



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΥΠΟΛΟΓΙΣΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Αλγοριθμικές Τεχνικές και Εφαρμογές για Συστήματα Συστάσεων

Μπάρδης Σταμάτιος

Εισηγήτρια: Αναστασία Ν. Βελώνη Λέκτορας Εφαρμογών

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή

Τόπος:

.....

Ημερομηνία:

.....

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1.
2.
3.

Περιεχόμενα

Κατάλογος εικόνων.....	7
Περίληψη.....	9
1. Τι είναι τα συστήματα συστάσεων.....	11
1.1. Εισαγωγή.....	11
1.2. Ορισμός.....	14
2. Διαδικασίες και Μορφές Σύστασης στα ΣΣ.....	18
2.1. Διαδικασία Σύστασης.....	18
2.2. Μορφές της διαδικασίας σύστασης.....	24
2.2.1. Ένας προς έναν.....	24
2.2.2. Πολλοί προς πολλούς.....	24
2.2.3. Πολλοί προς έναν.....	25
2.2.4. Ένας προς πολλούς.....	25
3. Αλγόριθμοι και Τεχνικές Συστημάτων Συστάσεων.....	27
3.1. Αξιολόγηση Συστημάτων Συστάσεων.....	27
3.2. Τεχνικές Συστημάτων Συστάσεων.....	29
3.2.1. Συστήματα βασισμένα στο Περιεχόμενο.....	29
3.2.2. Συστήματα Βασισμένα στην Συνεργασία.....	32
3.2.3. Συστήματα Προσανατολισμένα Στην Μνήμη (Memory Based).....	35
3.2.4. Συστήματα Προσανατολισμένα Στο Μοντέλο (Model Based systems).....	36
3.2.5. Συστήματα βασισμένα σε Δημογραφικά δεδομένα (Demographic Based)....	43
3.2.6. Συστήματα Βασισμένα Στην γνώση (Knowledge- Based Systems).	44
3.2.7. Συστήματα Βασισμένα Στην Υπόθεση (Case- Based System).....	45
3.2.8. Συστήματα Βασισμένα Στον Στόχο (Goal-Based Systems).....	46
3.2.9. Υβριδικά Συστήματα (Hybrid Systems).....	51
4. Εφαρμοσμένος αλγόριθμος Συστήματος Συστάσεων.....	52
4.1. Παρουσίαση αλγορίθμου.....	52
4.2. Δομή Εφαρμογής.....	55
4.3. Παρουσίαση αποτελεσμάτων.....	56
5. Συμπεράσματα και προοπτικές.....	57
Βιβλιογραφία.....	58

Κατάλογος εικόνων

Εικόνα 3 1 Διαδικασία που ακολουθείται στα συστήματα βασισμένα στη συνεργασία	32
Εικόνα 3 2 Παράδειγμα μιας μήτρας από βαθμολογίες χρηστών για ένα αντικείμενο	33
Εικόνα 3 3 Κατηγοριοποίηση και σύσταση των βαθμολογούμενων ταινιών βάση του φύλλου.....	33
Εικόνα 3 4 Περιγραφή του τρόπου λειτουργίας των συστημάτων βασισμένα στην γνώση.....	45
Εικόνα 3 5 Αρχική σελίδα συστήματος Entree.....	48
Εικόνα 3 6 Knowledge-based αποτελέσματα συστήματος Entree.....	49
Εικόνα 4 1 Αξιολογήσεις των χρηστών για κάθε ένα είδος ταινίας.....	54

Περίληψη

Τα συστήματα συστάσεων είναι εφαρμογές λογισμικού που παρέχουν εξατομικευμένες προτάσεις στους χρήστες σχετικά με προϊόντα ή υπηρεσίες που μπορεί να τους ενδιαφέρουν. Προτείνουν στοιχεία σχετικά με τα ενδιαφέροντα των χρηστών, με βάση τις προτιμήσεις τους οι οποίες εκφράζονται άμεσα (explicitly), είτε έμμεσα (implicitly). Τα συστήματα αυτά με τη βοήθεια ειδικών αλγορίθμων επιχειρούν να προβλέψουν ποιες υπηρεσίες είναι πιθανόν να ενδιαφέρουν περισσότερο τον χρήστη.

