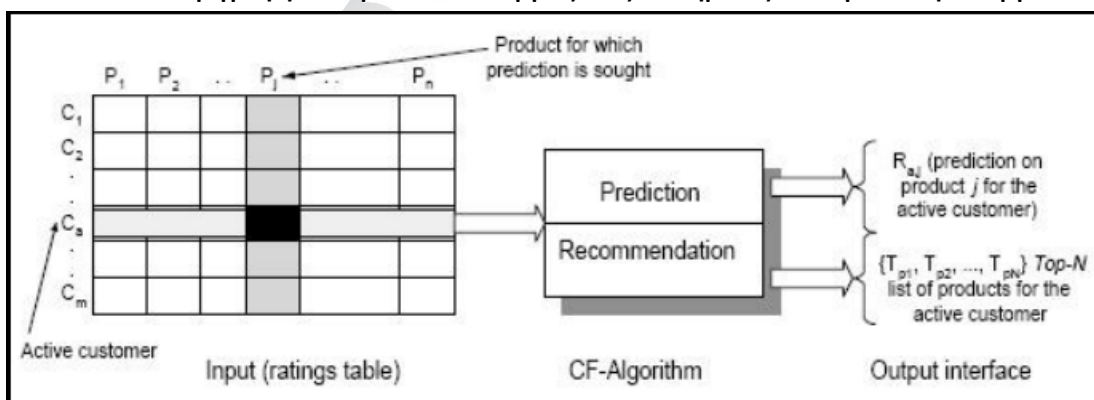
Η τεχνική του φιλτραρίσματος βασισμένα στην συνεργασία εξυπηρετεί το χρήστη που αναφέρει «πείτε μου τι είναι πιο δημοφιλές ανάμεσα στους γείτονές μου;» δηλαδή τι προτιμούν οι υπόλοιποι που είναι όμοιοι μαζί μου. Οι εικόνες παρακάτω περιγράφουν την λογική και τον τρόπο που λειτουργούν τα συστήματα αυτά.



Εικόνα 3: Περιγραφή Συστήματος Βασισμένο Στη Συνεργασία

Η χρησιμότητα $u(c,s)$ του αντικειμένου s ως προς τον χρήστη c βασίζεται στη χρησιμότητα $u(c_j,s)$ που ανατίθεται στο αντικείμενο s για κάθε χρήστη $c_j \in C$ όπου είναι παρόμοιο με το χρήστη c . Σε προσπάθεια πρότασης μιας ταινίας σε ένα χρήστη U , τα συστήματα φιλτραρίσματος με βάση το περιεχόμενο αρχικά θα έβρισκαν τους όμοιους χρήστες, θα δημιουργούσαν τους k πλησιέστερους γείτονες, και ύστερα θα συμπεραίναν την σύσταση για το χρήστη (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005). Οι προβλέψεις για κάποιο αντικείμενο βασίζονται στις βαθμολογίες-προτιμήσεις όμοιων χρηστών. Δηλαδή, τα συστήματα συνεργατικού φιλτραρίσματος, για να παράγουν συστάσεις για ένα χρήστη, χρησιμοποιούν πληροφορία από χρήστες που είναι όμοιοι με αυτόν (Λίκα). Η παρακάτω εικόνα δείχνει πως λειτουργούν τα συστήματα CF για να κάνουν τις συστάσεις τους.

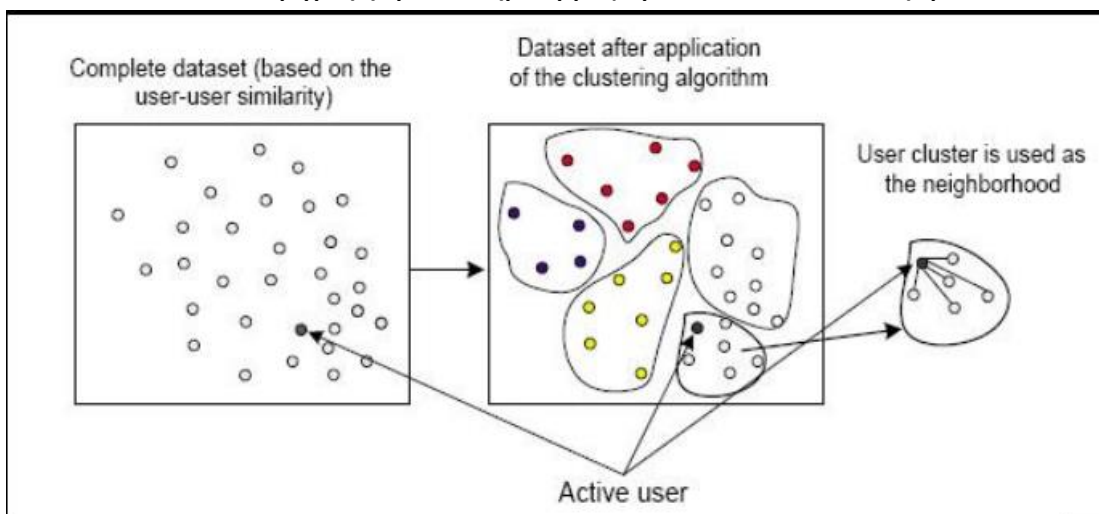
Εικόνα 4: Περιγραφή Του Τρόπου Λειτουργίας Ενός Συστήματος Βασισμένο Στη Συνεργασία



Γενικότερα τα συστήματα CF αναλύουν τις απόψεις των άλλων χρηστών σε σχέση με τα στοιχεία, έτσι παρέχουν ένα βαθμό αρεσκείας όχι με βάση τη φύση του στοιχείου, αλλά βάσει της ανθρώπινης κρίσης.

Η παρακάτω εικόνα δείχνει λεπτομερώς τον τρόπο όπου τα συστήματα με βάση την συνεργασία λειτουργούν. Πως δηλαδή κατά την αναζήτηση ενός αντικειμένου από τον χρήστη οι πληροφορίες που είναι στην διάθεση του συστήματος ομαδοποιούνται ανάλογα με τις προτιμήσεις και πως το σύστημα κάνει τις προτάσεις του μόνο από τους χρήστες που ανήκουν στην ίδια ομάδα.

Εικόνα 5: Περιγραφή Τρόπου Δημιουργίας Ομάδων Με Κοινά Ενδιαφέροντα



Εμπορικές εφαρμογές που υλοποιούν τέτοιο φιλτράρισμα είναι (Μουζακίδης):

- Netflix εκπομπή βίντεο μέσω διαδικτύου
- StumbleUpon κοινωνική δικτύωση• Directed Edge εταιρεία που παρέχει συστάσεις πάνω σε ιστοσελίδες κοινωνικής δικτύωσης, ηλεκτρονικών καταστημάτων και ιστοσελίδων πληροφόρησης.
- eBay διαδικτυακή πύλη ηλεκτρονικού εμπορίου
- Google News υπηρεσία της Google που παρέχει εξατομικευμένα νέα
- Gravity R&D εταιρεία παροχής ψηφιακού περιεχομένου
- half.ebay.com
- Heei i Παρέχει συστάσεις για ιστοσελίδες
- Hollywood Video Διαδικτυακή πύλη παρακολούθησης
- Hulu τηλεόραση μέσω διαδικτύου
- Internet Movie Database παρέχει πληροφορίες για ταινίες
- iTunes διαδικτυακή πύλη με εμπόριο πολυμετοχικού υλικού
- Last.fm διαδικτυακό ραδιόφωνο
- Threadless διαδικτυακή ιστοσελίδα για παραγγελίες t-shirt
- Baynote μηχανή αναζήτησης
- ChoiceStream εταιρεία παροχής διαδικτυακό διαφημίσεων
- Collarity εταιρεία παροχής διαδικτυακών διαφημίσεων και μηχανής αναζήτησης
- Digg.com ιστοσελίδα κοινωνικής δικτύωσης• Amazon ηλεκτρονικό πολυκατάστημα
- Amie Stree t ηλεκτρονικό πολυκατάστημα μουσικής
- Barilliance εφαρμογή υποστήριξης ηλεκτρονικού εμπορίου
- Barnes and Noble ηλεκτρονικό κατάστημα βιβλίων
- Musicmatch Εφαρμογή αναπαραγωγής μουσικής
- MyStrands ιστοσελίδα κοινωνικής δικτύωσης
- LibraryThing πληροφορίες βιβλίων
- Loomia παροχή συστάσεων ιστοσελίδων

Οι αλγόριθμοι των συστημάτων φιλτραρίσματος διακρίνονται σε δύο κύριες κατηγορίες ανάλογα με τον τρόπο εύρεσης της γειτνίασης του χρήστη. Η πρώτη βασίζεται στη μνήμη (*memory/heuristic - based*) και η δεύτερη βασίζεται σε μοντέλο (*model- based*) τα οποία θα αναλύσουμε παρακάτω (Breese, Heckerman, & Kadie, Empirical Analysis Of Predictive Algorithms For CollaBorative Filtering , 1998).

2.5.1 Πλεονεκτήματα Των Συστημάτων Βασισμένα στην Συνεργασία

Τα συστήματα που βασίζονται στην συνεργασία έχουν ένα βασικό πλεονέκτημα που κάνει πολύ πιο απλή και γρήγορη την διαδικασία δημιουργίας και ανάπτυξης ενός τέτοιου συστήματος. Το πλεονέκτημα αυτό είναι πως δεν χρειάζεται να καταναλώνονται πόροι του συστήματος για την συγκέντρωση πληροφοριών σχετικά με το αντικείμενο. Δεν είναι ανάγκη δηλαδή να υπάρχει καμιά λεπτομέρεια σχετικά με το αγαθό καθώς το μόνο που χρειάζεται το σύστημα για να κάνει σωστές συστάσεις είναι οι αξιολογήσεις που δίνουν οι χρήστες του συστήματος για το συγκεκριμένο προϊόν.

Επιπλέον επειδή ο αριθμός των συστάσεων που είναι διαθέσιμες για ένα αντικείμενο αυξάνετε συνεχώς, αυξάνετε έτσι και η ποιότητα των συστάσεων του συστήματος με αποτέλεσμα με το πέρασμα του χρόνου το σύστημα να γίνετε όλο και πιο αξιόπιστο.

2.5.2 Συστήματα Προσανατολισμένα Στην Μνήμη (Memory Based)

Οι αλγόριθμοι με βάση τη μνήμη χαρακτηρίζονται ως «ευρετικοί» και βασίζονται σε μια ολόκληρη συλλογή από προηγούμενα εκτιμημένα αντικείμενα που έχουν αξιολογήσει οι χρήστες. Με τα δεδομένα που έχει στην μνήμη του, το σύστημα υπολογίζει την ομοιότητα ανάμεσα σε προϊόντα ή χρήστες και εξάγει την πρόβλεψη για τον χρήστη από τον σταθμισμένο μέσο όρο των βαθμολογιών. Ο υπολογισμός της ομοιότητας γίνεται με μηχανισμούς όπως η συσχέτιση Pearson (Pearson correlation) ή η ομοιότητα με βάση το συνημιτονικό διάνυσμα (*cosine-based similarity*) που θα αναλυθούν παρακάτω.

Η τιμή της άγνωστης αξιολόγησης $r_{c,s}$ για τον χρήστη c και το αντικείμενο s υπολογίζεται από τη συνάθροιση των αξιολογήσεων των κοντινότερων γειτόνων, για το ίδιο αντικείμενο. Γενικά ο τύπος της μετρικής ορίζεται ως εξής (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005):

$$r_{c,s} = \text{aggr}_{c' \in \hat{C}} r_{c',s},$$

Όπου \hat{c} οι πλησιέστεροι γείτονες που έχουν βαθμολογήσει το στοιχείο s .

Η ειδική φορμαλιστική προσέγγιση αποδίδεται ως εξής:

$$(a) r_{c,s} = \frac{1}{N} \sum_{c' \in \hat{C}} r_{c',s},$$

$$(b) r_{c,s} = k \sum_{c' \in \hat{C}} sim(c, c') \times r_{c',s},$$

$$(c) r_{c,s} = \bar{r}_c + k \sum_{c' \in \hat{C}} sim(c, c') \times (r_{c',s} - \bar{r}_{c'}),$$

Όπου k πολλαπλασιαστής που χρησιμεύει ως παράγοντας ομαλοποίησης και ορίζεται ως

$$k = 1 / \sum_{c' \in \hat{C}} |sim(c, c')|,$$

και όπου \bar{r}_c ο μέσος όρος αξιολογήσεων των χρηστών που ορίζεται ως

$$\bar{r}_c = (1/|S_c|) \sum_{s \in S_c} r_{c,s}, \text{ where } S_c = \{s \in S | r_{c,s} \neq \emptyset\}.$$

Βέβαια, η πιο συνηθισμένη μέθοδος είναι το άθροισμα των βαρών, η οποία προκύπτει από τον (b) τύπο. Η ομοιότητα $sim(c, c')$ μεταξύ δύο χρηστών c, c' βασίζεται στην υπολογιστική μέθοδο της εύρεσης των αποστάσεων. Όσο μεγαλύτερο είναι το βάρος, τόσο μικρότερη είναι η απόσταση και τόσο μεγαλύτερη η ομοιότητα. Υπάρχουν πολλές μέθοδοι για να υπολογιστεί η απόσταση.

Η πιο απλή μέθοδος για να υπολογιστεί το βάρος είναι μέσο της απόστασης από τους τύπους (Adomavicius & Kwon, New Recommendation Techniques for Multicriteria Rating Systems, 2007):

Απόσταση Manhattan

$$w(a, i) = \sum_{j=0}^k |r_{a,j} - r_{i,j}|$$

Η Ευκλείδεια Απόσταση

$$w(a, i) = \sqrt{\sum_{j=0}^k |r_{a,j} - r_{i,j}|^2}$$

Απόσταση Chebyshev

$$w(a, i) = \max_{j=0, \dots, k} |r_{a,j} - r_{i,j}|$$

- Η συσχέτιση κατά Pearson ορίζεται ως:

$$\text{sim}(x, y) = \frac{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)(r_{y,s} - \bar{r}_y)}{\sqrt{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)^2 \sum_{s \in S_{xy}} (r_{y,s} - \bar{r}_y)^2}}$$

Όπου x,y οι δύο χρήστες και $S_{x,y}$ το σύνολο των στοιχείων που αξιολογήθηκαν και από τους δύο χρήστες.

$$S_{xy} = \{s \in S \mid r_{x,s} \neq \emptyset \ \& \ r_{y,s} \neq \emptyset\}.$$

Υπολογίζει την γραμμική συσχέτιση δύο διανυσμάτων και επιστρέφει τιμές από -1 έως 1. Όσο η τιμή πλησιάζει το 1 τα διανύσματα μοιάζουν όλο και περισσότερο ενώ όσο η τιμή πλησιάζει στο -1 τα διανύσματα είναι όλο και πιο αντίθετα. Η τιμή 0 δηλώνει ότι οι τιμές των συνιστωσών των διανυσμάτων είναι γραμμικώς ανεξάρτητες (Μουζακίδης).

Ο τύπος της μετρικής είναι:

$$\text{similarity}(A, B) = \frac{\sum_{i=1}^N (A_i - \mu(A)) * (B_i - \mu(B))}{\sqrt{\sum_{i=1}^N (A_i - \mu(A))^2} * \sqrt{\sum_{i=1}^N (B_i - \mu(B))^2}}$$

Όπου τα $\mu(A)$ και $\mu(B)$ συμβολίζουν τον μέσο όρο των τιμών A,B αντίστοιχα. Οι τιμές A και B μπορεί να συμβολίζουν είναι τους χρήστες και θα δείχνουν τις προτιμήσεις του κάθε χρήστη ενώ αν είναι αντικείμενα τότε οι τιμές συμβολίζουν τις προτιμήσεις των χρηστών στα αντικείμενα.

- Στην συνημιτονική ομοιότητα οι δύο χρήστες x, y εκφράζονται ως δύο διανύσματα σε ένα n-διάστατο χώρο, όπου $m=|S_{x,y}|$. Επομένως η ομοιότητα μεταξύ των διανυσμάτων μπορεί να μετρηθεί με τον υπολογισμό του συνημίτονου της γωνίας μεταξύ τους (Δηλαδή, είναι το συνημίτονο της γωνίας που σχηματίζουν δύο διανύσματα που συγκρίνονται):

$$\text{sim}(x, y) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|_2 \times \|\vec{y}\|_2} = \frac{\sum_{s \in S_{xy}} r_{x,s} r_{y,s}}{\sqrt{\sum_{s \in S_{xy}} r_{x,s}^2} \sqrt{\sum_{s \in S_{xy}} r_{y,s}^2}},$$

Όπου τα διανύσματα $x \cdot y$ δηλώνουν το γινόμενο των διανυσμάτων x, y .

Επιπρόσθετα μια άλλη προσέγγιση που υπολογίζει την ομοιότητα μεταξύ των χρηστών χρησιμοποιεί το μέσο τετραγωνικό μέτρο διαφοράς. *Επιστρέφει τιμές ανάμεσα στο -1 και το 1 με το 1 να δηλώνει την απόλυτη ταύτιση, το -1 την απόλυτη απόκλιση των διανυσμάτων.* Όταν τα διανύσματα είναι κάθετα μεταξύ τους η επιστρεφόμενη τιμή είναι 0 και συνήθως αυτό λαμβάνεται ως ανεξαρτησία.

Επίσης όταν οι προτιμήσεις των χρηστών δηλώνονται μόνο με θετικές τιμές τότε το σύνολο των επιστρεφόμενων τιμών περιορίζεται στο $[0,1]$.

Κάθε σύστημα συστάσεων διαφέρει στην προσέγγιση του μέτρου της ομοιότητας και της αξιολόγησης. Κοινή τακτική όλων είναι να υπολογίζουν την ομοιότητα όλων των χρηστών αρχικά και να την επαναυπολογίζουν εκ νέου όταν αυτό απαιτείται. Έτσι όταν ένας χρήστης ζητάει σύσταση, οι αξιολογήσεις μπορούν να υπολογιστούν αποτελεσματικά χρησιμοποιώντας προϋπολογισμένες ομοιότητες.

- Προσαρμοσμένη Ομοιότητα Συνημιτόνου (Adjusted Cosine Similarity):

Η προσαρμοσμένη ομοιότητα συνημιτόνου είναι μια άλλη οπτική της απόστασης συνημιτόνου η οποία υπολογίζει τον μέσο όρο των τιμών που υπάρχουν σε κάθε διάσταση. Σκοπός της μετρικής είναι να αντιμετωπίσει την διαφορετική αντίληψη που έχουν οι χρήστες για την κλίμακα των βαθμολογιών όπου αποτελεί ένα από τα μειονεκτήματα των τεχνικών του συνεργατικού φιλτραρίσματος.

$$\text{similarity}(A, B) = \frac{\sum_{i=1}^N (A_i - \mu_o(i)) * (B_i - \mu_o(i))}{\sqrt{\sum_{i=1}^N (A_i - \mu_o(i))^2} * \sqrt{\sum_{i=1}^N (B_i - \mu_o(i))^2}}$$

Όπου $\mu_o(i)$ είναι ο μέσος όρος των τιμών της i -οστής διάστασης των διανυσμάτων. Αν συγκρίνουμε χρήστες θα είναι ο μέσος όρος των τιμών των προτιμήσεων που έχουν δείξει όλοι οι χρήστες στο αντικείμενο i ενώ αν συγκρίνουμε αντικείμενα θα είναι ο μέσος των τιμών από τις προτιμήσεις που έχει εκφράσει ο i χρήστης.

Αρκετές τροποποιήσεις-τεχνικές που βασίζονται στο συνημίτονο έχουν προταθεί στα παραπάνω ως επεκτάσεις προς βελτιστοποίηση όπως: προεπιλεγμένο ψήφου (default voting), αντίστροφη συχνότητα χρήσης (inverse user frequency), περίπτωση ενίσχυσης (case amplification) και σταθμισμένη πλειοψηφία πρόβλεψης (weighted-majority prediction) (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005).

- Ομοιότητα Jaccard:

Η μετρική αυτή χρησιμοποιείται για να υπολογιστεί το ποσοστό που δείχνει τις αλληλεπικαλύψεις των συνόλων και όχι τις αποστάσεις τους. Για την σωστή χρήση της πρέπει να χρησιμοποιηθούν τα αντικείμενα για τα οποία οι χρήστες έχουν εκφράσει ενδιαφέρον ανεξάρτητα αν τους ενδιαφέρει πολύ ή λίγο. Ο τύπος είναι:

$$\textit{similarity}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Πανεπιστήμιο Πειραιώς

Ακόμα ο τύπος της απόστασης Jaccard ο οποίος χρησιμοποιείται για τον υπολογισμό της διαφοράς δύο συνόλων είναι:

$$\text{dissimilarity}(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

Οι μετρικές που αναλύθηκαν παραπάνω δεν χρησιμοποιούνται σε όλες τις εφαρμογές με την ίδια απόδοση. Πειραματικές μέθοδοι δείχνουν πως ο συντελεστής συσχέτισης του Pearson χρησιμοποιείται καλύτερα για τον υπολογισμό ομοιότητας χρηστών η προσαρμοσμένη ομοιότητα συνημίτονου χρησιμοποιείται για τον υπολογισμό ομοιότητας αντικειμένων (Meyer & Boulle, 2007) (Sarwar B. , Karypis, Konstan, & Riedl, 2001).

2.5.3 Συστήματα Προσανατολισμένα Στο Μοντέλο (Model Based systems)

Εν αντιθέσει με τους αλγόριθμους που βασίζονται στη μνήμη, οι αλγόριθμοι βασιζόμενοι στο μοντέλο χρησιμοποιούν τις αξιολογήσεις προκειμένου να «διαβάσουν» το μοντέλο, το οποίο στη συνέχεια θα χρησιμεύσει ώστε να προτείνει συστάσεις.

Οι αλγόριθμοι που λειτουργούν βάσει μοντέλου χρησιμοποιούν τα δεδομένα των προτιμήσεων που είναι αποθηκευμένα στο σύστημα, ως σύνολο εκπαίδευσης αλγόριθμων μηχανικής μάθησης ώστε να παράγουν μοντέλα πρόβλεψης βαθμολογιών.

Η γενική ιδέα είναι να μοντελοποιηθούν οι αλληλεπιδράσεις χρηστών-προϊόντων με παράγοντες που εκπροσωπούν τα ελλειπή χαρακτηριστικά των χρηστών και των προϊόντων στο σύστημα, όπως η κατηγορία προτιμήσεων του χρήστη ή η κλάση κατηγορίας στην οποία ανήκει το προϊόν.

Η επεξεργασία γίνεται σε μη πραγματικό χρόνο και έτσι μειώνεται το πρόβλημα κλιμάκωσης. Επίσης μπορεί να γίνει επεξεργασία μεγάλου αριθμού από προφίλ καθώς δεν απαιτείται απάντηση σε πραγματικό χρόνο (Μουζακίδης).

Το μοντέλο αυτό εκπαιδεύεται χρησιμοποιώντας τα διαθέσιμα δεδομένα και στη συνέχεια εφαρμόζεται για να προβλέψει τις βαθμολογήσεις των χρηστών σε καινούρια προϊόντα.

Για παράδειγμα, η ομαδοποίηση με βάση τη μέθοδο συνεργατικού φιλτραρίσματος χτίζει ένα μοντέλο συνόλου δεδομένων ως συστάδες χρηστών και στη συνέχεια χρησιμοποιεί τις αξιολογήσεις των χρηστών μέσα σε μια συστάδα για να δημιουργήσει την πρόβλεψη (Sachan). Μια πολύ επιτυχημένη μέθοδος βάσει μοντέλου είναι η Singular Value Decomposition (SVD), την οποία προαναφέραμε σε προηγούμενη ενότητα, όπου αντιπροσωπεύει τα δεδομένα με ένα σύνολο διανυσμάτων, ένα για κάθε στοιχείο και ένα για κάθε χρήστη, έτσι ώστε το εσωτερικό γινόμενο του φορέα χρήστη και του φορέα ταινία να αποτελεί τη καλύτερη προσέγγιση για την εκπαίδευση του αλγορίθμου (Breese, Heckerman, & Kadie, Empirical Analysis Of Predictive Algorithms For Collaborative Filtering, 1998).

Τυπικά η διαδικασία οικοδόμησης του μοντέλου είναι υπολογιστικά δαπανηρή και απαιτεί πολλή μνήμη. Ύστερα, αφού κατασκευαστούν τα μοντέλα, οι προβλέψεις υλοποιούνται πολύ γρήγορα με μικρή απαίτηση μνήμης.

Τα μοντέλα με βάση τις μεθόδους συνεργατικού φιλτραρίσματος συνήθως επιτυγχάνουν λιγότερο ακριβή πρόβλεψη από ότι τα μοντέλα με βάση τη μνήμη, που βασίζονται σε πυκνά σύνολα δεδομένων, όπου ένα μεγάλο ποσοστό τιμών διατίθεται ως σύνολο εκπαίδευσης, αλλά αποδίδουν καλύτερα σε αραιά σύνολα δεδομένων (Sachan).

Με την πάροδο του χρόνου ερευνητές έχουν προτείνει διάφορες προσεγγίσεις για την σύσταση αντικειμένων όπως την Bayesian Clustering, την Latent Semantic Analysis(LSI), την Maximum Entropy,

τις Boltzmann Machines, τις Support Vector Machines (SVM) και τη Singular Value Decomposition(SVD).

Βασικά οι αλγόριθμοι βάσει μοντέλου όπως Bayesian model, clustering models, και dependency networks-δίκτυα εξάρτησης ανακαλύφθηκαν προκειμένου να λύσουν τις ελλείψεις των τεχνικών προσέγγισης βάσει μνήμης (Xiaoqian & Taghi, 2009). Συνήθως, οι αλγόριθμοι ταξινόμησης μπορούν να χρησιμοποιηθούν ως μοντέλα (CF) φιλτραρίσματος εάν οι αξιολογήσεις των χρηστών είναι κατηγοριοποιημένες, ενώ μοντέλα παλινδρόμησης και μέθοδοι SVD χρησιμοποιούνται για αριθμητικές βαθμολογίες.

2.5.4 Βελτίωση Των Αποτελεσμάτων Των Συστημάτων Βασισμένα Στην Συνεργασία

Υπάρχουν πολλές μέθοδοι που μπορούν να χρησιμοποιηθούν για να βελτιωθούν τα αποτελέσματα που προκύπτουν από τα συστήματα βασισμένα στην συνεργασία. Οι μέθοδοι αυτοί είναι:

- Default Voting:

Η μέθοδος default voting είναι μια μέθοδος που χρησιμοποιείται για την βελτίωση των αποτελεσμάτων όταν υπάρχουν λίγες αξιολογήσεις. Με την χρήση αυτής της μεθόδου εισάγονται αξιολογήσεις οι οποίες είναι προκαθορισμένες στα αντικείμενα όπου δεν υπάρχει κάποια αξιολόγηση. Με αυτόν τον τρόπο το σύνολο των αξιολογημένων αποτελεσμάτων μεγαλώνει και κατ' επέκταση βελτιώνονται και τα αποτελέσματα σε ορισμένες περιπτώσεις.

- Inverse User Frequency:

Η μεθοδολογία inverse user frequency ισχυρίζεται πως οι βαθμολογίες των αντικειμένων οι οποίες είναι υψηλές από μια μεγάλη μερίδα χρηστών δεν βοηθάνε στον υπολογισμό της συσχέτισης μεταξύ των χρηστών γιατί δεν δίνουν αρκετές πληροφορίες. Έτσι χρησιμοποιείται ο συντελεστής συσχέτισης:

$$f_j = \log \frac{n}{n_j}$$

Όπου το n_j συμβολίζει τον αριθμό των χρηστών που ψήφισαν το j και n είναι ο αριθμός των χρηστών που υπάρχουν.

- Case Amolification (Breese, Heckerman, & Kadie, Empirical Analysis Of Predictive Algorithms For Collaborative Filtering, 1998):

Η μέθοδος αυτή αλλάζει τον αριθμό από τα βάρη για να βελτιωθούν οι προβλέψεις που γίνονται. Για να γίνει καλύτερη η πρόβλεψη τα βάρη που είναι κοντά στην μονάδα ενισχύονται και μειώνονται αυτά που είναι κοντά στο μηδέν.

2.5.5 Προβλήματα Των Συστημάτων Βασισμένα Στην Συνεργασία

Τα συστήματα βασισμένα σε συνεργασία έχουν κάποιους περιορισμούς στην λειτουργία τους οι οποίοι είναι ίδιοι με τους περιορισμούς που έχουν και τα συστήματα βασισμένα στο περιεχόμενο αλλά έχουν και κάποιους ακόμα. Οι πιο σημαντικοί περιορισμοί είναι:

- Το πρόβλημα των αραιών δεδομένων (Sparsity Problem):

Ένα πρόβλημα το οποίο κάνει συχνά την εμφάνισή του είναι το γεγονός πως εμφανίζεται ένας μεγάλος αριθμός εναλλακτικών αποτελεσμάτων, το πρόβλημα αυτό ονομάζεται ως πρόβλημα χαμηλής κάλυψης. Το ίδιο πρόβλημα έχουν ακόμα και οι χρήστες που έχουν αρκετές αξιολογήσεις. Για παράδειγμα αν ένας χρήστης με πολλές αξιολογήσεις κάνει μια αναζήτηση σε μια βιβλιοθήκη με μεγάλο αριθμό βιβλίων τα αποτελέσματα που θα πάρει θα είναι πάρα πολλά για να μπορέσει να τα ψάξει όλα. Αυτό συμβαίνει γιατί τα συστήματα βασισμένα στην συνεργασία δεν μπορούν να εντοπίσουν γειτονικούς χρήστες και έτσι δεν μπορούν να δώσουν ορθά αποτελέσματα.

Για να λυθεί αυτό το πρόβλημα προτάθηκε μια λύση η οποία αναφέρει πως η ύπαρξη περισσότερων στοιχείων για τους χρήστες όπως το φύλο, η ηλικία, τα ενδιαφέροντα και άλλα μπορούν να δώσουν την δυνατότητα στα συστήματα να κάνουν καλύτερους συσχετισμούς στα δεδομένα τους και να εμφανίζουν πιο αξιόπιστα δεδομένα (Pazzani, A Framework For Collaborative, Content-Based And Demographic Filtering, 1999).

Υπάρχουν και άλλες προτεινόμενες λύσεις οι οποίες προϋποθέτουν την ύπαρξη ενός αλγορίθμου. Μια τέτοια λύση είναι και η χρήση του αλγορίθμου SVD (Singular Value Decomposition) ο οποίος χρησιμοποιείται για να εντοπίσει σχέσεις μεταξύ των χρηστών και των εναλλακτικών με σκοπό να χρησιμοποιηθούν για τον υπολογισμό των εναλλακτικών για τους χρήστες. Αφού γίνει αυτό τότε ο αλγόριθμος περιορίζει τα αποτελέσματα σε έναν αριθμό όπου να μπορούν να γίνουν οι κατάλληλες συστάσεις στους χρήστες (Sarwar, Karypis, & Riedl, Analysis Of Recommendation Algorithms For E-Commerce, 2000).

- Το πρόβλημα της Ψυχρής Εκκίνησης (Cold Start Problem):

Το πρόβλημα της ψυχρής εκκίνησης είναι όμοιο με αυτό με το πρόβλημα των αραιών δεδομένων. Το πρόβλημα αυτό εμφανίζεται όταν εμφανίζονται νέοι χρήστες ή νέες εναλλακτικές (Schein, Popescul, Ungar, & Pennock, 2002).

- Το πρόβλημα Του Νέου Χρήστη (New User Problem):

Κατά την εγγραφή και την είσοδο ενός νέου χρήστη στο σύστημα πρέπει αυτός να δημιουργήσει ένα προφίλ το οποίο θα περιέχει στοιχεία γι' αυτόν που θα βοηθούν το σύστημα να του κάνει αξιόπιστες προτάσεις. Όμως ακόμα και αν ο χρήστης δώσει στο σύστημα ότι στοιχεία χρειάζεται τα αποτελέσματα και πάλι δεν θα είναι αξιόπιστα μέχρι ο χρήστης να κάνει έναν αριθμό αξιολογήσεων ώστε να μπορούν να γίνουν οι κατάλληλες συγκρίσεις. Το πρόβλημα αυτό ονομάζεται το πρόβλημα του νέου χρήστη.

Για την αντιμετώπιση αυτού του προβλήματος έχουν αναφερθεί πολλές διαφορετικές λύσεις όπως η χρήση των δημογραφικών προβλέψεων προτιμήσεων και άλλες οι οποίες χρησιμοποιούν έναν συνδυασμό από τεχνικές συστημάτων περιεχομένου και συστημάτων βασισμένα στην συνεργασία.

Ακόμα υπάρχουν και άλλες τεχνικές οι οποίες δεν βοηθούν στο να γίνουν καλύτεροι οι υπολογισμοί των εναλλακτικών αλλά βοηθούν στην συμπλήρωση του προφίλ του χρήστη με πιο γρήγορο τρόπο. Οι τεχνικές αυτές χρησιμοποιούν διάφορες μεθόδους για να πετύχουν τον σκοπό τους όπως την χρήση των εναλλακτικών που εμφανίζουν μεγάλη δημοτικότητα και άλλες (Rashid, Albrt, Cosley, & Lam).

- Το Πρόβλημα του Νέου Αντικειμένου (New Item Problem)

Στα συστήματα όπου υπάρχουν προτάσεις το φαινόμενο να υπάρχουν νέα αντικείμενα όπως νέα βιβλία, νέα τραγούδια και άλλα είναι πολύ συχνό. Όμως βάση των συστημάτων βασισμένα στην συνεργασία για να εμφανιστεί μια εναλλακτική ως αποτέλεσμα της αναζήτησης ενός χρήστη θα πρέπει να υπάρχει τουλάχιστον μια αξιολόγηση γι' αυτήν την εναλλακτική. Αυτό έχει σαν αποτέλεσμα οι νέες εναλλακτικές να μην εμφανίζονται ως αποτελέσματα στο σύστημα μέχρι να αξιολογηθούν (Breese, Heckerman, & Kadie, Empirical Analysis Of Predictive Algorithms For CollaBorative Filtering , 1998)

Για να αντιμετωπιστεί αυτό το πρόβλημα θα πρέπει να χρησιμοποιηθεί ένας συνδυασμός από τα συστήματα με βάση το περιεχόμενο και τα συστήματα με βάση την συνεργασία.

- Το Πρόβλημα Της Κλίμακας (Scalability Problem)

Εκτός από τα άλλα προβλήματα που εμφανίζονται κατά την λειτουργία του συστήματος βασισμένο στην συνεργασία είναι η έλλειψη υπολογιστικής ισχύος. Τα περισσότερα συστήματα βασισμένα στην

συνεργασία διαχειρίζονται εκατομμύρια αντικείμενα και χρήστες και κάθε φορά που κάποιος χρήστης κάνει μια αναζήτηση τότε το σύστημα θα πρέπει να ψάξει όλους τους χρήστες για να μπορέσει να εμφανίσει τις κατάλληλες αξιολογήσεις. (Resnick & Varian, 1997). Για να γίνουν οι κατάλληλες συστάσεις χρησιμοποιούνται κάποιες μέθοδοι για την αναζήτηση των εναλλακτικών και την εμφάνισή τους στους χρήστες. Οι μέθοδοι αυτοί χρησιμοποιούν πολλούς υπολογιστικούς πόρους για να εμφανίσουν τα αποτελέσματά τους και αν ο όγκος των δεδομένων που έχουν να επεξεργαστούν αυτές οι μέθοδοι είναι αρκετά μεγάλος τότε οι χρόνοι όπου τα συστήματα βγάζουν τα αποτελέσματά τους αυξάνονται αισθητά. Δημιουργείται έτσι το πρόβλημα της ύπαρξης αποτελεσμάτων σε χρόνους αποδεκτούς για τα δεδομένα του διαδικτύου. Το πρόβλημα αυτό δεν μπορεί να αντιμετωπιστεί ποτέ εξ' ολοκλήρου καθώς κάθε φορά που θα ανεβαίνει ο αριθμός των χρηστών και των αντικειμένων αρκετά θα εμφανίζεται ξανά. Για την αντιμετώπιση αυτών των προβλημάτων απαιτούν την ύπαρξη σύγχρονων δομών δεδομένων και αλγορίθμων που λειτουργούν σε μεγάλες εφαρμογές.

Ένας τρόπος για την αντιμετώπιση αυτού του προβλήματος είναι η χρησιμοποίηση της μεθόδου SVD ή κάποιας άλλης παρόμοιας ή οποια θα μειώνει αρκετά τον αριθμό των δεδομένων που έχει να επεξεργαστεί το σύστημα και έτσι να μειώνει τον χρόνο εμφάνισης των αποτελεσμάτων (Sarwar, Karypis, & Riedl, Analysis Of Recommendation Algorithms For E-Commerce, 2000).

Αυτή η μέθοδος όμως στοιχίζει αρκετά στην ορθότητα και στην ακρίβεια των αποτελεσμάτων γιατί με την μεγάλη μείωση του όγκου των δεδομένων υπάρχει μεγάλη περίπτωση να χαθούν πολύ χρήσιμες πληροφορίες που πιθανόν να ενδιέφεραν τον χρήστη.

- Το Πρόβλημα των Συνωνύμων (Synonymy)

Τα συστήματα βασισμένα στην συνεργασία αδυνατούν να κάνουν συσχετισμούς μεταξύ των λέξεων κλειδιά που χρησιμοποιεί ο χρήστης και όμοιων λέξεων που χρησιμοποιούνται για το ίδιο θέμα, έτσι δεν μπορούν να εξεταστούν όλες οι πιθανές εναλλακτικές που μπορεί να ενδιαφέρουν τον χρήστη και τα αποτελέσματα είναι ελλιπή.

Ένα βασικό μειονέκτημα των συστημάτων βασισμένα στην συνεργασία είναι πως απαιτούν ένα βασικό αριθμό αξιολογήσεων για να μπορέσει να βγάλει ένα ασφαλές αποτέλεσμα. Έτσι στην αρχή της λειτουργίας του συστήματος και μέχρι να μαζευτούν οι αξιολογήσεις που είναι απαραίτητες, οι συστάσεις που θα κάνει δεν θα είναι αξιόπιστες (Jennings & Higuchi, 1993).