Η δημιουργία προφίλ αποτελεί το πλέον σημαντικό στοιχείο ενός συστήματος συστάσεων. Τα συστήματα συστάσεων εκμεταλλεύονται τις ιδιαιτερότητες των χρηστών με σκοπό να διευκολύνουν στο να προσδιορίζουν ακριβέστερα τις πληροφορίες ή τα προϊόντα για τα οποία ενδιαφέρονται περισσότερο ή σχετίζονται με τις ανάγκες τους.

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 γίνεται ευκολότερα ο διαχωρισμός των περιπτώσεων από τις πληροφορίες που είναι χρήσιμες στον τελικό χρήστη. 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.

Στην παρούσα εργασία θα γίνει μια εκτενής μελέτη των συστημάτων συστάσεων καθώς επίσης και μια βιβλιογραφική ανασκόπηση αλγοριθμικών μεθόδων και εφαρμογών που έχουν αναπτυχθεί προς αυτή την κατεύθυνση.

1. Τι είναι τα συστήματα συστάσεων

1.1. Εισαγωγή

Κατά τα τελευταία έτη, η δραματική αύξηση στη χρήση του Διαδικτύου και η ταχύτητα με την οποία εξαπλώνεται ο τεράστιος όγκος πληροφορίας που προέρχεται από αυτό, έχουν οδηγήσει στο φαινόμενο της υπερπληροφόρησης (γνωστό και ως Information overload ή information explosion). Μέσα σε αυτόν τον κατατρεγισμό πληροφοριών οι χρήστες του Διαδικτύου δυσκολεύονται ολοένα και περισσότερο να ξεχωρίσουν το περιεχόμενο που τους ενδιαφέρει γρήγορα και αποτελεσματικά. Στην προσπάθεια δημιουργίας συστημάτων που θα επέλυαν αυτό το πρόβλημα, αναπτύχθηκαν τα συστήματα συστάσεων –ΣΣ (Recommender Systems). Τα συστήματα αυτά χρησιμοποιούνται σε διάφορους τομείς όπου είναι απαραίτητο το φιλτράρισμα της πληροφορίας. Λόγου χάρη, χρησιμοποιούνται σε εφαρμογές ψυχαγωγίας για προτάσεις ταινιών, μουσικής κλπ, σε διαδικτυακές υπηρεσίες για συστάσεις ταξιδιών και εστιατορίων, σε κοινωνικά μέσα για συστάσεις βίντεο, άρθρων και για on-line αγορές διαφόρων ειδών καθημερινής χρήσης π.χ. ενδυμασίας. Τα Recommender Systems είναι ιδιαίτερος σημαντικά σε εφαρμογές ηλεκτρονικού εμπορίου (Reischach et al., 2009).

Δεδομένου ότι η λήψη αποφάσεων των χρηστών για αγορά μέσα από μια μεγάλη ποσότητα διαθέσιμων προϊόντων στο ηλεκτρονικό εμπόριο αποτελεί φλέγον ζήτημα για τα ηλεκτρονικά καταστήματα, τα Recommender Systems έχουν υιοθετηθεί ευρέως για να βοηθήσουν τους πελάτες να εντοπίσουν προϊόντα που τους ταιριάζουν και πιθανώς θα ήθελαν να αγοράσουν. Μπορεί ακόμη να βοηθήσουν χρήστες που έχουν ελλιπείς γνώσεις σχετικά με κάποιον τομέα προϊόντος. Στην ουσία, αυτά τα συστήματα εφαρμόζουν τεχνικές ανάλυσης δεδομένων σχετικά με τις προτιμήσεις κάθε πελάτη τα οποία συσχετίζονται την προηγούμενη συμπεριφορά του. Ως αποτέλεσμα, παρέχουν μια λίστα με προτεινόμενα προϊόντα για εκείνον στην επόμενη αλληλεπίδρασή του με το σύστημα τα οποία δεν είχε σκεφτεί να αγοράσει μέχρι εκείνη τη στιγμή (Adomavicius & Tuzhilin, 2005) και ταιριάζουν απόλυτα στο προφίλ του εκάστοτε αγοραστή/χρήστη.