- Το πρόβλημα των πολλών αντικειμένων

Επίσης ένα πρόβλημα που μπορεί να δημιουργηθεί είναι το πρόβλημα που αναφέρεται ως sparsity. Το sparsity αναφέρεται ως το πρόβλημα που δημιουργείται όταν τα συστήματα που βασίζονται στην συνεργασία έχουν πάρα πολλά αντικείμενα για αξιολόγηση σε σύγκριση με τον αριθμό των χρηστών που χρησιμοποιούν το συγκεκριμένο σύστημα. Έτσι είναι πολύ δύσκολο να βρεθούν συσχετίσεις μεταξύ των αντικειμένων και των χρηστών που τα αξιολόγησαν (ERCIM, 1997).

- Το πρόβλημα των χρηστών με μη αναμενόμενες προτιμήσεις

Τα συστήματα συστάσεων βασισμένα στην συνεργασία λειτουργούν βάση των αξιολογήσεων των χρηστών. Έτσι όσο πιο συνηθισμένες είναι οι απαιτήσεις ενός χρήστη τόσο πιο αξιόπιστα θεωρούνται τα αποτελέσματα. Αντιθέτως αν οι απαιτήσεις ενός χρήστη είναι ιδιαίτερες τα συστήματα θα μπορέσουν να συσχετίσουν τις αξιολογήσεις λίγων ατόμων και έτσι τα αποτελέσματα δεν θα είναι το ίδιο αξιόπιστα (Breese, Heckerman, & Kadie, Empirical Analysis Of Predictive Algorithms For Collaborative Filtering, 1998).

- Το στατιστικό πρόβλημα

Ακόμα είναι γνωστό πως τα αποτελέσματα των συστημάτων βασισμένα στην συνεργασία είναι αποτελέσματα που προκύπτουν από στατιστικές μεθόδους. Αυτό σημαίνει πως οι διαχειριστές και οι δημιουργοί των συστημάτων γνωρίζουν πως μπορούν να υπάρξουν αρκετές στατιστικές ανωμαλίες που θα επηρεάζουν τα αποτελέσματα του συστήματος.

- Το πρόβλημα της αλλαγής των προτιμήσεων των χρηστών

Τέλος ένα πρόβλημα που εμφανίζεται με τις αξιολογήσεις των συστημάτων βασισμένα στην συνεργασία είναι ότι όταν οι χρήστες είναι στο σύστημα πολλά χρόνια είναι λογικό να αλλάζουν οι προτιμήσεις τους με αποτέλεσμα κάποια αντικείμενα που τους άρεσαν στο παρελθόν να μην τους άρεσουν στο παρόν ή και το αντίθετο. Αυτό επηρεάζει γενικά τον τρόπο λειτουργίας των συστημάτων καθώς τα αποτελέσματα των αναζητήσεων που στηρίζονται σε αυτούς τους χρήστες δεν θα μπορούν να είναι αξιόπιστα. Υπάρχει όμως και δυσκολία στην αντιμετώπιση αυτού του γεγονότος καθώς η αλλαγή των προτιμήσεων των ανθρώπων καθώς αυτοί μεγαλώνουν είναι φυσιολογική και ταυτόχρονα πολύ δύσκολο στο να προβλεφτεί.

2.6 Χρήση Δημογραφικών Δεδομένων (Demographic-Based)

Ο όρος δημογραφικά δεδομένα αναφέρεται στα στοιχεία που υπάρχουν στο προφίλ κάθε χρήστη και έχουν να κάνουν με στοιχεία όπως την ηλικία του, το φύλο του, τα ενδιαφέροντά του και άλλα τα οποία βοηθούν να γίνει καλύτερη συσχέτιση των εναλλακτικών με τον χρήστη. Τα δημογραφικά δεδομένα χρησιμοποιούνται με την ιδέα πως άτομα που έχουν τα ίδια ενδιαφέροντα, την ίδια ηλικία και τις ίδιες ασχολίες είναι πολύ πιθανόν να αναζητούν το ίδιο αντικείμενο.

Τα συστήματα που χρησιμοποιούν αλγόριθμους που βασίζονται στα δημογραφικά δεδομένα εκτιμούν πως μια εναλλακτική για έναν χρήστη θα του είναι τόσο ενδιαφέρον όσο ενδιαφέρον είναι η ίδια εναλλακτική σε χρήστες με παρόμοια δημογραφικά δεδομένα.

2.6.1 Περιορισμοί Των Δημογραφικών Προβλέψεων

Το μεγαλύτερο πρόβλημα που υπάρχει με αυτήν την μέθοδο εύρεσης εναλλακτικών προτάσεων είναι πως πολλοί χρήστες ιδιαίτερα αυτοί που είναι μεγάλης ηλικίας δεν θέλουν να εισάγουν τα προσωπικά τους στοιχεία μέσα στο διαδίκτυο και έτσι τα στοιχεία που χρησιμοποιούν οι αλγόριθμοι είναι ελλιπή και δεν μπορούν να βγάλουν αξιόπιστα αποτελέσματα. Ένας ακόμα λόγος που δεν υπάρχουν αρκετά δημογραφικά στοιχεία είναι πως οι χρήστες δεν θέλουν να συμπληρώνουν μεγάλα ερωτηματολόγια σχετικά με αυτούς θεωρώντας πως είναι άσκοπα και πως δεν θα ωφεληθεί σε τίποτα αν το συμπληρώσουν.

Για να αντιμετωπιστεί αυτό το πρόβλημα έχουν δημιουργηθεί αλγόριθμοι στους οποίους ο χρήστης δίνει μια προσωπική σελίδα και από εκεί μπορούν να χρησιμοποιήσουν πολλά στοιχεία από την γλώσσα που χρησιμοποιεί, τις εκφράσεις του, τα θέματα με τα οποία ασχολείται και άλλα.

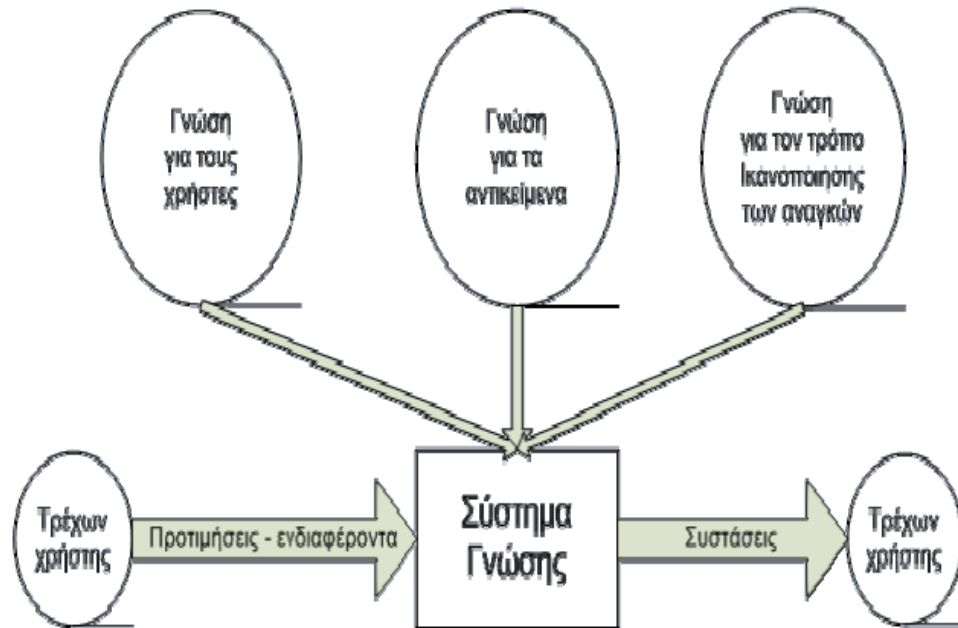
2.7 Συστήματα Βασισμένα Στην γνώση (Knowledge- Based Systems)

Η κατηγορία αυτή των συστημάτων είναι τα συστήματα τα οποία βασίζονται πάνω στην γνώση τους για κάποια χαρακτηριστικά είτε του αντικειμένου είτε του χρήστη και βασίζεται πάνω σε αυτά για να κάνει στην κατάλληλη σύσταση. Η γνώση που μπορεί να διαθέτουν τα συστήματα αυτά χωρίζεται σε τρεις κατηγορίες και ανάλογα με το ποια διαθέτει το κάθε σύστημα κάνει τις ανάλογες ενέργειες. Οι τρεις κατηγορίες είναι (Burke, Knowledge-based recommender systems, 2000):

- **Γνώση για τα αντικείμενα:** Το σύστημα πρέπει να γνωρίζει λεπτομέρειες για το κάθε αντικείμενο που περιέχει για να μπορέσει να κάνει τις απαραίτητες συγκρίσεις και να κάνει τις απαραίτητες συστάσεις ενός προϊόντος αλλά και όλα τα άλλα προϊόντα που ανήκουν στην ίδια κατηγορία.
- **Γνώση για τους χρήστες:** Το σύστημα ακόμα πρέπει να γνωρίζει λεπτομέρειες σχετικά με τον κάθε χρήστη που είναι μέλος του για να καταφέρει να δημιουργήσει ένα ικανοποιητικό προφίλ που θα το βοηθήσει κατά την διάρκεια της δημιουργίας συστάσεων.
- **Γνώση του τρόπου με τον οποίο καλύπτονται οι ανάγκες:** Το σύστημα πρέπει να γνωρίζει ποια προϊόντα μπορούν να καλύψουν τις ανάγκες των χρηστών και να κάνει την αντιστοίχιση με τα διαθέσιμα προϊόντα που υπάρχουν.

Η διαδικασία με την οποία τα συστήματα αποκτούν την γνώση διαφέρει ανάλογα με την κατηγορία. Για την γνώση σχετικά με τα αντικείμενα το σύστημα πρέπει να αντλήσει τις πληροφορίες του μέσω τεχνικών άντλησης γνώσης που γίνονται μέσα στις βάσεις δεδομένων που περιέχει το σύστημα. Στην συνέχεια το σύστημα αποφασίζει αν θα χρησιμοποιήσει τις πληροφορίες που έχει αποκτήσει ή όχι. Αυτός ο τρόπος απόκτησης γνώσης είναι ο έμμεσος τρόπος. Παράλληλα υπάρχει και ο άμεσος τρόπος στον οποίο οι διαχειριστές του συστήματος εισάγουν τα δεδομένα που έχουν στην διάθεσή τους από τους παραγωγούς ή τους πωλητές των προϊόντων. Ο άμεσος τρόπος χρησιμοποιείται συνήθως σε συστήματα όπου περιέχουν προϊόντα προς πώληση. Η παρακάτω εικόνα περιγράφει τον τρόπο με τον οποίο γίνονται οι συστάσεις στα συστήματα βασισμένα στην γνώση.

Εικόνα 6: Περιγραφή Του Τρόπου Λειτουργίας Των Συστημάτων Βασισμένα Στην Γνώση



Τα συστήματα βασισμένα στην γνώση χωρίζονται σε δύο κατηγορίες τα συστήματα βασισμένα στην υπόθεση και τα συστήματα βασισμένα στον στόχο.

2.7.1 Συστήματα Βασισμένα Στην Υπόθεση (Case- Based System)

Τα συστήματα αυτής της κατηγορίας κάνουν την αντιστοίχιση των προϊόντων και των αναγκών του χρήστη με βάση τις προτιμήσεις και τις ανάγκες που έβαλε ο χρήστης κατά την δημιουργία του προφίλ του. Τα συστήματα βασισμένα στην υπόθεση ακολουθούν τα παρακάτω βήματα για την δημιουργία των συστάσεων (Burke, The Wasabi Personal Shopper: A Case-Based Recommender System, 1999):

- Κατά την εγγραφή του χρήστη και την δημιουργία του προφίλ του το σύστημα τον ρωτάει κατά πόσο τον ενδιαφέρουν ορισμένες κατηγορίες προϊόντων. Οι προτιμήσεις αυτές καταχωρούνται στο σύστημα και χρησιμοποιούνται στο μέλλον για να γίνουν οι συστάσεις.
- Στην συνέχεια το σύστημα ελέγχει τα δεδομένα που έχει στην διάθεση του από τα προϊόντα και συγκρίνει τα χαρακτηριστικά τους με τις απαιτήσεις του χρήστη και κάνει τις συστάσεις που πιστεύει πως θα ικανοποιήσουν τον χρήστη. Στην συνέχεια ο χρήστης αναφέρει αν είναι ικανοποιημένος με τις συστάσεις που του έγιναν ή όχι. Αν οι συστάσεις ενδιαφέρουν τον χρήστη τότε η διαδικασία σταματάει αλλιώς συνεχίζει στο τρίτο βήμα.
- Στο βήμα αυτό το σύστημα επανεξετάζει ξανά τα δεδομένα που έχει κάνοντας κάποιες αλλαγές στις προτιμήσεις του χρήστη και τέλος εμφανίζει νέες προτάσεις προς τον χρήστη.

2.7.2 Συστήματα Βασισμένα Στον Στόχο (Goal- Based Systems)

Η κατηγορία αυτή ονομάζεται συστήματα βασισμένα στο στόχο και χρησιμοποιούν μέτρα ομοιότητας για να μπορέσουν να κάνουν τις συστάσεις που επιθυμούν. Η λειτουργία των συστημάτων αυτών περιγράφεται στα παρακάτω βήματα (Burke, Hammond, & Young, The FindMe Approach to Assisted Browsing, 1997):

- Στο στάδιο αυτό το σύστημα χρησιμοποιεί τα προϊόντα που είδε ή που αγόρασε ο χρήστης. Γνωρίζοντας τα προϊόντα αυτά το σύστημα μπορεί να καταλάβει τις προτιμήσεις και τις ανάγκες του χρήστη και έτσι μπορεί πιο εύκολα να κάνει πιο εύστοχες συστάσεις.

- Στο επόμενο στάδιο το σύστημα αναζητά τα προϊόντα που εξυπηρετούν τον ίδιο σκοπό με αυτά τα προϊόντα που έχει δείξει ενδιαφέρον ο χρήστης στο παρελθόν. Στο στάδιο αυτό χρησιμοποιούνται τα κριτήρια της ομοιότητας για να μπορέσει να ξεχωρίσει το σύστημα αν δύο αντικείμενα που είναι όμοια μεταξύ τους εξυπηρετούν τους ίδιους σκοπούς ή όχι.
- Στο τελευταίο στάδιο ο χρήστης αξιολογεί τις συστάσεις που του έγιναν ανάλογα με το αν τον ενδιαφέρουν ή όχι. Η αξιολόγηση μπορεί να γίνει είτε άμεσα είτε έμμεσα. Άμεσα γίνεται όταν ο χρήστης βαθμολογεί τις συστάσεις που του προτάθηκαν ενώ έμμεσα όταν το σύστημα παρακολουθεί τις κινήσεις του χρήστη για να διαπιστώσει εάν έδειξε ενδιαφέρον στις προτάσεις που του έγιναν ή όχι. Αν ο χρήστης δεν μείνει ικανοποιημένος τότε του γίνονται νέες προτάσεις με διαφορετικές προδιαγραφές και χαρακτηριστικά.

Ένα παράδειγμα των συστημάτων συστάσεων με βάση την γνώση είναι το σύστημα Entrée από το Chicago το οποίο δίνει πληροφορίες στους χρήστες για παροχές και υπηρεσίες οι οποίες παρέχονται από εστιατόρια. Είναι χαρακτηριστικό παράδειγμα γιατί αποτελείται από μια πολύ καλά οργανωμένη βάση δεδομένων που λειτουργεί με πολύ αποτελεσματικό τρόπο. Η παρακάτω εικόνα δείχνει τα στοιχεία που πρέπει να εισάγει ο χρήστης στην εφαρμογή Entrée για να πάρει τα αποτελέσματα που επιθυμεί (Burke, Knowledge-based recommender systems, 2000).

Εικόνα 7: Το Σύστημα Entrée Από Το Chicago



I would like to eat at a restaurant that has:

I would like to eat at a restaurant just like:

2.7.3 Πλεονεκτήματα Των Συστημάτων Βασισμένα Στην Γνώση

Τα συστήματα βασισμένα στην γνώση δεν αντιμετωπίζουν το πρόβλημα της ψυχρής εκκίνησης που αντιμετωπίζουν τα συστήματα βασισμένα στην συνεργασία. Αυτό συμβαίνει γιατί τα αποτελέσματα που βγάζουν τα συστήματα που βασίζονται στην γνώση δεν εξαρτώνται από τις αξιολογήσεις των άλλων χρηστών το οποίο απαιτεί πολύ χρόνο και αποτελεί ένα από τα σημαντικότερα προβλήματα των συστημάτων βασισμένα στην γνώση. Έτσι τα συστήματα βασισμένα στην γνώση μπορούν να δώσουν αξιόπιστες προτάσεις και να ανταποκρίνονται σωστά στις αλλαγές των προτιμήσεων των χρηστών.

Ακόμα τα συστήματα που βασίζονται στην γνώση έχουν την ικανότητα να προσαρμόζονται γρήγορα στις αλλαγές των προτιμήσεων των χρηστών γιατί αντλούν πληροφορίες για τους χρήστες συνεχώς είτε με έμμεσο είτε με άμεσο τρόπο.

Επιπλέον σε αντίθεση με τα συστήματα βασισμένα στην συνεργασία δεν απαιτούν μεγάλη επεξεργαστική ισχύ για να υπολογίσουν τους κοντινούς χρήστες με βάση τα ενδιαφέροντα κάθε φορά που υπάρχουν νέες αξιολογήσεις στο σύστημα γιατί υπολογίζει και κάνει τις συστάσεις σε πραγματικό χρόνο.

Επίσης τα συστήματα βασισμένα στην γνώση δεν επηρεάζονται από τις στατιστικές ανωμαλίες που πιθανόν να υπάρχουν στο σύστημα. Ακόμα μπορούν να κάνουν αξιόπιστες προτάσεις σε χρήστες όπου οι προτιμήσεις τους διαφέρουν πολύ σε σχέση με τον μεγαλύτερο πληθυσμό σε αντίθεση με τα συστήματα βασισμένα στην συνεργασία που αντιμετώπιζαν μεγάλο πρόβλημα με το συγκεκριμένο κομμάτι των χρηστών. Αυτό συμβαίνει γιατί τα συστήματα βασισμένα στην γνώση δεν εξαρτώνται από τις προτιμήσεις των υπόλοιπων χρηστών αλλά μόνο από τις πληροφορίες που εισάγουν οι ίδιοι στο σύστημα. Με αυτό τον τρόπο μπορούν να τους γίνουν αξιόπιστες συστάσεις.

Ένα σημαντικό τμήμα των αντικειμένων που δεν επηρεάζονται τα συστήματα βασισμένα στην γνώση είναι ότι έχουν την δυνατότητα να προτείνουν προϊόντα τα οποία είναι καινούργια στο σύστημα και δεν έχουν αξιολογηθεί ακόμα. Αυτό συμβαίνει γιατί το σύστημα γνωρίζει τα χαρακτηριστικά του αντικειμένου. Σε περίπτωση που κρίνει το ίδιο το σύστημα πως ταιριάζουν με τις απαιτήσεις του χρήστη τότε δημιουργεί την πρόταση, εν αντίθεση με τα συστήματα βασισμένα στην συνεργασία που προτείνουν μόνο ότι έχει αξιολογηθεί ως ενδιαφέρον από άτομα που έχουν τα ίδια ενδιαφέροντα.

Τέλος τα συστήματα βασισμένα στη γνώση έχουν την δυνατότητα να δείξουν στον χρήστη το σκεπτικό τους που τους οδήγησε στην δημιουργία αυτών των προτάσεων. Αυτό γίνεται εφόσον ο χρήστης το ζητήσει και ο λόγος που είναι διαθέσιμη αυτή η λειτουργία είναι για να γίνουν οι διεργασίες με διαφάνεια και χωρίς να υπάρχουν υποψίες από τους χρήστες πως τα συστήματα προσπαθούν να προωθήσουν συγκεκριμένα αντικείμενα.

2.7.3 Μειονεκτήματα Των Συστημάτων Βασισμένα Στην Γνώση

Για την κατασκευή ενός συστήματος συστάσεων βασισμένο στην γνώση απαιτείται μια προεργασία η οποία είναι μια μακροχρόνια και επίμονη διαδικασία. Θα πρέπει για την δημιουργία της γνώσης να δημιουργηθεί ένα σύστημα που να μπορεί να αντιληφθεί ένα αντικείμενο μαζί με τα κύρια χαρακτηριστικά του και τις ανάγκες τις οποίες ικανοποιεί. Επιπλέον θα χρειαστεί να κατασκευαστεί και μια βάση δεδομένων στην οποία θα αποθηκεύονται όλα αυτά τα χαρακτηριστικά και να μπορούν να αναζητηθούν με εύκολο και γρήγορο τρόπο. Σε περιπτώσεις όμως που το σύστημα δεν περιέχει μεγάλη ποικιλία αντικειμένων η πολυπλοκότητα των συστημάτων μειώνεται αρκετά.

Τα συστήματα βασισμένα στην γνώση χρησιμοποιούν την γνώση που έχουν ήδη από την δημιουργία του συστήματος για την δημιουργία συστάσεων. Δεν μπορεί όμως να αποκτήσει νέα γνώση από τα δεδομένα που υπάρχουν στο σύστημα παρά μόνο εάν τα εισάγουν οι διαχειριστές. Αυτό απαιτεί ανθρώπινο δυναμικό για να ανανεώνεται η γνώση που διαθέτει το σύστημα και να κάνει πιο εύστοχες συστάσεις.

Τέλος τα δεδομένα που διαθέτει το σύστημα θα πρέπει να είναι πολύ καλά δομημένα και οργανωμένα έτσι ώστε το σύστημα να είναι σε θέση να τα διαχειριστεί.

2.8 Αξιολόγηση Των Συστημάτων Συστάσεων

Σε κάθε τεχνική που αναλύθηκε παραπάνω τα αποτελέσματά της μπορούν να αξιολογηθούν και στην συνέχεια να βελτιωθούν. Υπάρχουν πολλοί τρόποι με τους οποίους ένα σύστημα μπορεί να αξιολογηθεί και γι' αυτόν τον λόγο υπάρχουν και πολλές μέθοδοι. Οι πιο αποτελεσματικοί είναι (Herlocker, 2004):

- Μέσο Απόλυτο Σφάλμα (Mean Absolute Error)

Ο δείκτης του μέσου απόλυτου σφάλματος είναι ένα μέτρο σύγκρισης της απόλυτης διαφοράς που υπάρχει από την πραγματική βαθμολογία που έβαλε ο χρήστης σε ένα αντικείμενο και την βαθμολογία που προτάθηκε από το σύστημα. Όταν ελεγχθεί ο δείκτης αυτός στην συνέχεια τον χρησιμοποιεί ο χρήστης για να κάνει τις αλλαγές που θεωρεί πως πρέπει να γίνουν. Ο δείκτης του μέσου απόλυτου σφάλματος υπολογίζεται από τον τύπο (Breese, Heckerman, & Kadie, Empirical Analysis Of Predictive Algorithms For Collaborative Filtering, 1998):

$$MAE = \frac{\sum_{i=1}^N |p_i - r_i|}{N}$$

Όπου MAE είναι το μέσο απόλυτο σφάλμα, p_i είναι η βαθμολογία που πρότεινε το σύστημα για το αντικείμενο i , r_i είναι η βαθμολογία που έβαλε ο χρήστης και N είναι το πλήθος όλων των εναλλακτικών που έχει στην διάθεση του το σύστημα.

- Ακρίβεια

Ο δείκτης ακρίβειας είναι ένας δείκτης ο οποίος υπολογίζει το κατά πόσο το αποτέλεσμα του αλγορίθμου σύστασης είναι ικανοποιητικό ή όχι. Υπολογίζει δηλαδή αν οι επιλογές που επέλεξε ο αλγόριθμος είναι ορθές ή όχι.

Για να γίνει πιο κατανοητή η χρήση του δείκτη ακρίβειας χωρίζεται σε τέσσερις κατηγορίες, οι δύο αφορούν την προτίμηση του χρήστη και οι άλλες δύο τα αποτελέσματα του αλγορίθμου.

Για την προτίμηση από τον χρήστη υπάρχουν οι σχετικές εναλλακτικές στις οποίες συμπεριλαμβάνονται όλες οι εναλλακτικές που ενδιαφέρουν τον χρήστη και στις μη σχετικές εναλλακτικές οι οποίες αφορούν όλες τις εναλλακτικές που δεν ενδιαφέρουν τον χρήστη (Herlocker, 2004).

Ως αποτέλεσμα του αλγορίθμου υπάρχουν οι επιλεγμένες εναλλακτικές που περιέχουν όλες τις εναλλακτικές που ο αλγόριθμος θεωρεί πως ενδιαφέρουν τον χρήστη, και σαν αποτέλεσμα τις προτείνει σε αυτόν, και τις μη επιλεγμένες εναλλακτικές στις οποίες υπάρχουν όλες οι εναλλακτικές που εξετάστηκαν από τον αλγόριθμο και θεώρησε πως δεν ενδιαφέρουν τον χρήστη, ώστε να μην τις εμφανίσει ποτέ σε αυτόν.

Έτσι αφού ορίστηκε ο δείκτης ακρίβειας μπορεί να υπολογιστεί ο δείκτης ακρίβειας μια εναλλακτικής η οποία επιλέχθηκε και είναι ταυτόχρονα και σχετική με την αναζήτηση του χρήστη από τον τύπο:

$$P = \frac{N_{rs}}{N_s}$$

Όπου N_{rs} είναι οι εναλλακτικές που έχουν επιλεγεί από τον αλγόριθμο και είναι και σχετικές ταυτόχρονα και N_s είναι το σύνολο των εναλλακτικών που έχουν επιλεγεί.

- Δείκτης Ανάκλησης:

Ο δείκτης ανάκλησης είναι συμπληρωματικός του δείκτη ακρίβειας και χρησιμοποιείται για να συμβολίσει τον λόγο μεταξύ των επιλεγμένων σχετικών εναλλακτικών N_{rs} προς το σύνολο των εναλλακτικών που υπάρχουν και υπολογίζεται από τον τύπο:

$$R = \frac{N_{rs}}{N_r}$$

2.9 Υβριδικά Συστήματα (Hybrid Systems)

Τα συστήματα βασισμένα στο περιεχόμενο και τα συστήματα συνεργασίας που αναλύθηκαν παραπάνω είχαν κάποιους περιορισμούς και αν παρατηρηθούν πιο προσεκτικά πολλοί από αυτούς μπορούν να αντιμετωπιστούν με την χρήση τεχνικών από μια άλλη κατηγορία. Γι' αυτόν τον λόγο έχει δημιουργηθεί μια τρίτη κατηγορία συστημάτων η οποία χρησιμοποιεί και τις δύο προαναφερθείσες κατηγορίες και παράλληλα είναι αυτόνομη από αυτές τις δύο. Τα νέα συστήματα ονομάζονται υβριδικά συστήματα και ονομάζονται έτσι καθώς είναι ο συνδυασμός των δύο συστημάτων. Σε γενικές γραμμές τα υβριδικά συστήματα είναι πιο αξιόπιστα σε σύγκριση με τα υπόλοιπα συστήματα που αναφέρθηκαν.

2.9.1 Παράδειγμα Υβριδικών Συστημάτων

Τα τελευταία χρόνια πολλές ιστοσελίδες ηλεκτρονικών αγορών έχουν χρησιμοποιήσει πολλές μεθόδους σύστασης για τα μέλη τους. Η ιστοσελίδα amazon.com είναι ένα χαρακτηριστικό παράδειγμα τέτοιου είδους ιστοσελίδων που χρησιμοποιεί πολύ ισχυρά συστήματα συστάσεων.

Στην παρακάτω εικόνα φαίνεται μια σύσταση της ιστοσελίδας amazon η οποία είναι βασισμένη στις προτιμήσεις των υπόλοιπων πελατών. Η σύσταση που γίνεται είναι βασισμένη στην συνεργασία και περιέχει και τους στατιστικούς λόγους που πρέπει ο χρήστης να προτιμήσει το προϊόν.

Εικόνα 8: Σύσταση Της Ιστοσελίδας Amazon Βασισμένη Στην Συνεργασία



Η παρακάτω εικόνα είναι μια σύσταση που γίνεται στον χρήστη με βάση τις παλαιότερες αγορές που έχει κάνει και η ιστοσελίδα πιστεύει πως μπορεί να τον ενδιαφέρουν. Τα αποτελέσματα προκύπτουν με βάση το φιλτράρισμα βασισμένο στο περιεχόμενο και υπάρχει η δυνατότητα να δει ο χρήστης τους λόγους για τους οποίους του έγινε αυτή η σύσταση.

Εικόνα 9: Σύσταση Της Ιστοσελίδας Amazon Βασισμένο Στο Περιεχόμενο



Τέλος στην παρακάτω εικόνα δείχνει σε έναν πίνακα όλες τις προτάσεις που γίνονται στο χρήστη. Οι συστάσεις που του γίνονται χωρίζονται σε δύο κατηγορίες, τις συστάσεις με βάση το ιστορικό των αγορών του και τις συστάσεις με βάση τις αναζητήσεις του στην σελίδα. Επιπλέον υπάρχει και η επιλογή στην οποία ο χρήστης μπορεί να λάβει νέες προτάσεις βασισμένες στις αγορές που έκαναν σε παρόμοια προϊόντα άλλοι χρήστες.

Εικόνα 10: Σύσταση Της Ιστοσελίδας Amazon Βασισμένα Στο Ιστορικό Του Χρήστη

Amazon Personalized Recommendations System	
Your Browsing History	Your Purchase History
Actual Items	New releases (Item recommendation)
Related Items (Item recommendation)	Related Items (Item recommendation)
Others Purchased (Social Recommendation)	Others Purchased (Social recommendation)

ΚΕΦΑΛΑΙΟ 3: ΠΕΡΙΓΡΑΦΗ ΥΒΡΙΔΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Υπάρχουν πολλοί συνδυασμοί συστημάτων που χρησιμοποιούν τα υβριδικά συστήματα για να δημιουργήσουν νέα συστήματα τα οποία λειτουργούν με μεγαλύτερη αξιοπιστία. Κάποια από τα είδη των υβριδικών συστημάτων περιγράφονται στην συνέχεια.

3.1 Ταυτόχρονη Εφαρμογή Των Συστημάτων Βασισμένα Στο Περιεχόμενο Και Των Συστημάτων Βασισμένα Στην Συνεργασία

Το πρώτο είδος των υβριδικών συστημάτων είναι αυτό που χρησιμοποιεί ξεχωριστά τα συστήματα βασισμένα στο περιεχόμενο και το δεύτερο είδος είναι τα συστήματα βασισμένα στην συνεργασία, όμως τα δύο συστήματα χρησιμοποιούνται ταυτόχρονα. Με τον τρόπο αυτό τα συστήματα μπορούν να αξιοποιήσουν τα αποτελέσματα με διαφορετικούς τρόπους. Τα αποτελέσματα αυτά μπορούν να συγκριθούν και από τις δύο μεθόδους. Ως τελικό αποτέλεσμα θα χρησιμοποιηθεί ένας συνδυασμός τους που όμως είναι πιο ακριβής και από τα δύο αποτελέσματα ξεχωριστά (Claypool, Cokhale, & Miranda, 1999).

Μπορεί επίσης σε ορισμένες περιπτώσεις τα αποτελέσματα να αξιολογηθούν με την χρήση κάποιων συστημάτων αξιολόγησης και να χρησιμοποιηθούν τα αποτελέσματα της μεθόδου που είναι πιο κατάλληλα σύμφωνα με τις ανάγκες του κάθε χρήστη (Tran & Cohen, 2000).

3.2 Χρήση Συστημάτων Με Βάση Την Συνεργασία Με Στοιχεία Των Συστημάτων Με Βάση Το Περιεχόμενο

Υπάρχουν αρκετά συστήματα βασισμένα στην συνεργασία που χρησιμοποιούν τεχνικές των συστημάτων με βάση το περιεχόμενο για να έχουν καλύτερα αποτελέσματα σε ορισμένους τομείς. Ένα χαρακτηριστικό παράδειγμα είναι το σύστημα Fab το οποίο χρησιμοποιεί τεχνικές βασισμένες στην συνεργασία όμως για

την αξιολόγηση των προφίλ των χρηστών χρησιμοποιούνται συστήματα βασισμένα στο περιεχόμενο. Έτσι για την συσχέτιση των αποτελεσμάτων δεν χρησιμοποιούνται οι αξιολογήσεις που έκανε ο χρήστης στο παρελθόν αλλά χρησιμοποιούνται τα αποτελέσματα από το σύστημα με βάση το περιεχόμενο (Balabanovic & Shoham, <https://www.ischool.utexas.edu/>, 1997).

Επιπλέον με την χρήση αυτής της μεθόδου στο χρήστη δεν εμφανίζονται οι εναλλακτικές με την μεγαλύτερη βαθμολογία από χρήστες που έχουν κοινά χαρακτηριστικά αλλά εμφανίζονται τα αποτελέσματα που έχουν υψηλή βαθμολογία από χρήστες που το προφίλ τους έχει μεγάλη συσχέτιση με το προφίλ του χρήστη που κάνει την αναζήτηση.

3.3 Χρήση Συστημάτων Βασισμένα Στην Συνεργασία Σε Συστήματα Με Βάση Το Περιεχόμενο

Τα συστήματα αυτά δεν χρησιμοποιούν τόσο συχνά τον προηγούμενο συνδυασμό συστημάτων. Η τεχνική αυτή χρησιμοποιείται για την μείωση των προφίλ που υπάρχουν από τα αποτελέσματα των συστημάτων με βάση των περιεχόμενο. Ένα παράδειγμα αυτού του συστήματος είναι η μέθοδος LSI (Latent Semantic Indexing) το οποίο χρησιμοποιεί τα συστήματα με βάση τη συνεργασία για να βελτιώσει τα αποτελέσματα των συστημάτων με βάση το περιεχόμενο (Soboroff, 1999).

3.4 Νέο Μοντέλο Με Χρήση Συστημάτων Με Βάση Το Περιεχόμενο Και Συστημάτων Βασισμένα Στην Συνεργασία

Πρόκειται για μια τεχνική η οποία έχει αρχίσει να χρησιμοποιείται αρκετά τα τελευταία χρόνια και για την λειτουργία της τεχνικής αυτής χρησιμοποιείται ένα νέο μοντέλο το οποίο αποτελεί συνδυασμό πολλών γνωστών τεχνικών με σκοπό την μείωση των εμποδίων που θέτει η χρήση της κάθε μεθόδου ξεχωριστά.

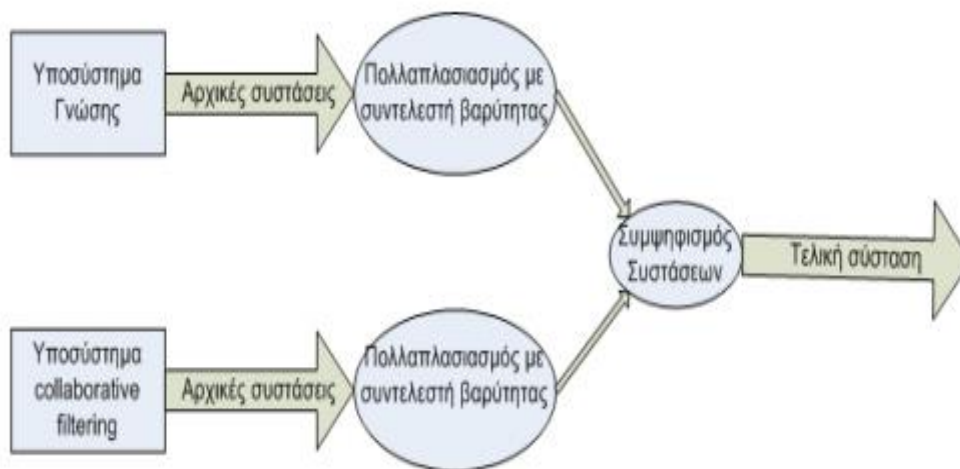
Ένα παράδειγμα αυτής της κατηγορίας είναι τα συστήματα που βασίζονται στην γνώση. Τα συστήματα αυτά όμως δεν μπορούν να παράξουν ή να αποκτήσουν γνώση άρα λειτουργούν μόνο σε τομείς όπου η γνώση υπάρχει ήδη και είναι στην κατάλληλη μορφή ώστε να την επεξεργαστούν.

3.5 Υβριδικά Συστήματα Συντελεστών Βαρύτητας

Το υβριδικό συστήματα συντελεστών βαρύτητας εκτελεί τις τεχνικές των συστημάτων βασισμένα στην συνεργασία και των συστημάτων βασισμένα στην γνώση προκειμένου να συγκρίνει τα αποτελέσματά του και να κάνει καλύτερες συστάσεις στον χρήστη. Η λειτουργία των συστημάτων χωρίζεται σε δύο στάδια. Η παρακάτω εικόνα βοηθάει στην καλύτερη κατανόηση της λειτουργίας των συστημάτων συντελεστών βαρύτητας (Resnick & Varian, 1997):

- Στο πρώτο στάδιο τους τα συστήματα συντελεστών βαρύτητας χρησιμοποιούν ξεχωριστά τα συστήματα βασισμένα στην συνεργασία και τα συστήματα βασισμένα στην γνώση χρησιμοποιώντας και από τα δύο τα διαθέσιμα στοιχεία που υπάρχουν για τον χρήστη και για τα αντικείμενα. Έτσι δίνουν δύο διαφορετικά αποτελέσματα. Τα αποτελέσματα αυτά είναι βαθμολογημένα ανάλογα με την πεποίθηση του κάθε συστήματος ότι ένα αντικείμενο θα ικανοποιεί τον χρήστη.
- Στην συνέχεια το σύστημα συντελεστών βαρύτητας δέχεται τις βαθμολογίες των αντικειμένων που του έδωσαν τα δύο προηγούμενα συστήματα και τις πολλαπλασιάζει με ένα συγκεκριμένο συντελεστή βαρύτητας. Οι βαθμολογίες που προκύπτουν είναι τα τελικά αποτελέσματα των συστημάτων συντελεστών βαρύτητας και τα αντικείμενα με την υψηλότερη τελική βαθμολογία είναι αυτά τα οποία προτείνονται στον χρήστη.