Τα Recommender Systems παρουσιάζουν πολλά ενδιαφέροντα πλεονεκτήματα. Πέραν από τις εξατομικευμένες προτάσεις αγοράς που κάνουν στους χρήστες τους, μπορούν να προωθήσουν εξειδικευμένα προϊόντα που ενδεχομένως δεν έχουν

μεγάλη ζήτηση και να ενισχύσουν την αλυσιδωτή πρόταση προϊόντων που σχετίζονται μεταξύ τους. Μπορούν ακόμη να δώσουν την ευκαιρία στους χρήστες να αξιολογήσουν τα προϊόντα εκφράζοντας δημοσίως την άποψή τους και κατ' επέκταση να βοηθήσουν άλλους χρήστες στη λήψη αποφάσεων (Papagelis et al., 2005).

Όσο περισσότερο ένας χρήστης χρησιμοποιεί μια ιστοσελίδα ηλεκτρονικού καταστήματος και αγοράζει αντικείμενα, τόσο περισσότερες πληροφορίες λαμβάνει το σύστημα για το χρήστη και συνεπώς βελτιώνονται οι συστάσεις που κάνει. Το γεγονός αυτό βοηθά στο να οικοδομηθεί μια σχέση προστιθέμενης αξίας μεταξύ της ιστοσελίδας και του χρήστη. Ένα αξιόπιστο Recommender System συμβάλλει στην ικανοποίηση του χρήστη καθώς βελτιώνει την εμπειρία του και αυξάνει την εμπιστοσύνη του στο σύστημα. Αυτό μεταφράζεται σε αύξηση των πωλήσεων και των κερδών για τις εταιρείες καθώς θα επιτυγχάνονται περισσότερες αγορές προϊόντων και έτσι ωθείται η ανταγωνιστικότητα μεταξύ των ηλεκτρονικών καταστημάτων. Για το λόγο αυτό, υπάρχει μεγάλη ερευνητική δραστηριότητα γύρω από την εύρεση καινοτόμων και αποτελεσματικών αλγορίθμων που θα δίνουν ακριβή αποτελέσματα. Τα συστήματα συστάσεων είναι μια ενεργή ερευνητική περιοχή στους τομείς της εξόρυξης δεδομένων και της μηχανικής μάθησης και αποτελεί κύριο θέμα συνεδρίων όπως το ACM RecSys, το SIGIR και το KDD (Adomavicius & Tuzhilin, 2005).

Μια νέα τάση που διέπει τον Παγκόσμιο Ιστό την τελευταία δεκαετία είναι η μετατροπή των παθητικών χρηστών του Διαδικτύου σε ενεργούς χρήστες που οι ίδιοι παράγουν περιεχόμενο. Πλέον, έχουν δημιουργηθεί πλατφόρμες, που ενισχύουν την αλληλεπίδραση μεταξύ των χρηστών δίνοντάς τους τη δυνατότητα να παρουσιάσουν τις απόψεις τους δημοσίως και να συνεισφέρουν στο περιεχόμενο του ιστοτόπου. Ιδιαίτερα δημοφιλή καθίστανται τα λεγόμενα κοινωνικά δίκτυα (social networks), δηλαδή ιστοχώροι που προσομοιώνουν την κοινωνική λειτουργικότητα των χρηστών. Ένας χρήστης που έχει λογαριασμό σε κάποιο κοινωνικό δίκτυο μπορεί να αλληλεπιδράσει με όμοιους χρήστες, να επικοινωνήσει μαζί τους, να ανεβάσει φωτογραφίες, να δημοσιεύσει μια γνώμη, να σχολιάσει την άποψη ενός άλλου χρήστη, να διαθέσει δημοσίως πληροφορίες σχετικά με προτιμήσεις του, να συμμετέχει σε ομάδες, κ.λπ. (Papagelis et al., 2005).