Εικόνα 11: Περιγραφή Λειτουργίας Συστήματα Συντελεστών Βαρύτητας



Όπως φαίνεται για να είναι εύστοχα τα αποτελέσματα των συστημάτων συντελεστών βαρύτητας θα πρέπει να υπολογιστεί ο συντελεστής βαρύτητας για το κάθε σύστημα που θα χρησιμοποιηθεί. Επειδή ο υπολογισμός των συντελεστών είναι μια δύσκολη διαδικασία και πολλές φορές μπορεί να μην είναι ακριβή τα αποτελέσματα που βγαίνουν προστίθεται και ένα τρίτο βήμα το οποίο λειτουργεί ως ανατροφοδότηση. Έχουν την δυνατότητα δηλαδή αυτά τα συστήματα να ελέγχουν την αντίδραση του χρήστη και να μεταβάλλουν τους συντελεστές ώστε να πετύχουν καλύτερα αποτελέσματα. Το τελευταίο βήμα είναι ένα σημαντικό πλεονέκτημα που έχουν τα συστήματα συντελεστών βαρύτητας καθώς μπορούν απλά μεταβάλλοντας τον συντελεστή βαρύτητας τους να μεταβάλουν την αποδοτικότητά τους.

Τα βασικό μειονέκτημα που έχουν τα συστήματα συντελεστών βαρύτητας είναι πως θεωρούν σταθερή την απόδοση των συστημάτων βασισμένα στην γνώση και των συστημάτων βασισμένα στην συνεργασία. Αυτό όμως δεν είναι ακριβές γιατί αν το σύστημα βασισμένο στην συνεργασία δεν περιέχει αρκετές αξιολογήσεις τότε ο συντελεστής βαρύτητας των συστημάτων βασισμένα στην γνώση θα έχει μεγαλύτερο βαθμό από τον αντίστοιχο του συστήματος βασισμένα στην συνεργασία. Αντιθέτως αν οι γνώσεις που έχουν τα συστήματα βασισμένα στην γνώση για κάποιο αντικείμενο η για μια ομάδα αντικειμένων δεν είναι αρκετές τότε ο συντελεστής βαρύτητας των συστημάτων βασισμένα στην συνεργασία θα είναι μεγαλύτερος. Έτσι γίνεται φανερό πόσο δύσκολη και επίπονη διαδικασία είναι ο υπολογισμός των συντελεστών βαρύτητας (Towle & Quinn, 2000).

3.6 Συστήματα Εναλλαγής Τεχνικών

Τα συστήματα εναλλαγής τεχνικών είναι μια ακόμα μορφή υβριδικών συστημάτων τα οποία μπορούν να χρησιμοποιήσουν είτε τα συστήματα βασισμένα στην συνεργασία είτε τα συστήματα βασισμένα στην γνώση ανάλογα με το ποια μέθοδος λειτουργεί πιο αποδοτικά σε κάθε περίπτωση. Επιπλέον στα συστήματα εναλλαγής τεχνικών υπάρχει και ένα σύστημα το οποίο περιέχει κάποια κριτήρια και ελέγχει ποια από αυτά πληρεί κάθε μέθοδος και αποφασίζει ποια από τις δύο μεθόδους θα χρησιμοποιηθεί. Και σε αυτήν την περίπτωση τα δύο συστήματα λειτουργούν ανεξάρτητα από το άλλο. Υπάρχουν δύο τρόποι με τους οποίους αποφασίζεται ποια μέθοδος θα χρησιμοποιηθεί (Burke, Hybrid Recommender Systems: Survey and Experiments, 2000):

- Η πρώτη μέθοδος εκτελεί και τα δύο συστήματα και στην συνέχεια συγκρίνει τις προτάσεις του κάθε συστήματος με τις αξιολογήσεις που έχει κάνει στο παρελθόν ο χρήστης. Όποιο σύστημα έχει τα μεγαλύτερα κοινά με τις απαιτήσεις του χρήστη τότε τα δικά του αποτελέσματα παρουσιάζονται στο χρήστη.

- Η δεύτερη μέθοδος χρησιμοποιεί ένα από τα δύο διαθέσιμα συστήματα για την παραγωγή συστάσεων στον χρήστη. Αν τα αποτελέσματα που βγάλει το πρώτο σύστημα δεν είναι ικανοποιητικά ή τα απορρίπτει ο χρήστης κατά την παρουσίασή τους τότε εκτελείται το δεύτερο σύστημα. Στα περισσότερα συστήματα εναλλαγής τεχνικών χρησιμοποιείται πρώτα το σύστημα βασισμένο στην γνώση και στην συνέχεια αν δεν μπορεί αυτό να κάνει ικανοποιητικές συστάσεις χρησιμοποιείται το σύστημα βασισμένο στην συνεργασία.

Το πλεονέκτημα της μεθόδου είναι πως γίνεται το σύστημα πιο ισχυρό δίνοντας του την δυνατότητα να προτείνει στον χρήστη αντικείμενα που τον ενδιαφέρουν ακόμα και αν αυτά δεν έχουν μεγάλες ομοιότητες. Το μειονέκτημα της είναι πως αυξάνεται πολύ η πολυπλοκότητα του συστήματος γιατί πρέπει από μόνο του το σύστημα να επιλέξει την μέθοδο που θα επακολουθήσει.

Ένα χαρακτηριστικό παράδειγμα των συστημάτων εναλλαγής τεχνικών είναι το Gico το οποίο χρησιμοποιεί την ομοιότητα κειμένου για την δημιουργία συστάσεων. Όπως φαίνεται από την παρακάτω εικόνα το σύστημα Gico προτείνει στο χρήστη να διαβάσει κάποια άρθρα. Δίπλα από τις επικεφαλίδες των άρθρων έχει ένα εικονίδιο με μια μπλε μπάρα που δείχνει πόσο δημοφιλές είναι το καθένα με κλίμακα από το ένα έως το τέσσερα. Στο ίδιο εικονίδιο υπάρχει και μια κόκκινη μπάρα με την ίδια κλίμακα που δείχνει πόσο ταιριάζει το κάθε άρθρο με τα προηγούμενα που έχει διαβάσει ο χρήστης (Pazzani & Billsus, Content-based Recommendation Systems, 2007).

Εικόνα 12: Το Σύστημα Gixo

Gixo Gixo is the most personalized news service available. Gixo instantly recommends current news based on the articles you've read in the past. It will organize the headlines to match your interests. You can also customize the layout yourself. ([Learn More...](#))

You have clicked: 17 articles. Review your clicked articles to get better recommendations.

Columns: 1 2 3 Font: A A A Intro Text: on off Images: on off View Articles In: game window new window More Options

by Jade Official Site Login or Register to save your preferences: (Optional, Free, & Anonymous) Username: Password: Login/Register

Technology

[Another Hidden Gem: The Windows Disk Management Tool](#)
InformationWeek | Windows Disk Management | Langa Letter: Another Hidden Gem: The Windows Disk... (informationweek.com)

[Last post: 3/31/2006 1:20PM](#)
I have almost decided to stop buying games for the PC. I bought and installed a Ubisoft game... (extremetech.com)

[eBay Urges High Court to 'Buy It Now'](#)
Intel delivers virtualization technology to address critical IT challenges- the need for increased... (internetnews.com)

[Microsoft's Linux site](#)
Microsoft on Thursday at LinuxWorld is expected to unveil a new Web site for users to find... (infoworld.com)

3.7 Μεικτά Συστήματα

Τα μεικτά συστήματα συστάσεων είναι μια μορφή συστημάτων συστάσεων τα οποία χρησιμοποιούν τα συστήματα βασισμένα στην γνώση και τα συστήματα βασισμένα στην συνεργασία ταυτόχρονα και ανεξάρτητα το ένα από το άλλο. Τα δύο συστήματα καταλήγουν σε έναν αριθμό προτάσεων τα οποία παρουσιάζονται στον χρήστη. Με την χρήση των μεικτών συστημάτων σύστασης ο χρήστης έχει την δυνατότητα να παίρνει προτάσεις από δύο διαφορετικά συστήματα και να επωφελείται από τα θετικά στοιχεία και των δύο. Πρέπει να σημειωθεί όμως πως για να χρησιμοποιηθεί αυτή η μέθοδος θα πρέπει να υπάρχει χώρος για να παρουσιαστούν πολλά αποτελέσματα (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005).

Τα πλεονεκτήματα της μεθόδου είναι πως στα σημεία που αδυνατεί να βγάλει ασφαλή αποτελέσματα η μια μέθοδος βγάζει η άλλη. Αναλυτικά όταν τα συστήματα βασισμένα στην συνεργασία δεν μπορούν να αντιμετωπίσουν το πρόβλημα των νέων αντικειμένων ή χρηστών το σύστημα βασισμένο στην γνώση μπορεί να χρησιμοποιηθεί με αξιόπιστες προτάσεις (Adomavicius & Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005).

ΓΡΑΦΟΘΕΩΡΗΤΙΚΕΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΗΣ

Το μειονέκτημα των μεικτών συστημάτων είναι το γεγονός πως οι προτάσεις που κάνει στον χρήστη είναι πολλές σε αριθμό και πως αν δεν υπάρχει χώρος εμφάνισης τους ή ο χρήστης επιθυμεί πιο λίγες θα πρέπει να δημιουργηθεί ένα σύστημα αξιολόγησης των προτάσεων για να εμφανίζονται μόνο αυτές που έχουν τις περισσότερες πιθανότητες να ενδιαφέρουν τον χρήστη. Ένα ακόμα μειονέκτημα της μεθόδου αυτής εμφανίζεται στην περίπτωση όπου τα δύο συστήματα προτείνουν δύο αντικείμενα τα οποία είναι αντικρουόμενα το ένα από το άλλο. Σε αυτήν την περίπτωση θα πρέπει να αξιολογηθούν τα αποτελέσματα και να επιλεγεί ποια από τις δύο προτάσεις θα προβληθεί στον χρήστη και ποια θα διαγραφεί.

Χαρακτηριστικό παράδειγμα των μεικτών συστημάτων συστάσεων είναι το σύστημα PTV. Στο σύστημα αυτό όπως φαίνεται και στην παρακάτω εικόνα προτείνει στο χρήστη προγράμματα για να παρακολουθήσει στην τηλεόραση. Έχει προβλέψει το ενδεχόμενο τα δύο συστήματα να προτείνουν ένα πρόγραμμα το οποίο να παίζει στην ίδια ώρα και ως λύση προτείνεται το πρόγραμμα που προτείνει το σύστημα βασισμένο στην γνώση.

Εικόνα 13: Το Σύστημα PTV



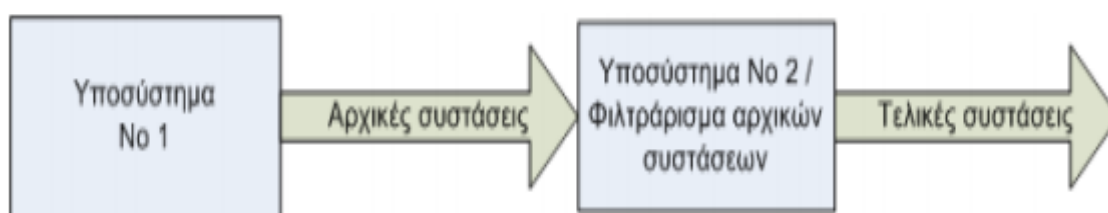
3.8 Υβριδικά Συστήματα Τύπου Καταρράκτη

Στα συστήματα τύπου καταρράκτη χρησιμοποιούνται και πάλι τα συστήματα βασισμένα στην συνεργασία και τα συστήματα βασισμένα στην γνώση. Η διαφορά όμως με τις υπόλοιπες τεχνικές είναι πως εδώ τα δύο συστήματα δεν λειτουργούν ανεξάρτητα μεταξύ τους αλλά το ένα λειτουργεί με τα αποτελέσματα του άλλου. Εκτελείτε δηλαδή η μία τεχνική και στην συνέχεια χρησιμοποιείται η άλλη για να ταξινομήσει τα

αποτελέσματά της. Η λειτουργία του αποτελείται από δύο στάδια και περιγράφεται από την παρακάτω εικόνα (Alspector, Kolcz, & Karunanithi, 1997):

- Στο πρώτο στάδιο του υπολογισμού των συστάσεων λειτουργεί το πρώτο από τα δύο συστήματα συνήθως το σύστημα βασισμένο στην γνώση. Τα αποτελέσματα του πρώτου συστήματος πρέπει να είναι πολλά σε αριθμό.
- Στο δεύτερο στάδιο του υπολογισμού των συστάσεων τα αποτελέσματα του πρώτου συστήματος γίνονται είσοδοι του δεύτερου συστήματος. Τα αποτελέσματα του δεύτερου συστήματος είναι οι ίδιες συστάσεις ταξινομημένες και βελτιωμένες.

Εικόνα 14: Περιγραφή λειτουργίας Συστημάτων Τύπου Καταρράκτη



Το πλεονέκτημα της μεθόδου αυτής είναι ότι η απόδοση του συστήματος είναι πολύ μεγάλη. Αυτό συμβαίνει γιατί το δεύτερο σύστημα εφαρμόζεται πάνω σε αντικείμενα τα οποία έχουν ήδη αξιολογηθεί και κατά πολύ μεγάλη πιθανότητα θα ενδιαφέρουν τον χρήστη. Άρα και τα τελικά αποτελέσματα που θα παρουσιαστούν στον χρήστη θα έχουν πολύ μεγαλύτερη πιθανότητα να τον ενδιαφέρουν. Ακόμα το σύστημα τύπου καταρράκτη είναι πιο γρήγορο από άλλα συστήματα γιατί το δεύτερο υποσύστημα που εκτελείτε λειτουργεί για ένα μικρό αριθμό αντικειμένων και όχι για όλα τα αντικείμενα που υπάρχουν καταχωρημένα στο σύστημα.

Το μειονέκτημα της εισόδου είναι πως το δεύτερο υποσύστημα ουσιαστικά φιλτράρει τα αποτελέσματα του πρώτου υποσυστήματος και βελτιώνει την αξιοπιστία των αποτελεσμάτων. Έτσι αν το πρώτο υποσύστημα βγάλει λανθασμένες προτάσεις το δεύτερο δεν θα μπορέσει να τις αλλάξει και να κάνει νέες.

3.9 Υβριδικά Συστήματα Συνδυασμού Χαρακτηριστικών

Τα υβριδικά συστήματα συνδυασμού χαρακτήρων είναι άλλη μια κατηγορία υβριδικών συστημάτων τα οποία περιέχουν δύο υποσυστήματα που το ένα εξαρτάται από το άλλο. Σε αυτού του είδους τα συστήματα χρησιμοποιούνται τα στοιχεία που έχει στην διάθεσή του το σύστημα βασισμένο στην συνεργασία για να παραχθεί γνώση για το συγκεκριμένο αντικείμενο. Στην συνέχεια αυτή την γνώση που παράχθηκε την χρησιμοποιεί το σύστημα βασισμένο στην γνώση για να κάνει τις τελικές του προτάσεις στον χρήστη. Αναλυτικά η μέθοδος λειτουργεί με τα δύο ακόλουθα βήματα (Basu, Hirsh, & Cohen, 1998):

- Στο πρώτο βήμα το σύστημα μαζεύει τα στοιχεία που έχει στην διάθεση του από το σύστημα βασισμένα στην συνεργασία τα οποία κατά κύριο λόγο είναι αξιολογήσεις από άλλους χρήστες. Τα στοιχεία αυτά προστίθενται στα στοιχεία γνώσης που περιέχονται ήδη στις βάσεις δεδομένων στο σύστημα βασισμένα στην γνώση και έτσι δημιουργείται μια νέα βάση δεδομένων με περισσότερα στοιχεία για το κάθε αντικείμενο ξεχωριστά.
- Στην συνέχεια στο δεύτερο βήμα της μεθόδου εκτελείται το σύστημα βασισμένο στην γνώση αλλά ως γνώση το σύστημα χρησιμοποιεί την νέα βάση δεδομένων που δημιουργήθηκε από το πρώτο βήμα. Έτσι τα αποτελέσματα που συστήνονται στον χρήστη είναι προϊόν του συστήματος που είναι βασισμένο στην γνώση αλλά έχει όμως χρησιμοποιήσει στοιχεία από τα συστήματα βασισμένα στην συνεργασία.

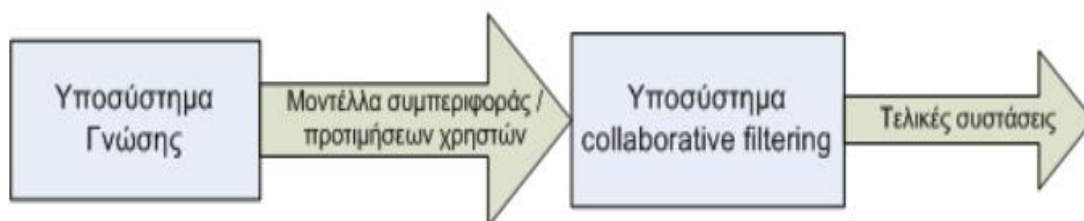
Το πλεονέκτημα αυτής της μεθόδου είναι πως αυξάνετε η αξιοπιστία της μεθόδου βασισμένα στην γνώση αλλά δεν εξαρτάται άμεσα από τον αριθμό των αξιολογήσεων που υπάρχει για κάθε αντικείμενο.

3.10 Υβριδικά Συστήματα Μεταφοράς Μοντέλου

Το σύστημα μεταφοράς μοντέλου είναι ένα σύστημα που όπως και τα προηγούμενα συστήματα που έχουν αναφερθεί περιέχει δύο υποσυστήματα που το ένα επιδρά με το άλλο. Έτσι εκτελείται το πρώτο υποσύστημα και τα αποτελέσματά του χρησιμοποιούνται για την δημιουργία ενός μοντέλου το οποίο θα χρησιμοποιηθεί ως βασικό στοιχείο για την λειτουργία του δευτέρου υποσυστήματος. Η διαφορά του συστήματος αυτού από την μέθοδο του συνδυασμού χαρακτήρων είναι πως δεν μεταφέρονται απλά τα στοιχεία από το ένα μοντέλο στο άλλο αλλά υπάρχει μεταφορά ενός ολοκληρωμένου μοντέλου που περιγράφει την διάθεση των χρηστών. Η λειτουργία του συστήματος χωρίζεται σε δύο βήματα που περιγράφονται από την παρακάτω εικόνα (Balabanovic, Exploring versus Exploiting when Learning User Models for Text Representation, 1998):

- Το πρώτο βήμα του συστήματος είναι η δημιουργία του μοντέλου. Αυτό γίνεται από το υποσύστημα βασισμένο στην γνώση το οποίο αφού λάβει υπ' όψιν του τις προτιμήσεις των χρηστών δημιουργείται ένα μοντέλο που περιγράφει την συμπεριφορά και τις προτιμήσεις του κάθε χρήστη.
- Στην συνέχεια στο δεύτερο βήμα το μοντέλο περνάει στο δεύτερο υποσύστημα για την αξιοποίηση του για την παραγωγή συστάσεων στους χρήστες. Έτσι το μοντέλο που έχει ήδη δημιουργηθεί από το υποσύστημα βασισμένο στην γνώση γίνεται είσοδος για το υποσύστημα βασισμένο στην συνεργασία. Αφού το σύστημα βασισμένο στην συνεργασία δεχθεί την είσοδο του τότε λειτουργεί ως ένα κλασσικό σύστημα βασισμένο στην συνεργασία χρησιμοποιώντας κατάλληλα το μοντέλο που έχει στην διάθεσή του και κάνει τις συστάσεις στο χρήστη.

Εικόνα 15: Περιγραφή Λειτουργίας Συστημάτων Μεταφοράς Μοντέλου



ΚΕΦΑΛΑΙΟ 4: ΥΛΟΠΟΙΗΣΗ ΑΛΓΟΡΙΘΜΩΝ

Σε αυτό το κεφάλαιο θα αναφερθούμε συνοπτικά στην πειραματική διαδικασία που ακολουθήσαμε στο σύστημά μας για την πρόβλεψη των αποτελεσμάτων του αλγορίθμου μας. Θα αναλύσουμε την πειραματική διαδικασία, ενώ στο τέλος θα προβληθούν τα αποτελέσματα της μελέτης μας.

4.1 Διατύπωση του προβλήματος

Στα πειράματα που πραγματοποιούμε σε αυτή τη διπλωματική, χρησιμοποιούμε ένα σύνολο από δεδομένα που περιλαμβάνουν αξιολογήσεις των χρηστών σε ταινίες. Οποιοδήποτε σύνολο δεδομένων που περιλαμβάνει αντικείμενα, χρήστες και ανάδραση των χρηστών σε αντικείμενα με την μορφή αξιολογήσεων μπορούμε να το θεωρήσουμε έτοιμο για να χρησιμοποιηθεί από την προσέγγισή μας. Εδώ, τα δεδομένα εισόδου μπορεί να να εκπροσωπούνται από τις αξιολογήσεις μιας μήτρας, η οποία έχει μία γραμμή για κάθε χρήστη και μία στήλη για κάθε στοιχείο και στην οποία τα στοιχεία δείχνουν τις βαθμολογίες που δόθηκαν από τους χρήστες για τα διάφορα θέματα.

Η εργασία αυτή προτείνει μια προσέγγιση που ασχολείται με άγνωστες τιμές σε μια πιθανολογική αναπαράσταση του προβλήματος. Δηλαδή, αντί να εργάζονται με τις αξιολογήσεις, οι αλγόριθμοι μας εργάζονται με πιθανολογικές κατανομές πάνω από το πεδίο των πιθανών αξιολογήσεων. Εάν μια αξιολόγηση ενός χρήστη για ένα στοιχείο είναι γνωστή, τότε η αντίστοιχη κατανομή πιθανότητας έχει μηδενική διακύμανση και τοποθετεί την τιμή της πιθανότητας για τη συγκεκριμένη βαθμολογία. Εάν δεν είναι γνωστή τότε οι αλγόριθμοι κάνουν χρήση μιας ενιαίας, παγκόσμιας, προβλεπόμενης κατανομής.

Στην κορυφή αυτής της πιθανοτικής εκπροσώπησης, προτείνουμε ένα εκτιμώμενο μέτρο απόστασης για να υπολογίσουμε την ομοιότητα μεταξύ του προφίλ των δύο χρηστών.

Αυτό το μέτρο μπορεί να χρησιμοποιηθεί για τον πλησιέστερο γείτονα βάση των αλγορίθμων συνεργατικής διήθησης. Όπως θα δούμε, αυτό βελτιώνει σημαντικά την απόδοση της προσέγγισης του πλησιέστερου γείτονα (KNN).

Δεύτερη προσέγγιση μας για να βελτιώσουμε την κατάσταση είναι η προσέγγιση KNN που αποτελείται από έναν αλγόριθμο γράφημα ο οποίος καλείται γράφημα βάση φιλτραρίσματος συνεργασίας (PGBCF). Αυτός ο αλγόριθμος βασίζεται σε ένα γράφημα χρήστη στο οποίο οι χρήστες είναι οι κόμβοι και οι ακμές επισημαίνονται με βάση το μέτρο απόστασης. Ο αλγόριθμος προσπαθεί να βελτιωθεί με τη χρήση πληροφοριών που υπάρχουν στους έμμεσους γείτονες.

4.2 Σχετικές συναφείς εργασίες

Μια πληθώρα διαφορετικών τεχνικών συλλογικού φιλτραρίσματος

περιγράφονται και συγκρίνονται, συμπεριλαμβανομένης

της προσέγγισης του πλησιέστερου γείτονα, τον οποίο θα χρησιμοποιήσουμε για να

αξιολογήσουμε το προτεινόμενο μέτρο απόστασης. Οι μέθοδοι συνήθως σχεδιάζονται για να καταπολεμήσουν το πρόβλημα της σποραδικότητας, και το πρόβλημα της ψυχρής εκκίνησης. Συχνά αποτελούνται από μια υβριδική μέθοδο μεταξύ ενός συνεργατικού φίλτρου και ενός συστήματος με βάση το περιεχόμενο. Η προσέγγισή μας διαφέρει με αυτό που εννοούμε συνεργατικό φίλτρο. Η προτεινόμενη μέθοδος μας χρησιμοποιεί μια πιθανολογική παράσταση.

Μια σειρά από προσεγγίσεις που βασίζονται γράφημα έχουν προταθεί στο παρελθόν , καθώς και μερικά από τα οποία ήταν σχεδιασμένα για να λύσουν το πρόβλημα της σποραδικότητας.

Ο Huang et al. [4] κατασκεύασε ένα διμερές γράφημα , που περιέχει τους χρήστες σε μία πλευρά και τα στοιχεία από την άλλη πλευρά. Μια άκρη μεταξύ ενός χρήστη και ένα στοιχείο προστίθεται όταν ο χρήστης θέλει το προϊόν . Ένα στοιχείο στη συνέχεια κατατάσσεται για ένα χρήστη με βάση τον αριθμό των διαδρομών μεταξύ αυτού του στοιχείου. Αν και το σύστημα δίνει καλά αποτελέσματα, είναι περιορισμένο στη χρήση επειδή μπορεί να χειριστεί μόνο δυαδικά δεδομένα. Ωστόσο όταν ενός χρήστη , είτε αρέσει είτε αντιπαθεί ένα στοιχείο , τότε τα άκρα δεν θα έχουν βάρος .

Ο Παπαγγελής προτείνει έναν άλλο αλγόριθμο βασισμένο σε γράφημα όπου περιέχονται μόνο χρήστες . Το σύστημα χρησιμοποιεί μόνο δυαδικά δεδομένα. Οι ακμές υπολογίζονται με την βοήθεια του βάρους που προκύπτει από τις βαθμολογίες μεταξύ δύο χρηστών (Papangelis, 2005).

Στην έρευνα Alexandros περιγράφεται μία μέθοδος η οποία εκμεταλλεύεται τις μεταβατικές συσχετίσεις μεταξύ των στοιχείων, δηλαδή ορισμένα στοιχεία μπορεί να συσχετίζονται σε μεγάλο βαθμό μεταξύ τους. Για παράδειγμα η βαθμολογία για ένα στοιχείο θα μπορούσε να μας πει κάτι σχετικά με την βαθμολογία για άλλα στοιχεία αντίστοιχα. Ο αλγόριθμος συνιστά το top- k (μεταβατικά) συσχετίζονται στοιχεία . Αυτό μπορεί να θεωρηθεί ως μια υβριδική προσέγγιση για το πρόβλημα της σποραδικότητας. (Alexandros Nanopoulos, 2007)

Ο Huang et al. παρουσιάζει επίσης ένα σύστημα συστάσεων που βασίζεται στο γράφημα.

Σε αντίθεση με το σύστημά μας , το σύστημα αυτό προτείνει μόνο μια σειρά από αντικείμενα και καμία άλλη πρόβλεψη δεν δημιουργείται για άλλα στοιχεία. Αυτό το σύστημα χρησιμοποιεί μια διστρώματική γραφική αναπαράσταση , που περιέχει το χρήστη - χρήστη , στοιχείο - στοιχείο και το χρήστη με τις ακμές των στοιχείων, όπου αυτό το γράφημα στη συνέχεια ψάχνει για αξιόσυστατα αντικείμενα (Zan Huang, 2002).

Καμία όμως από αυτές τις μεθόδους βασισμένες στο γράφημα δεν μπορεί να συγκριθεί άμεσα με τη μέθοδό μας, είτε επειδή προβλέπουν μια κατάταξη αντί της απόλυτης τιμής ή επειδή μπορούν να χειριστούν μόνο δυαδικά δεδομένα. Η κύρια διαφορά μεταξύ EMD - Uni και EMD - Ποινής βρίσκεται στα στοιχεία που έχουν αξιολογηθεί από ένα μόνο από τους χρήστες, για αυτά τα στοιχεία η MD - Ποινή αποδίδει επίσης ποινή C , ενώ EMD-Uni υπολογίζει την αναμενόμενη απόσταση με βάση τις δοθείσες κατανομές.

4.3 Μέτρα απόστασης και Εκπροσώπηση

Σε αυτή την ενότητα, θα συζητήσουμε πρώτα μια δημοφιλή μέθοδο η οποία ασχολείται με τις αξιολογήσεις που λείπουν κατά τον υπολογισμό της ομοιότητας του προφίλ δύο χρηστών. Στη συνέχεια θα παρουσιάσουμε το προφίλ ενός χρήστη και θα καθορίσουμε την έννοια της αναμενόμενης απόστασης, η οποία οδηγεί στην καλύτερη διαχείριση των χαμένων αξιολογήσεων.

Μέτρο απόστασης(Distance Measures)

Η ομοιότητα δύο χρηστών μπορεί να υπολογιστεί με τα μέτρα απόστασης κατά βάση είτε με το μέτρο ομοιότητας ή αλλιώς συντελεστή συσχέτισης όπως ονομάζεται. Προς απλοποίηση της μελέτης μας αρχικά θα θεωρήσουμε ότι δεν έχουμε καμία απουσία βαθμολογίας για κανένα στοιχείο από τον κάθε χρήστη, κατά συνέπεια θα έχουμε να μελετήσουμε την περίπτωση ενός ιδανικού πίνακα. Στην συνέχεια θα ορίσουμε πως τα μέτρα των αποστάσεων μπορούν να χρησιμοποιηθούν ώστε να χειριστούμε βαθμολογίες που λείπουν.

Ας θεωρήσουμε ότι μελετάμε δύο χρήστες p και q , όπου καλούνται να βαθμολογήσουν σε συγκεκριμένο εύρος τιμών R από το ένα έως το πέντε, δηλαδή $\{R = 1, 2, 3, 4, 5\}$.

Ως προς τον συντελεστή συσχέτισης μπορούμε να ορίσουμε ότι:

$$pc(\mathbf{p}, \mathbf{q}) = \frac{\sum_i (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum_i (p_i - \bar{p})^2 \sum_i (q_i - \bar{q})^2}}$$

Ως προς τον αλγόριθμο Manhattan μπορούμε να ορίσουμε το μέτρο της απόστασης ως:

$$d_{MD}(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^n |p_i - q_i|$$

Στην περίπτωση της απουσίας βαθμολογίας ο συντελεστής συσχέτισης μπορεί να υπολογιστεί μόνο από τα γνωστά στοιχεία των δύο χρηστών p, q .

Κάτι τέτοιο δεν μπορεί να γίνει στον αλγόριθμο Manhattan εφόσον εδώ το εύρος των μέτρων εξαρτάται από το πλήθος των στοιχείων. Αν υποθέσουμε ότι έχουμε την ίδια απόσταση σε κάθε διάσταση θα καταλήγαμε σύμφωνα με τον παραπάνω υπολογισμό σε ένα πολύ μικρό αριθμό για όσο πιο πολλά στοιχεία έλειπαν ή παρέμεναν αβαθμολόγητα. Αποτέλεσμα το οποίο δεν θα βοηθούσε καθόλου, και παράλληλα θα ήταν ανακριβές.

Κατά συνέπεια θα υπολογίζουμε το μέτρο στον αλγόριθμο Manhattan με άλλες παραμέτρους ως εξής:

$$d_{MD-Scaled}(p, q) = \frac{n}{|K(p) \cap K(q)|} \sum_{i \in (K(p) \cap K(q))} |p_i - q_i|$$

Με $K(u) = \{i | u_i \neq 0\}$ δηλαδή για το ζεύγος των θέσεων του πίνακα των οποίων τα στοιχεία είναι γνωστά. Για παράδειγμα ας θεωρήσουμε ότι έχουμε δύο χρήστες p, q με τις βαθμολογίες τους για πέντε προϊόντα

$$\left. \begin{array}{l} P = \{3.1.?.0.?\} \\ Q = \{1.2.?.?.?\} \end{array} \right\} d_{MD-Scaled} = 5/2(|3-1| + |1-2|) = 7.5$$

Όπως παρατηρούμε η απόδοση των μέτρων δεν αποδίδεται βέλτιστα με πενιχρές αξιολογήσεις με αποτέλεσμα και οι δύο μέθοδοι να χαρακτηρίζονται ως απροσδιόριστοι όταν οι χρήστες δεν έχουν συνεκτιμημένα στοιχεία.

Για να χειριστούμε λοιπόν την προαναφερθείσα αστάθεια προτείνεται μια μεθοδολογία η οποία βασίζεται σε πιθανολογικές κατανομές πάνω στο πεδίο των πιθανών αξιολογήσεων. Εάν αντικαταστήσουμε κάθε διανυσματική συνιστώσα p_i με την οριακή κατανομή πιθανότητας τότε θα έχουμε:

$$Pr(p_i) \quad (p_i \in R \quad Pr(p_i = u) = 1)$$

Οπότε η αναμενόμενη υπολογίσιμη απόσταση Manhattan (EMD) μεταξύ δύο πιθανών χρηστών ορίζεται ως:

$$d_{EMD}(p, q) = E[\|p - q\|_1] = \sum_{i=1}^n \sum_{u_p \in R} \sum_{u_q \in R} |u_p - u_q| Pr(p_i = u_p) Pr(q_i = u_q)$$

όπως έχουμε ήδη αναφέρει.

Το αποτέλεσμα είναι τώρα η διανυσματική συνιστώσα για κάθε στοιχείο που δεν έχει αξιολόγηση, να είναι διαφορετική και να μπορούμε να επιλέξουμε ένα πλήθος πιθανών υπολογισμών για τα στοιχεία που δεν έχουν κριθεί από το χρήστη.

4.4 Πιθανοτική Εκπροσώπηση

Για την αντιμετώπιση της προαναφερθείσας αστάθειας της ομοιότητας και της απόστασης των μέτρων προτείνουμε μια πιθανολογική αναπαράσταση με το ενδεχόμενο προφίλ του χρήστη, στην οποία οι αξιολογήσεις θα αντικαθίστανται από τις κατανομές πιθανοτήτων πάνω από την περιοχή των πιθανών βαθμολογιών. Δηλαδή, αντικαθιστούμε κάθε συνιστώσα p_i με την οριακή κατανομή πιθανότητας δηλαδή υπολογίζοντας τη μετρική:

$$Pr(p_i) \quad (p_i \in R \quad Pr(p_i = u) = 1).$$

Μπορούμε να υπολογίσουμε την αναμενόμενη απόσταση Manhattan (EMD) μεταξύ δύο προφίλ χρηστών:

$$\begin{aligned} d_{EMD}(p, q) &= E[\|p - q\|_1] \\ &= \sum_{i=1}^n \sum_{u_p \in R} \sum_{u_q \in R} |u_p - u_q| Pr(p_i = u_p) Pr(q_i = u_q) \end{aligned}$$

Υποθέτουμε εδώ ότι οι τυχαίες μεταβλητές αντιστοιχούν σε διαφορετικές αξιολογήσεις και είναι ανεξάρτητες μεταξύ τους.

$$Pr(p_i = u) = \begin{cases} 1 & \text{if } u = p \\ 0 & \text{if } u \neq p \end{cases}$$

Η κατανομή για τις διανυσματικές συνιστώσες που αντιστοιχούν σε άγνωστες αξιολογήσεις θα είναι διαφορετικές. Στη συνέχεια θα εξετάσουμε διάφορους τρόπους για να ορίσουμε μια κατανομή για αυτή την περίπτωση.

Κατανομή για Άγνωστες Βαθμολογίες

Ποια διανομή θα πρέπει να χρησιμοποιούμε για να υπολογίσουμε το EMD αν μια αξιολόγηση για ένα αντικείμενο είναι άγνωστη; Δύο δυνατότητες υπάρχουν:

- Ας υποθέσουμε ότι η βαθμολογία του στοιχείου ακολουθεί την ομοιόμορφη

κατανομή $Pr(p_i = u) = 1/|R|$, με $u \in R$. Σε αυτή την περίπτωση θα καλέσουμε την αναμενόμενη απόσταση Manhattan.

- Ας υποθέσουμε ότι η αξιολόγηση για το αντικείμενο i για j χρήστη είναι

άγνωστη. Υποθέτοντας ότι οι τιμές του χρήστη βαθμολογούνται παρόμοια με άλλους χρήστες μπορούμε να χρησιμοποιήσουμε την κατανομή όλων των γνωστών αξιολογήσεις για το συγκεκριμένο αντικείμενο i αντί της ομοιόμορφης κατανομής.

Αν και οι δύο χρήστες δεν βαθμολογήσουν το συγκεκριμένο αντικείμενο τότε, ο όρος στον τύπο d_{EMD} που αντιστοιχεί σε αυτό το αντικείμενο εξαρτάται μόνο από τον τομέα αξιολογήσεων R . Δεν εξαρτάται από τον χρήστη αλλά ούτε και από το αντικείμενο.

Για παράδειγμα εάν οι αξιολογήσεις έχουν εύρος τιμών $R\{1, \dots, N\}$ τότε αυτή η τιμή είναι μια σταθερά όπου: $c = r(r^2 - 1) = (3r^2)$.

Η τιμή αυτή μπορεί να θεωρηθεί ως ποινή που προστίθεται στην απόσταση Manhattan για κάθε στοιχείο που δεν έχει αξιολογηθεί από τους χρήστες.

Στην ουσία ακολουθεί την παρακάτω διαδικασία για τον υπολογισμό της απόστασης του μέτρου: υπολογίζουμε το Manhattan απόσταση πάνω από τα στοιχεία που είναι γνωστά για τους χρήστες και προστίθεται για κάθε στοιχείο που δεν έχει βαθμολογηθεί από τους χρήστες η ποινή c . Καλούμε αυτό το μέτρο απόσταση Manhattan απόσταση - ποινή ή MD - Ποινή

4.5 Πιθανοτικό γράφημα με βάση το φιλτράρισμα συνεργασίας

Στο γράφημα χρήστη Εκπροσωπούνται τα δεδομένα από ένα χρήστη $G = (V, E)$. Κάθε κόμβος v που ανήκει στο V αντιπροσωπεύει ένα χρήστη. Εάν η απόσταση d μεταξύ δύο χρηστών - ακμών μπορούν να υπολογιστούν, ένα υπολογισμένο σταθμισμένο βάρος των ακμών e που ανήκει στο E προστίθεται μέσα στο γράφο. Το βάρος w του e λόγω άκρου είναι αντιστρόφως ανάλογο προς την απόσταση μεταξύ των κόμβων που τα συνδέει έτσι ώστε $w_e(v_1, v_2) = 1/d(v_1, v_2)$.

Η γραφική παράσταση δεν θα περιέχει ακμές μεταξύ των χρηστών χωρίς επαναξιολογούμενα αντικείμενα κατά τη χρήση του Pearson Correlation ή MD -κλίμακα όμως όταν χρησιμοποιείτε EMD ή MD - ποινής, η διαδικασία θα έχει ως αποτέλεσμα ένα πλήρως συνδεδεμένο γράφημα. Για να χρησιμοποιηθεί το γράφημα σε μια συνεργατική μέθοδο φιλτραρίσματος, θα πρέπει να συμπληρωθούν τα στοιχεία σε κάθε κόμβο, με βάση τις πληροφορίες που υπάρχουν στους συνδεδεμένους κόμβους.