Το πλούσιο περιεχόμενο που παρέχεται μέσω των κοινωνικών δικτύων θα μπορούσε να αποβεί πολύ χρήσιμο για τη βελτίωση των Recommender Systems που αφορούν το ηλεκτρονικό εμπόριο. Στις περισσότερες περιπτώσεις, η κοινωνική επιρροή διαδραματίζει σπουδαίο ρόλο στο marketing των προϊόντων. Άτομα τα οποία συνδέονται με κάποιον κοινωνικό δεσμό, μπορούν να επηρεάσουν ο ένας τον άλλο σχετικά με κάποια λήψη απόφασης. Διαισθητικά, όταν κάποιος επιθυμεί να αγοράσει ένα προϊόν με το οποίο δεν είναι εξοικειωμένος, συχνά συμβουλευτεί τους φίλους ή τους γνωστούς του έστω, οι οποίοι είχαν ήδη κάποια εμπειρία με το εν λόγω προϊόν. Όταν κάποιος φίλος συστήσει ένα προϊόν, το άτομο τείνει να αποδεχθεί την πρότασή του, επειδή θεωρεί τη γνώμη του αξιόπιστη. Το γεγονός αυτό δίνει μια άλλη διάσταση στη δημιουργία των Recommender Systems. Οι παραδοσιακές εφαρμογές συστάσεων δε, λαμβάνουν υπόψη τους τις ρητές κοινωνικές σχέσεις μεταξύ των χρηστών, μια πληροφορία που πλέον είναι διαθέσιμη μέσα από τα κοινωνικά δίκτυα. Εάν κάποιο σύστημα παράγει αποτελέσματα προσομοιώνοντας τον τρόπο με τον οποίο γίνονται οι συστάσεις στην πραγματική ζωή, από στόμα σε στόμα μεταξύ γνωστών, έχει ενδεχομένως μεγάλες πιθανότητες να βελτιωθεί από πλευράς απόδοσης. Αναφορικά με την ακρίβεια πρόβλεψης, οι συμπληρωματικές πληροφορίες σχετικά με τους χρήστες και τους φίλους τους, που προέρχονται από τα κοινωνικά δίκτυα, βελτιώνει την κατανόηση της συμπεριφοράς των χρηστών (Paragelis et al., 2005). Ως εκ τούτου, είναι δυνατή η μοντελοποίηση και η ερμηνεία των προτιμήσεων των χρηστών με μεγαλύτερη ακρίβεια, βελτιώνοντας έτσι την αξιοπιστία της πρόβλεψης. Επιπροσθέτως, έχοντας γνώση σχετική με τους φίλους του χρήστη μπορεί να προτείνει σε εκείνον προϊόντα που αρέσουν σε εκείνους τους χρήστες. Αυτό συμβαίνει στην περίπτωση που ο ίδιος δεν έχει αλληλεπιδράσει αρκετά με το σύστημα και κατά συνέπεια δεν μπορεί να βγάλει ακριβή συμπέρασμα για τις προτιμήσεις του. Η οπτική αυτή μπορεί να έχει ισχύ εάν δύο άνθρωποι είναι πραγματικοί φίλοι, και όχι μόνο «διαδικτυακοί φίλοι», οπότε είναι πιθανό να έχουν και αρκετά κοινά γνωρίσματα και παρόμοιες απόψεις (Adomavicius & Tuzhilin, 2005).

1.2. Ορισμός

Τα συστήματα συστάσεων είναι εφαρμογές λογισμικού που παρέχουν εξατομικευμένες προτάσεις στους χρήστες σχετικά με προϊόντα ή υπηρεσίες που μπορεί να τους ενδιαφέρουν. Προτείνουν στοιχεία σχετικά με τα ενδιαφέροντα των χρηστών, με βάση τις προτιμήσεις τους οι οποίες εκφράζονται άμεσα (explicitly), είτε έμμεσα (implicitly). Τα συστήματα αυτά με τη βοήθεια ειδικών αλγορίθμων επιχειρούν να προβλέψουν ποιες υπηρεσίες είναι πιθανόν να ενδιαφέρουν περισσότερο τον χρήστη. Οι τεχνικές που χρησιμοποιούνται λαμβάνουν ως είσοδο τα χαρακτηριστικά και τις προτιμήσεις του χρήστη (προσωπικά στοιχεία, ιστορικό περιήγησης), τις σχέσεις μεταξύ των χρηστών, και τα γνωρίσματα των προς σύσταση αντικειμένων, και υπολογίζουν το εκτιμώμενο ενδιαφέρον του χρήστη για κάθε αντικείμενο. Στη συνέχεια φιλτράρουν ή ταξινομούν τα αντικείμενα με κριτήριο το εκτιμώμενο ενδιαφέρον. Τυπικά, τέτοια συστήματα χρησιμοποιούν προφίλ χρηστών και έχουν σαν στόχο τη σύσταση στοιχείων που ταιριάζουν περισσότερο στο προφίλ αυτό (Sarwar et al.,2000).

Η δημιουργία προφίλ αποτελεί το πλέον σημαντικό στοιχείο ενός συστήματος συστάσεων. Τα συστήματα συστάσεων εκμεταλλεύονται τις ιδιαιτερότητες των χρηστών με σκοπό να διευκολύνουν στο να προσδιορίζουν ακριβέστερα τις πληροφορίες ή τα προϊόντα για τα οποία ενδιαφέρονται περισσότερο ή σχετίζονται με τις ανάγκες τους. Πιο συγκεκριμένα ένα σύστημα συστάσεων μπορεί να κρατάει ιστορικό από τα άρθρα που έχει διαβάσει κάποιος χρήστης, οπότε την επόμενη φορά που θα επισκεφθεί τον ιστότοπο, το σύστημα θα του προτείνει νέα άρθρα σύμφωνα με την θεματολογία αυτών που είχε διαβάσει στο παρελθόν και πιθανόν να τον ενδιέφεραν. Τα συστήματα συστάσεων αποτελούν ένα σημαντικό πεδίο με μεγάλο ενδιαφέρον σε ερευνητικό επίπεδο, από την εμφάνιση των πρώτων δημοσιεύσεων για το συνεργατικό φιλτράρισμα στα μέσα της δεκαετίας του 90. Υπάρχει μεγάλη δραστηριότητα τόσο στη βιομηχανία όσο και στην ακαδημαϊκή κοινότητα για την ανάπτυξη νέων προσεγγίσεων κατά την τελευταία δεκαετία (Reischach et al., 2009).

Το ενδιαφέρον σε αυτό τον τομέα παραμένει ακόμα υψηλό, δεδομένου ότι πρόκειται για ένα περίπλοκο πρόβλημα σε ερευνητικό επίπεδο, και λόγω του μεγάλου αριθμού πρακτικών εφαρμογών που βοηθούν τους χρήστες να αντιμετωπίσουν το πρόβλημα της υπερφόρτωσης πληροφοριών με το να παρέχουν εξατομικευμένες συστάσεις.

Παραδείγματα τέτοιων εφαρμογών περιλαμβάνουν συστάσεις για βιβλία, CD και άλλα προϊόντα από το δικτυακό τόπο Amazon.com, ταινίες από το δικτυακό τόπο MovieLens, και ειδήσεις στην Versif. Παρά το μεγάλο ενδιαφέρον των εταιρειών και το σημαντικό όγκο ερευνητικής δραστηριότητας για τα συστήματα συστάσεων, απαιτούνται περαιτέρω βελτιώσεις για να καταστήσουν αποτελεσματικές και εφαρμόσιμες σε ακόμα μεγαλύτερο αριθμό εφαρμογών, όπως ορισμένους τύπους οικονομικών υπηρεσιών σε επενδυτές, σε αγορές από κατάσταση που θα γίνονται με έξυπνο καλάθι αγορών. Οι βελτιώσεις αυτές περιλαμβάνουν καλύτερες μεθόδους αναπαράστασης του προφίλ των χρηστών, και των στοιχείων που προτείνονται, πιο εξελιγμένες μεθόδους δημιουργίας συστάσεων, ενσωμάτωση των διάφορων βασισμένων στα συμπραζόμενα πληροφοριών στη διαδικασία συστάσεων, και ανάπτυξη πιο ευέλικτων μεθόδων οι οποίες θα στηρίζονται σε μέτρα που καθορίζουν αποτελεσματικότερα την παραγωγή συστάσεων (Sarwar et al., 2000).