4.6 Γράφημα Κοντινότερου γείτονα-Αλγόριθμος_1-KNN-USERS

Ας υποθέσουμε ότι έχουμε ένα κόμβο $v = (3, -1, 0, ?)$. Στόχος μας, στη συνέχεια, είναι να προβλέψουμε τις ελλειπείς τιμές $V3$ και $V4$. Επομένως θα χρησιμοποιήσουμε το γράφημα για να δημιουργήσουμε τις προβλέψεις. Δεδομένου ότι το υψηλότερο άκρο βάρους σημαίνει υψηλότερη ομοιότητα, όπου αυτό προμυνεί το κοντινότερο γείτονα, οι τιμές που λείπουν είναι πιο πιθανό να είναι ίσες με τις αντίστοιχες τιμές των συνδεδεμένων κόμβων στις οποίες το βάρος της ακμής σε σχέση με την συνδεδεμένη ακμή είναι υψηλό. Μια τιμή που λείπει μπορεί στη συνέχεια να προβλεφθεί ως εξής: να επιλέξουμε k κόμβους που συνδέονται με τα υψηλότερα βάρη των ακμών, για τους οποίους η αξία που πρέπει να προβλεφθεί είναι γνωστή, και να λάβει το μέσο όρο αυτών των γνωστών τιμών. Αυτό δηλώνει την προσέγγιση των k πλησιέστερων γειτόνων.

Για τις ανάγκες της παρούσας διπλωματικής εργασίας αναπτύχθηκε και υλοποιήθηκε η παρακάτω συνάρτηση η οποία σκοπό έχει να προβλέψει τις τιμές που λείπουν από τον πίνακα A .

```

procedure kNN(A,E,w, k)
1: for each (i; j) such that  $A_{i,j} = ?$  do
2:    $N := kNN\text{-Users}(i; j, A,E,w, k)$ 
3:    $Ratings := \{A_{x,j} \mid x \in N\}$ 
4:    $A_{i,j} := \text{Most frequent value in Ratings}$ 
procedure kNN-Users(I, j, A,E,w, k)
1:  $N := \{(x,wx) \mid (i, x) \in E \wedge wx = w(i; x) \wedge A_{x,j} \neq ?\}$ 
2: return Top-k users from N sorted by wx

```

Όπου E είναι το σύνολο των ακμών και όπου k ο αριθμός των πλησιέστερων γειτόνων. Η ενέργειά της είναι η εξής για κάθε στοιχείο που βρίσκει αβαθμολόγητο (γρ.1) δηλαδή για κάθε χρήστη i που δεν έχει βαθμολογήσει την j ταινία, και όπου υπάρχει ακμή (γρ.3) ψάχνουμε να βρούμε την μεγαλύτερη συχνότητα στις αξιολογήσεις των γειτόνων. Από την συνάρτηση στον πίνακα N επιστρέφονται δύο στήλες η μία είναι το X και η άλλη είναι το w_x που συμβολίζει το βάρος. Για τις τιμές αυτές λοιπόν θα ταξινομήσει με πρώτη προτίμηση την πρώτη στήλη του πίνακα N και μετά την δεύτερη στήλη των βαρών. Αυτή η διαδικασία θα μας επιστρέψει τον πίνακα $topkusers$. Ένα δείγμα της πειραματικής διαδικασίας φαίνεται παρακάτω ενώ όλα τα αποτελέσματα του πειράματος θα παρατεθούν στην ενότητα των αποτελεσμάτων.

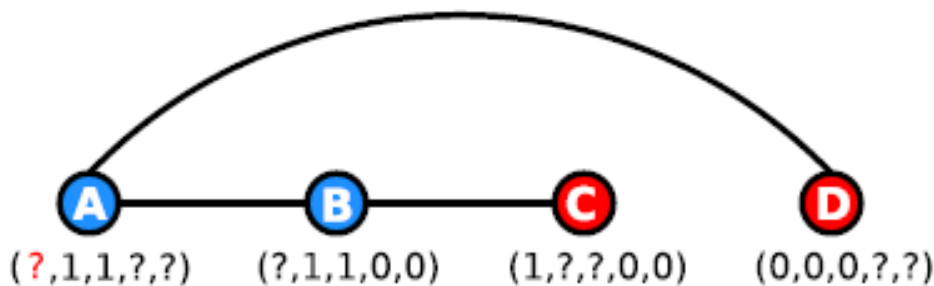
18	0,142857
58	0,166667
62	0,066667
76	0,111111
85	0,083333
92	0,1
128	0,071429
160	0,111111

ΕΙΚΟΝΑ Αποτελέσματα Πίνακα N

Ενώ αυτή η μέθοδος δίνει καλά αποτελέσματα στις περισσότερες περιπτώσεις υποφέρει από το λεγόμενο πρόβλημα της ψυχρής εκκίνησης. Χρήστες που βαθμολογούν μόνο μερικά στοιχεία δεν θα επιστρέψουν καλές αξιολογήσεις, καθώς και τα νέα στοιχεία, που έχουν βαθμολογηθεί από λίγους χρήστες, σπανίως συνιστώνται.

4.7 ΓΡΑΦΗΜΑ ΜΕ ΒΑΣΗ ΤΟ ΦΙΛΤΡΑΡΙΣΜΑ ΣΥΝΕΡΓΑΣΙΑΣ_ΑΛΓΟΡΙΘΜΟΣ 2

Για να διαχειριστούμε τα πρόβλημα της ψυχρής εκκίνησης και της χαμηλής πυκνότητας προτείνεται να χρησιμοποιηθούν κοντινοί, αλλά έμμεσοι γείτονες αντί άσχετο γείτονες με μη σχετικές βαθμολογίες που δεν δηλώνουν ίδια προτίμηση με τον χρήστη για ένα συγκεκριμένο αντικείμενο. Αυτό απεικονίζεται στις εικόνες 17 και 18 παρακάτω .



Στην εικόνα 17 στόχος είναι να προβλέφθει η βαθμολογία για τη θέση 1 του χρήστη A, το κόκκινο ερωτηματικό.Ο μόνος χρήστης που έχει βαθμολογήσει το πρώτο αντικείμενο όμως,στο σύνολο των χρηστών είναι ο χρήστης D.Συνεπώς το αποτέλεσμα της πρόβλεψης θα έπρεπε να είναι 0 αφού ο χρήστης D αξιολόγησε το αντικείμενο με 0.

Αν παρατηρήσουμε πιο προσεκτικά την βαθμολογία μεταξύ των χρηστών A και D , παρατηρούμε ότι έδωσαν εντελώς αντίθετες αξιολογήσεις ως προς το σύνολο της βαθμολογίας τους. Ο χρήστης A στα σημεία 2 και 3 ,βαθμολόγησε με 1 , ενώ ο D έδωσε την τιμή 0.Επομένως η πρόβλεψη που δημιουργήθηκε δεν ήταν αρκετά ακριβής.

Αν παρατηρήσουμε όμως τους χρήστες A και B βλέπουμε ότι έχουν κοινούς δείκτες μεταξύ τους και ότι οι αξιολογήσεις τους είναι πανομοιότυπες. Με αυτό τον τρόπο λοιπόν θα μπορούσαμε να υποθέσουμε ότι και οι δύο θα αξιολογούσαν και άλλα στοιχεία με όμοιες τιμές.Το ίδιο ισχύει και για το ζεύγος B και C. Μπορούμε έτσι να υποστηρίξουμε επίσης ότι ο B θα αξιολογούσε στοιχείο 1 με βαθμό 1επειδή ο χρήστης C είναι γείτονάς του με όμοιες τις δύο αξιολογήσεις του.Ετσι εάν ο B βαθμολογούσε το στοιχείο με 1 κατά συνέπεια και ο A που είναι γείτονας του B θα βαθμολογούσε κι αυτός το πρώτο στοιχείο με 1.

Συμπερασματικά με αυτό τον τρόπο μπορούμε σε μεγάλο βαθμό αποφύγουμε το πρόβλημα του αραιού πίνακα και να αποκτήσουμε μια πιο ακριβή πρόβλεψη.Αυτή η μέθοδος αποτελεί την βάση για συτήματα με βάση το φιλτραρισμα.

Ένα πρόβλημα για την προσέγγισή μας είναι το εξής, κάνοντας προβλέψεις και θεωρώντας κάποιες τιμές γνωστές ,οι ίδιες οι τιμές είναι επιρρεπής σε απόκλίσεις.Τα λάθη πολλαπλασιάζονται και αρχίζουν να συσσωρεύονται. Για να διορθωθεί αυτό το έχουμε

πρέπει να ληφθούν υπόψιν δύο μέτρα όπου θα αξιοποιηθούν στο έπακρο ορισμένες πρόβλεψη RST και θα λάβουν την αβεβαιότητα των προηγούμενων προβλέψεων υπόψη κατά τη χρήση τους για άλλους προβλέψεις.Στόχος μας είναι να κάνουμε προβλέψεις με βεβαιότητα. Ας υποθέσουμε ότι κάνουμε μια πρόβλεψη με βάση k γείτονες και όλοι αυτοί οι γείτονες βαθμολόγησαν το αντικείμενο για το οποίο θέλουμε να δημιουργήσουμε την πρόβλεψη ακριβώς το ίδιο. Αυτό θα μας δώσει μια πρόβλεψη με μέγιστη εμπιστοσύνη εάν από την άλλη πλευρά όλοι οι γείτονες αξιολογήσουν το στοιχείο διαφορετικά τότε δεν θα μπορεί να υπάρξει έμπιστη πρόβλεψη. Η διακύμανση του γείτονα – αξιολόγηση χρησιμοποιείται για να γίνει η πρόβλεψη, μας δίνει έτσι ένα μέτρο εμπιστοσύνης της πρόβλεψης. Μια χαμηλή διακύμανση μας λέει ότι η πρόβλεψη είναι πιθανότατα σωστή . Μια υψηλή διακύμανση μας λέει το αντίθετο για αυτή την πρόβλεψη. Σε αυτό το σημείο μπορούμε να χρησιμοποιήσουμε τη διακύμανση με την οποία θα συμπληρώσουμε τις προβλέψεις στο γράφημα .

Ας υποθέσουμε ότι οι χρήστες μας μπορούν να αξιολογήσουν μόνο ένα στοιχείο θετικά ή αρνητικά και θέλουμε να κάνουμε μια πρόβλεψη για ένα αντικείμενο του χρήστη A βασισμένοι σε τρεις κοντινούς του χρήστες. Δύο από τους τρεις γείτονες βαθμολόγησαν το στοιχείο θετικά και ένας γείτονας έδωσε το στοιχείο α αρνητική βαθμολογία. Σε αυτή την περίπτωση, θα προβλέψουμε ότι χρήστης A θα βαθμολογούσε το στοιχείο θετικά, αλλά αυτή η πρόβλεψη είναι ανακριβής αφού μόνο τα 2/3 των γειτόνων έδωσαν μια θετική αξιολόγηση. Στον αλγόριθμο μας, οι προβλεπόμενες τιμές χρησιμοποιούνται για μετέπειτα προβλέψεις. Για παράδειγμα όταν κάνουμε μια

πρόβλεψη για τον χρήστη B, αυτός ο χρήστης μπορεί να έχει ως γείτονά του το χρήστη A. Αν οι γείτονες του B (εκτός από τον A) όλοι βαθμολόγησαν το στοιχείο θετικά, τότε και η πρόβλεψη μας για τον A θα ήταν θετική. Η αντιστοίχιση της προβλεπόμενης τιμής είναι τώρα λίγο πιο δύσκολη.

Αν θεωρήσουμε την προβλεπόμενη τιμή για το A ως μία γνωστή τιμή, το σύνολο των γειτόνων του B θα έχουν αξιολογήσει το στοιχείο θετικά και η προκύπτουσα τιμή θα είναι το 1. Κάτι τέτοιο δεν λαμβάνει υπόψιν, το γεγονός δηλαδή ότι δεν είμαστε εντελώς σίγουροι για την πρόβλεψη μας για το χρήστη A.

Ως λύση σε αυτό το πρόβλημα, προτείνεται να αποθηκεύετε η κατανομή των αξιολογήσεων μεταξύ των γειτόνων αντί της πρόβλεψης που προκύπτει από αυτή την κατανομή. Δηλαδή για το πάνω παράδειγμα θα έπρεπε να αποθηκεύσουμε την κλασματική μετρική για την κάθε συχνότητα αξιολόγησης ξεχωριστά. Επί παραδείγματι για τον χρήστη A θα έπρεπε να αποθηκεύσουμε (+2/3, -1/3). Έτσι όταν θα γίνεται η πρόβλεψη για τον χρήστη B, θα χρησιμοποιούμε αυτή την κατανομή αντί της πρόβλεψης. Αυτό επιτυγχάνεται με διάσπαση της αξιολόγησης του A στην πρόβλεψη του B, σύμφωνα με την κατανομή. Το γεγονός ότι δεν είμαστε εντελώς σίγουροι για την πρόβλεψή μας για τον A έχει πια ληφθεί υπόψιν, όταν κάνει την πρόβλεψη για τον B. Καθώς ο αλγόριθμος μας λειτουργεί σε μη-δυναδικά δεδομένα, η μέθοδος που περιγράφεται ανωτέρω επεκτείνεται σε διακριτές κατανομές. Δεδομένου ότι η βαθμολογία μπορεί να είναι το αποτέλεσμα μιας πρόβλεψης, είναι δυνατόν να αναπαρασταθεί ως μια κατανομή όπως άλλοστε ζητάται και στον ψευδοκώδικα του αλγορίθμου 2.

```

procedure PGBCF(A, E, w, k, R)
1: { Find prediction order }
2: Todo := ∅
3: for each (i, j) such that Ai,j = ? do
4:   N := kNN-Users(i, j, A, E, w, k)
5:   v := Variance( {Ax,j | x ∈ N} )
6:   Todo := Todo ∪ { (i, j, |N|, v) }
7: Sort Todo by decreasing |N| and increasing v
8: { Initialize distributions for known ratings }
9: for each (i, j) such that Ai,j ≠ ? do
10:  for each r ∈ R do
11:    Pr(Ai,j = r) = { 1 if r = Ai,j
                     0 if r ≠ Ai,j
12: { Propagate (predicted) distributions }
13: for each (i, j, ·, ·) ∈ Todo do
14:   N := kNN-Users(i, j, A, E, w, k)
15:   Pr(Ai,j) := 1/|N| ∑x ∈ N Pr(Ax,j)
16:   Ai,j := Most frequent value Pr(Ai,j)

```

Έχοντας πάρει την ακόλουθη κατανομή $Pr = (0, 1/5, 2/5, 1/5, 1/5)$, η D κατανομή, με βάση τις κατανομές Pr^i των ν γειτόνων υπολογίζεται στη συνέχεια ως ακολούθως:

$$Pr_j = \sum_{i=1..n} D_j^i / n$$

Η τελική πρόβλεψη δημιουργείται βάση του υπολογισμού της μέσης τιμής της κατανομής αυτής. Για να συγκρίνουμε κατά πόσο είναι αξιόπιστο το αποτέλεσμα που έχουμε στη πρόβλεψη μπορούμε να χρησιμοποιήσουμε την διακύμανση της κατανομής. Όσο χαμηλότερη είναι η διακύμανση τόσο πιο αξιόπιστη είναι η πρόβλεψή μας.

4.8 Αποτελέσματα

Έχουμε υλοποιήσει δύο αλγόριθμους με σκοπό την εύρεση της εκτιμώμενης τιμής για ένα σύνολο στοιχείων ενός αραιού πίνακα. Επίσης έχουμε υπολογίσει το μέσο απόλυτο σφάλμα μεταξύ αρχικών και τελικών τιμών το οποίο μας βγήκε 2.7 και $K=4$ γείτονες, έτσι ώστε να επαληθεύεται η πειραματική διαδικασία. Παρακάτω παραθέτονται τα αποτελέσματα των πινάκων τοδο που πήραμε από τον αλγόριθμο 2 ενώ οι υπόλοιποι πίνακες θα παρατεθούν σε αρχεία .xlsx.

18	0,142857
58	0,166667
62	0,066667
76	0,111111
85	0,083333
92	0,1
128	0,071429
160	0,111111
201	0,066667
244	0,090909
262	0,142857
268	0,090909
286	0,166667
293	0,111111
299	0,076923
334	0,066667
344	0,1
345	0,055556
354	0,166667
360	0,25
389	0,333333
405	0,5
416	0,029412
437	0,111111
456	0,142857
468	0,142857
487	0,052632
524	0,142857
535	0,142857
537	0,035714
551	0,045455
561	0,142857
618	0,2
642	0,0625

645	0,333333
655	0,047619
711	1
712	1
715	0,2
758	0,05
762	0,125
763	0,5
828	0,111111
846	0,333333
859	0,5
871	0,071429
877	0,111111
883	0,066667
889	0,0625

Πίνακας N

4.9 ΣΥΜΠΕΡΑΣΜΑΤΑ

Στόχος της παρούσας διπλωματικής ήταν η αντιμετώπιση ενός αραιού πίνακα καθώς και το πρόβλημα της ψυχρής εκκίνησης. Προτείναμε αρχικά μια σειρά μεθοδολογιών όπως η απόσταση Manhattan και η MD-Penalty. Στη συνέχεια, χρησιμοποιήσαμε την αναπαράσταση ενός γραφήματος ενός αλγορίθμου φιλτραρίσματος, υλοποιήσαμε την διαδικασία του αλγορίθμου 1 καθώς και το κύριο πρόγραμμα τον αλγόριθμο 2. Τα πειράματα ότι με τις μετρικές που προαναφέρθηκαν μπορούμε να προβλέψουμε αξιολογήσεις για τον σποραδικό πίνακα με σχετική ακρίβεια.