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 περιπτώσεων 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 et al., 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 αποτελούν την εξιδανίκευση των συστημάτων φιλτραρίσματος της πληροφορίας και στόχος τους είναι να παρουσιάσουν στον χρήστη τις πληροφορίες που τον ενδιαφέρουν προσωπικά. Τα συστήματα αυτά 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 δεδομένων οι οποίες έπειτα συγκεντρώνονται και κατευθύνονται σε συγκεκριμένους παραλήπτες.

Οι τομείς οι οποίοι ώθησαν την δημιουργία των συστημάτων συστάσεων είναι (Paragelis et al., 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 ακόμη τομείς ώθησαν στην δημιουργία των συστημάτων συστάσεων το οποίο εξαπλώθηκε με μεγάλη ταχύτητα και εφαρμόστηκε σε πολλές μορφές στο διαδίκτυο όπως το ηλεκτρονικό εμπόριο, συστήματα αξιολόγησης ταινιών ή μουσικής (Paragelis et al., 2005).

2. Διαδικασίες και Μορφές Σύστασης στα ΣΣ

2.1. Διαδικασία Σύστασης

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 αντικείμενου 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. Ομοίως 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 (Reischach et al., 2009).

Τα δεδομένα στο προφίλ του χρήστη μπορούν να εισαχθούν είτε άμεσα είτε έμμεσα. Άμεση είναι η διαδικασία κατά την οποία ο χρήστης βάζει ως δεδομένα στο σύστημα τα προσωπικά του στοιχεία 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 (Reischach et al., 2009).

Ο βαθμός αλληλεπίδρασης του χρήστη και του αντικείμενου μπορεί να γίνει με πολλούς τρόπους. Μπορεί για παράδειγμα να ζητηθεί από το σύστημα στον χρήστη να βαθμολογήσει τον βαθμό της χρησιμότητας ενός αντικείμενου 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 (Adomavicius & Tuzhilin, 2005).

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 et al., 2009).

Μια ακόμα μορφή σύστασης είναι όταν συγκεντρώνονται πολλοί χρήστες σε έναν ιστότοπο και αξιολογούν μαζικά ένα προϊόν και αυτή η αξιολόγηση είναι διαθέσιμη σε όλους τους ενδιαφερόμενους χρήστες. Η μορφή αυτή όπως είναι φανερό βασίζεται 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 (Malone et al., 1987).

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 φιλτράρισμα (Reischach et al., 2009).

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 θεάτρου 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 και από κριτικές που υπάρχουν στις σελίδες κοινωνικής δικτύωσης (Adomavicius & Tuzhilin, 2005).

Γενικά, κάθε Recommender System ακολουθεί μια συγκεκριμένη διαδικασία για την παραγωγή των συστάσεων. Αρχικά το σύστημα θα πρέπει να επεξεργαστεί κάποια δεδομένα εισόδου προτού να παράγει αποτελέσματα. Τρεις πιθανές πηγές πληροφορίας μπορούν να προσδιοριστούν ως είσοδος για τη διαδικασία σύστασης. Οι διαθέσιμες πηγές είναι (Sarwar et al., 2000):

- τα δεδομένα χρήστη,
- τα δεδομένα αντικειμένου και
- οι πιθανές αλληλεπιδράσεις ενός χρήστη με το αντικείμενο (αξιολογήσεις, αγορές κτλ).

Τα δημογραφικά στοιχεία για το χρήστη όπως το φύλο, η ηλικία κ.λ.π. αλλά και οι γνώσεις σχετικά με τις προτιμήσεις του, συνθέτουν ένα "προφίλ χρήστη".

Παράλληλα, κάθε αντικείμενο περιγράφεται με τη βοήθεια χαρακτηριστικών γνωρισμάτων, τα οποία συμβάλλουν στη διαμόρφωση ενός προφίλ αντικειμένου, που ονομάζεται "περιεχόμενο". Για παράδειγμα ένα βιβλίο μπορεί να έχει ως χαρακτηριστικά γνωρίσματα τον τίτλο του, το συγγραφέα του κ.λπ.

Η καταγραφή των δεδομένων του χρήστη όπως έχει ήδη προαναφερθεί μπορεί να γίνει είτε άμεσα (explicitly) είτε έμμεσα (implicitly). Άμεση Καταγραφή δεδομένων σημαίνει ότι ο χρήστης αλληλοεπιδρά με το σύστημα ειδικά για το σκοπό της παροχής των απαραίτητων πληροφοριών σε αυτό. Για παράδειγμα, τα δημογραφικά χαρακτηριστικά του χρήστη καταγράφονται άμεσα όταν ζητούνται κατά τη διάρκεια της εγγραφής του σε ένα ηλεκτρονικό κατάστημα. Αντίθετα, έμμεση καταγραφή δεδομένων σημαίνει πως το σύστημα λαμβάνει τα δεδομένα ενόσω ο χρήστης αλληλοεπιδρά με αυτό για άλλο σκοπό. Λόγου χάρη, τα δεδομένα του χρήστη αποκτώνται έμμεσα όταν συγκεντρώνονται από τα συστήματα καταγραφής συναλλαγών (π.χ. κατά την πληρωμή ενός προϊόντος με τη χρήση πιστωτικών καρτών) ή ακόμη μπορούν να συλλεχθούν με βάση τα δεδομένα της κάρτας SIM του κινητού τηλεφώνου (Paragelis et al., 2005).

Τα δεδομένα σχετικά με τα προϊόντα είναι εύκολο να ληφθούν με άμεσο τρόπο καθώς τα προϊόντα ως επί το πλείστον καταχωρούνται με όλα τα είδη των δεδομένων, των ετικετών ή των χαρακτηριστικών τους. Για παράδειγμα, σε έναν ιστότοπο που αφορά ταινίες αναφέρεται το είδος της κάθε ταινίας, δηλαδή επιστημονικής φαντασίας, δράση ή κωμωδία. Το είδος της ταινίας αποτελεί

παράδειγμα ενός στοιχείου του προϊόντος που παρέχεται από το/στο σύστημα. Τα δεδομένα πρέπει να παρέχονται από τον ιδιοκτήτη του ηλεκτρονικού καταστήματος ή οι χρήστες μπορούν να παρέχουν τα χαρακτηριστικά του προϊόντος δίνοντας τις κατάλληλες ετικέτες σε αυτό. Έμμεσα, τα στοιχεία μπορούν να εξαχθούν από έγγραφα χρησιμοποιώντας κατάλληλα εργαλεία (Adomavicius & Tuzhilin, 2005).

Η αλληλεπίδραση χρήστη-αντικειμένου μπορεί και αυτή να πραγματοποιείται άμεσα ή έμμεσα. Η άμεση αλληλεπίδραση συνίσταται στην ιδέα ότι κάθε αντικείμενο μπορεί να εκτιμηθεί από τους χρήστες μέσω “αξιολογήσεων” (ratings), οι οποίες αντιπροσωπεύουν τη χρησιμότητα ενός αντικειμένου για ένα χρήστη. Κάθε αξιολόγηση εκπροσωπεί ένα χρήστη μέσω ενός ή περισσότερων από τις ακόλουθες επιλογές (Paragelis et al., 2005):