Πανεπιστήμιο Πειραιώς

Βιβλιογραφία

- Adomavicius, G., & Kwon, Y. (2007). *New Recommendation Techniques for Multicriteria Rating Systems*. IEEE.
- Adomavicius, G., & Tuzhilin, A. (2005, June). *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*. Ανάκτηση 2014, από <http://people.stern.nyu.edu/http://people.stern.nyu.edu/atuzhili/pdf/TKDE-Paper-as-Printed.pdf>
- Alspector, J., Kolcz, A., & Karunanithi, N. (1997). *Comparing Feature-based and Clique-based User Models for Movie Selection*. Ανάκτηση 2014, από <http://sydney.edu/http://sydney.edu.au/engineering/it/~irena/alspector98.pdf>
- Balabanovic, M. (1998). *Exploring versus Exploiting when Learning User Models for Text Representation*. Stanford University.
- Balabanovic, M., & Shoham, Y. (1997). Ανάκτηση 20014, από https://www.ischool.utexas.edu/https://www.ischool.utexas.edu/~i385q/readings/Balabanovic_Shoham-1997-Fab.pdf
- Balabanovic, M., & Shoham, Y. (1997). *Content-Based, Collaborative Recommendation*.
- Basu, C., Hirsh, H., & Cohen, W. (1998). *Recommendation as Classification: Using Social and Content-Based Information in Recommendation*. Ανάκτηση 2014, από <http://www.aaai.org/http://www.aaai.org/Papers/Workshops/1998/WS-98-08/WS98-08-002.pdf>
- Breese, J., Heckerman, D., & Kadie, C. (1998). *Empirical Analysis Of Predictive Algorithms For Collaborative Filtering*. Ανάκτηση 2014, από <http://research.microsoft.com/http://research.microsoft.com/pubs/69656/tr-98-12.pdf>
- Breese, J., Heckerman, D., & Kadie, C. (1998). *Empirical Analysis Of Predictive Algorithms For Collaborative Filtering*. San Francisco.
- Burke, R. (2000). *Hybrid Recommender Systems: Survey and Experiments*. Ανάκτηση 2014, από <http://josquin.cs.depaul.edu/http://josquin.cs.depaul.edu/~rburke/pubs/burke-umuai02.pdf>
- Burke, R. (2000). *Knowledge-based recommender systems*. Ανάκτηση 2014, από <http://josquin.cti.depaul.edu/http://josquin.cti.depaul.edu/~rburke/pubs/burke-elis00.pdf>
- Burke, R. (1999). *The Wasabi Personal Shopper: A Case-Based Recommender System*. Ανάκτηση 2014, από <http://www.aaai.org/http://www.aaai.org/Papers/IAAI/1999/IAAI99-119.pdf>
- Burke, R., Hammond, K., & Young, B. (1997). *The FindMe Approach to Assisted Browsing*. Ανάκτηση 2014, από http://pdf.aminer.org/http://pdf.aminer.org/000/874/501/the_findme_approach_to_assisted_browsing.pdf
- Claypool, M., Cokhale, A., & Miranda, T. (1999). *Combining Content-Based And Collaborative Filters*. Ανάκτηση 2014, από <http://web.cs.wpi.edu/http://web.cs.wpi.edu/~claypool/papers/content-collab/content-collab.pdf>
- ERCIM. (1997). *Filtering And Collaborative Filtering*. Ανάκτηση 2014, από <http://www.ercim.eu/publication/ws-proceedings/DELOS5/delos5.pdf>
- Herlocker, J. (2004). *Evaluating Collaborative Filtering Recommender Systems*. New York.
- Jennings, N., & Higuchi, H. (1993). *A User Model Neural Network for a Personal News Service*. Kluwer Academic.
- Malone, T., Grant, K., Turbak, F., Brobst, S., & Cohen, M. (1987). *Intelligent Information-Sharing Systems*. Communications Of The ACM, New York.
- Meyer, F., & Boulle, M. (2007). *Comparing State-of-the-Art Collaborative Filtering Systems*.

- Papagelis, M. (2005). *Crawling The Algorithmic Foundations Of Recommendation Technologies*. Ηράκλειο.
- Pazzani, M. (1999). *A Framework For Collaborative, Content-Based And Demographic Filtering*. Kluwer Academic.
- Pazzani, M., & Billsus, D. (2007). *Content-based Recommendation Systems*. Ανάκτηση 2014, από <http://www.fxpall.com>: <http://www.fxpall.com/publications/FXPAL-PR-06-383.pdf>
- Rashid, A., Albrt, I., Cosley, D., & Lam, S. (n.d.). *Getting to Know You: Learning New User Preferences in Recommender Systems*. Ανάκτηση 2014, από <http://files.grouplens.org/>: <http://files.grouplens.org/papers/voi-final.pdf>
- Reischach, F., Michahelles, F., & Schmidt, A. (n.d.). *The Design Space of Ubiquitous Product Recommendation Systems*. Ανάκτηση 2014, από <http://www.im.ethz.ch>: http://www.im.ethz.ch/publications/MUM_FvR.pdf
- Resnick, P., & Varian, H. (1997). Recommender systems. *Communications of the ACM*. ACM.
- Resnick, P., & Varian, R. (1997). *Recommender Systems*. Ανάκτηση 2014, από <http://www.inf.unibz.it>: <http://www.inf.unibz.it/~ricci/ISR/papers/resnick-varian97.pdf>
- Sachan, A. *A Survey on Recommender Systems Based on Collaborative Systems*.
- Sarwar, B., Karypis, G., & Riedl, J. (2000). *Analysis Of Recommendation Algorithms For E-Commerce*. New York.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). *em-Based Collaborative Filtering Recommendation Algorithms*. Ανάκτηση 2014, από <http://files.grouplens.org/>: http://files.grouplens.org/papers/www10_sarwar.pdf
- Schein, A., Popescul, A., Ungar, L., & Pennock, D. (2002). *Methods And Metrics For Cold-Start Recommendations*. New York.
- Soboroff, I. (1999). *Combining Content And Collaboration In Text Filtering*. Ανάκτηση 2014, από <http://www.csee.umbc.edu>: <http://www.csee.umbc.edu/csee/research/cadip/1999Symposium/mlif.pdf>
- Towle, B., & Quinn, C. (2000). *Knowledge Based Recommender Systems Using Explicit User Models*. Ανάκτηση 2014, από <http://www.aaai.org>: <http://www.aaai.org/Papers/Workshops/2000/WS-00-04/WS00-04-011.pdf>
- Tran, T., & Cohen, R. (2000). *Hybrid Recommender Systems For Electronic Commerce*. Ανάκτηση 2014, από <http://aaai.org>: <http://aaai.org/Papers/Workshops/2000/WS-00-04/WS00-04-012.pdf>
- Xiaoyuan, S., & Taghi, K. (2009). *A Survey of Collaborative Filtering Techniques*. Florida: Florida Atlantic University.
- Zanker, M. (2010). *Introduction to Recommender Systems*. Sierre: University Klagenfurt.
- Λίκα, Μ. Αυτόματη Κατηγοριοποίηση Χρηστών ως Μέθοδος Επίλυσης του Προβλήματος Ψυχρής Εκκίνησης.
- Μουζακίδης, Α. ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ: Υβριδική πεσοαρμοστική μέθοδος πρόβλεψης μεταβαλλόμενων προτιμήσεων χρηστών.
- Adomavicius, G., & Kwon, Y. (2007). *New Recommendation Techniques for Multicriteria Rating Systems*. IEEE.
- Adomavicius, G., & Tuzhilin, A. (2005, June). *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*. Ανάκτηση 2014, από <http://people.stern.nyu.edu>: <http://people.stern.nyu.edu/atuzhili/pdf/TKDE-Paper-as-Printed.pdf>
- Alspector, J., Kolcz, A., & Karunanithi, N. (1997). *Comparing Feature-based and Clique-based User Models for Movie Selection*. Ανάκτηση 2014, από <http://sydney.edu>: <http://sydney.edu.au/engineering/it/~irena/alspector98.pdf>
- Balabanovic, M. (1998). *Exploring versus Exploiting when Learning User Models for Text Representation*. Stanford University.
- Balabanovic, M., & Shoham, Y. (1997). Ανάκτηση 20014, από <https://www.ischool.utexas.edu/>: https://www.ischool.utexas.edu/~i385q/readings/Balabanovic_Shoham-1997-Fab.pdf
- Balabanovic, M., & Shoham, Y. (1997). *Content-Based, Collaborative Recommendation*.

- Basu, C., Hirsh, H., & Cohen, W. (1998). *Recommendation as Classification: Using Social and Content-Based Information in Recommendation*. Ανάκτηση 2014, από <http://www.aaai.org/Papers/Workshops/1998/WS-98-08/WS98-08-002.pdf>
- Breese, J., Heckerman, D., & Kadie, C. (1998). *Empirical Analysis Of Predictive Algorithms For Collaborative Filtering*. Ανάκτηση 2014, από [http://research.microsoft.com/pubs/69656/tr-98-12.pdf](http://research.microsoft.com/http://research.microsoft.com/pubs/69656/tr-98-12.pdf)
- Breese, J., Heckerman, D., & Kadie, C. (1998). *Empirical Analysis Of Predictive Algorithms For Collaborative Filtering*. San Francisco.
- Burke, R. (2000). *Hybrid Recommender Systems: Survey and Experiments*. Ανάκτηση 2014, από <http://josquin.cs.depaul.edu>: <http://josquin.cs.depaul.edu/~rburke/pubs/burke-umuai02.pdf>
- Burke, R. (2000). *Knowledge-based recommender systems*. Ανάκτηση 2014, από <http://josquin.cti.depaul.edu>: <http://josquin.cti.depaul.edu/~rburke/pubs/burke-elis00.pdf>
- Burke, R. (1999). *The Wasabi Personal Shopper: A Case-Based Recommender System*. Ανάκτηση 2014, από <http://www.aaai.org>: <http://www.aaai.org/Papers/IAAI/1999/IAAI99-119.pdf>
- Burke, R., Hammond, K., & Young, B. (1997). *The FindMe Approach to Assisted Browsing*. Ανάκτηση 2014, από <http://pdf.aminer.org>: http://pdf.aminer.org/000/874/501/the_findme_approach_to_assisted_browsing.pdf
- Claypool, M., Cokhale, A., & Miranda, T. (1999). *Combining Content-Based And Collaborative Filters*. Ανάκτηση 2014, από <http://web.cs.wpi.edu>: <http://web.cs.wpi.edu/~claypool/papers/content-collab/content-collab.pdf>
- ERCIM. (1997). *Filtering And Collaborative Filtering*. Ανάκτηση 2014, από <http://www.ercim.eu/publication/ws-proceedings/DELOS5/delos5.pdf>
- Herlocker, J. (2004). *Evaluating Collaborative Filtering Recommender Systems*. New York.
- Jennings, N., & Higuchi, H. (1993). *A User Model Neural Network for a Personal News Service*. Kluwer Academic.
- Malone, T., Grant, K., Turbak, F., Brobst, S., & Cohen, M. (1987). *Intelligent Information-Sharing Systems*. Communications Of The ACM, . New York.
- Meyer, F., & Boulle, M. (2007). *Comparing State-of-the-Art Collaborative Filtering Systems*.
- Papagelis, M. (2005). *Crawling The Algorithmic Foundations Of Recommendation Technologies*. Ηράκλειο.
- Pazzani, M. (1999). *A Framework For Collaborative, Content-Based And Demographic Filtering*. Kluwer Academic.
- Pazzani, M., & Billsus, D. (2007). *Content-based Recommendation Systems*. Ανάκτηση 2014, από <http://www.fxpal.com>: <http://www.fxpal.com/publications/FXPAL-PR-06-383.pdf>
- Rashid, A., Albrt, I., Cosley, D., & Lam, S. (n.d.). *Getting to Know You: Learning New User Preferences in Recommender Systems*. Ανάκτηση 2014, από <http://files.grouplens.org/>: <http://files.grouplens.org/papers/voi-final.pdf>
- Reischach, F., Michahelles, F., & Schmidt, A. (n.d.). *The Design Space of Ubiquitous Product Recommendation Systems*. Ανάκτηση 2014, από <http://www.im.ethz.ch>: http://www.im.ethz.ch/publications/MUM_FvR.pdf
- Resnick, P., & Varian, H. (1997). *Recommender systems*. Communications of the ACM. ACM.
- Resnick, P., & Varian, R. (1997). *Recommender Systems*. Ανάκτηση 2014, από <http://www.inf.unibz.it/>: <http://www.inf.unibz.it/~ricci/ISR/papers/resnick-varian97.pdf>
- Sachan, A. *A Survey on Recommender Systems Based on Collaborative Systems*.
- Sarwar, B., Karypis, G., & Riedl, J. (2000). *Analysis Of Recommendation Algorithms For E-Commerce*. New York.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). *em-Based Collaborative Filtering Recommendation Algorithms*. Ανάκτηση 2014, από <http://files.grouplens.org/>: http://files.grouplens.org/papers/www10_sarwar.pdf

Schein, A., Popescul, A., Ungar, L., & Pennock, D. (2002). *Methods And Metrics For Cold-Start Recommendations*. New York.

Soboroff, I. (1999). *Combining Content And Collaboration In Text Filtering*. Ανάκτηση 2014, από <http://www.csee.umbc.edu>: <http://www.csee.umbc.edu/csee/research/cadip/1999Symposium/mlif.pdf>

Towle, B., & Quinn, C. (2000). *Knowledge Based Recommender Systems Using Explicit User Models*. Ανάκτηση 2014, από <http://www.aaai.org>: <http://www.aaai.org/Papers/Workshops/2000/WS-00-04/WS00-04-011.pdf>

Tran, T., & Cohen, R. (2000). *Hybrid Recommender Systems For Electronic Commerce*. Ανάκτηση 2014, από <http://aaai.org>: <http://aaai.org/Papers/Workshops/2000/WS-00-04/WS00-04-012.pdf>

Xiaoyuan, S., & Taghi, K. (2009). *A Survey of Collaborative Filtering Techniques*. Florida: Florida Atlantic University.

Zanker, M. (2010). *Introduction to Recommender Systems*. Sierre: University Klagenfurt.

Λίκα, Μ. Αυτόματη Κατηγοριοποίηση Χρηστών ως Μέθοδος Επίλυσης του Προβλήματος Ψυχρής Εκκίνησης.

Μουζακίδης, Α. ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ: Υβριδική πεσοαρμωστική μέθοδος πρόβλεψης μεταβαλλόμενων προτιμήσεων χρηστών.

Alexandros Nanopoulos. *Collaborative Filtering based on transitive correlations between items*. *Advances in Information Retrieval, 2007*.

Zan Huang, Wingyan Chung, Thian-Huat Ong, and Hsinchun Chen. *A graph-based recommender system for digital library*. In *JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 65-73, New York, NY, USA, 2002. ACM.

Παράρτημα

A. Αλγόριθμος_2

```
% Reads the ./h.mat file of the eachmovie dataset into
% a sparse matrix, whose rows indicate users and columns
% indicate movies.
%
% Thesis, Algorithm 2
%
% Baltzi Vicky, June 2014

clear all
close all
%path(path, '..\..\MATLAB')

%loading the data
load d.mat;
mat=[];
Users=948;
Movies=1682;
% E is the set of edges defining the user graph,
% w is the edge weight function,
% k is the number of neighbors to consider,
% R is the rating domain,
% PGBCF predicts the missing ratings in the ratings in the matrix A.
kmax=4; % number of Top-k users

A=-ones(Users,Movies);

for l=1:100000,
    A(data(l,1),data(l,2))=data(l,3);
end;
[Users,Movies]=size(A)
for i=1:Users,
    % g(i)=0;
    [r,g(i)]=size(find(A(i,:)~= -1));
    %for j=1:Movies,
    %     if A(i,j)~= -1,
    %         g(i)=g(i)+1;
    %     end;
%end;

end;
hist(g);
DG=sparse(data(:,1),data(:,2),true,943,1682);
% Create user graph
%view(biograph(DG));
```

ΓΡΑΦΟΘΕΩΡΗΤΙΚΕΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΗΣ

```

for i=1:Users,
    for l=1:Users,
        S=0;
        k=0;
        for j=1:Movies,
            if ((A(i,j)~= -1) & (A(l,j)~= -1)),
                S=S+abs(A(i,j)-A(l,j));
                k=1;
            end;
        end;
        if (S~=0),
            w(i,l)=1/S;
            E(i,l)=1;
        else
            if (k==0),
                w(i,l)=0;
                E(i,l)=0;
            else
                w(i,l)=1;
                E(i,l)=1;
            end;
        end;
    end;
end;

% wx=w(i,:);
% [Ix,Sx]=sort(wx,'descend');
% for i=1:k,
%     result(i)=Sx(1,i);
% end;
% result
% Algorithm 2
Todo=[];
z=0;
for i=1:Users,
    for j=1:Movies,
        if A(i,j)~= -1,
            z=z+1;
            N=kNN_Users(i,j,A,E,w,kmax);

            mn=mean(N(:,2));
            u=sqrt(sum(pow2(N(:,2)-mn)));           % u = Variance calculation
            metro_dian=sqrt(sum(pow2(N(:,2)))));    % vector measure
            Todo(z,1)=i;
            Todo(z,2)=j;
        end;
    end;
end;

```

```

        Todo(z,3)=metro_dian;
        Todo(z,4)=u;
    end;
end;
end;

for ir=2:z,    % Sort Todo by decreasing |N| and increasing u
    for jr=z:-1:ir,
        if Todo(jr-1,3)<Todo(jr,3),
            for jk=1:4,
                temp=Todo(jr-1,jk);
                Todo(jr-1,jk)=Todo(jr,jk);
                Todo(jr,jk)=temp;
            end;
        elseif (Todo(jr-1,3)==Todo(jr,3)) & (Todo(jr-1,4)>Todo(jr,4)),
            for jk=1:4,
                temp=Todo(jr-1,jk);
                Todo(jr-1,jk)=Todo(jr,jk);
                Todo(jr,jk)=temp;
            end;
        end;
    end;
end;

end;
end;

R=[1,2,3,4,5];

for i=1:Users,
    for j=1:Movies,
        if A(i,j)~= -1,
            for ct=1:5;
                if R(ct)==A(i,j),
                    Pr(i,j)=1;
                else
                    Pr(i,j)=0;
                end;
            end;
        end;
    end;
end;
end;

for i=1:Users,
    for j=1:Movies,
        if (Todo(i,1)==i) & (Todo(i,2)==j),
            N=kNN_Users(i,j,A,E,w,kmax);

```

```

metro_dian=sqrt (sum (pow2 (N (:,2)))) ; % vector measure
s=0;
for ct=1:size (N (:,2)),
    s=s+Pr (N (ct,1),j);
end;
Pr (i,j)=(1/metro_dian)*s;

k=0; % number of Top-k users of N
for f=1:size (N,1),
    if N (f,1)~=0,
        k=k+1; % number of Top-k users of N
    end;
end;
if k>2 & k>size (N,1),
    k=kmax; % maximum number of Top-k users is defined to kmax
end;

R=zeros (k,1);
count1=zeros (5,1);
for ct=1:k,
    R (ct)=Pr (N (ct,1),j);
end;
%R=sort (R);
for ct=1:k,
    if (R (ct)~=0),
        count1 (R (ct))=count1 (R (ct))+1;
    end;
end;
[Y,Ikt]=max (count1);
A (i,j)=Ikt;

%Variance of A(x,j)

end;
end;
end;
B=-ones (Users,Movies);

for l=1:100000,
    B (data (l,1),data (l,2))=data (l,3); % B the initial array without prediction
end;

% A is the prediction matrix
MAE=sum (sum (A (:,:) -B (:,:)))/(Users*Movies); % MAE = mean absolute error

% End of Algorithm 2

```

B.Συνάρτηση knn-users

```
function [N]=kNN_Users(i,j,A,E,w,k)
% kNN predicts the missing ratings in the ratings in a matrix A.
% E is the set of edges defining the user graph,
% w is the edge weight function,
% k is the number of neighbors to consider.
% checking the parameters given

Users=943;
ct=0;
for x=1:Users, % kNN-Users
    if (A(x,j)~= -1 && E(i,x)==1),
        ct=ct+1;
        N(ct,1)=x;
        N(ct,2)=w(i,x);
    end;
end;
if ct==0,
    N(1,1)=0;
    N(1,2)=0;
else
    sort(N,1,'descend');
    sort(N,2,'descend');
%     for ir=2:ct,
%         for jr=ct:-1:ir,
%             if N(jr-1,2)<N(jr,2),
%                 for jk=1:2,
%                     temp=N(jr-1,jk);
%                     N(jr-1,jk)=N(jr,jk);
%                     N(jr,jk)=temp;
%                 end;
%             end;
%         end;
%     end;
end;
end;
```