- Δυαδική (Binary): Η επιλογή αυτή βασίζεται στη λογική δύο δυνατών τιμών σχετικά με την προτίμηση του χρήστη (like / dislike).
- Αριθμητική (Numeric): Μία αξιολόγηση μπορεί να δοθεί σε μια συγκεκριμένη κλίμακα. Αυτό είναι συχνά μια κλίμακα πέντε αστέρων (κλίμακα Likert), όπου ένα αστέρι αντιπροσωπεύει τη χαμηλότερη κατάσταση και τα πέντε αστέρια την υψηλότερη κατάσταση. Επίσης, είναι πιθανόν να υπάρχουν κλίμακες άνω των πέντε αστέρων. Η Amazon.com λόγω χάρη χρησιμοποιεί 1 αστέρι για την επιλογή “Το μισώ”, 2 αστέρια για την επιλογή “Δεν μου αρέσει”, 3 αστέρια για την επιλογή “Είναι εντάξει”, 4 αστέρια για την επιλογή “Μου αρέσει” και 5 αστέρια για την επιλογή “Το αγαπώ”.
- Συμβολική (Symbolic): Η επιλογή σχετικά με την προτίμηση του χρήστη είναι μια εικόνα με την οποία ο αξιολογητής επευφημεί (π.χ. μια εικόνα “thumbs up”) ή αποδοκιμάζει το προϊόν (μια εικόνα “thumbs down”).
- Κείμενο (Textual): Σε αυτήν την περίπτωση η εκτίμηση για ένα προϊόν μπορεί να γίνει μέσω ενός γραπτού κειμένου του αξιολογητή σε έναν ιστότοπο (π.χ. μια κριτική ή ένα σχόλιο). Αυτός ο τύπος αξιολόγησης είθισται να συναντάται σε ιστοσελίδες κοινωνικής δικτύωσης.

- Ήχος (Audio): Η επιλογή αυτή αφορά την αξιολόγηση μέσω ενός ηχητικού μηνύματος π.χ. μια προφορική κριτική του προϊόντος.
- Οπτική (Visual): Σε αυτού του είδους επιλογές μπορεί να χρησιμοποιηθούν πολυμεσικές εφαρμογές για την εξαγωγή της αξιολόγησης (π.χ. μια βίντεο-ηχογραφημένη κριτική του προϊόντος με τη χρήση ενός κινητού τηλεφώνου).

Υπάρχουν πολλές άλλες μέθοδοι για την απόκτηση άμεσων αξιολογήσεων, όπως το να ζητηθεί από το χρήστη να αναφέρει ρητά τις προτιμήσεις του ή να επιλέξει τα πιο αγαπημένα αντικείμενα από μια λίστα στοιχείων. Είναι δυνατό να χρησιμοποιηθούν επίσης υβριδικές τεχνικές που συνδυάζουν γενικές ερωτήσεις σχετικά με τις προτιμήσεις ακολουθούμενες από αξιολογήσεις των επιμέρους αντικειμένων (Sarwar et al.,2000).

Όλοι οι χρήστες δεν αξιολογούν όλα τα αντικείμενα που έχουν αγοράσει ή έχουν δει, επειδή απλά δε θέλουν να ξοδεύουν το χρόνο τους αξιολογώντας τα αντικείμενα ή δεν βλέπουν το λόγο για να το κάνουν. Ακόμη, δεν είναι απαραίτητο ότι όλοι οι πελάτες που εγγράφονται στο ηλεκτρονικό κατάστημα είναι διατεθειμένοι να δώσουν όλες τις προσωπικές πληροφορίες τους. Για αυτούς τους χρήστες μια άλλη πηγή πληροφοριών απαιτείται για να ξεπεραστεί η έλλειψη των αξιολογήσεων και εν τέλει των χρήσιμων πληροφοριών. Μια προσέγγιση σε αυτό το πρόβλημα είναι να χρησιμοποιηθούν έμμεσοι τρόποι αξιολόγησης μέσω της παρατήρησης της συμπεριφοράς του χρήστη (Malone et al., 1987). Ένα σενάριο για τη σύλληψη του προϊόντος με έμμεσο τρόπο μπορεί να είναι ένας καταναλωτής που αγοράζει ένα προϊόν, ή η επιστροφή και επισκευή ενός προϊόντος. Στην περίπτωση που ο χρήστης έχει αγοράσει πολλές φορές το ίδιο προϊόν, το σύστημα μπορεί να συμπεράνει ότι το εν λόγω προϊόν είναι αρεστό. Αντίθετα, αν ο χρήστης επιστρέψει την επόμενη ημέρα το προϊόν που είχε αγοράσει και αγοράσει ένα άλλο προϊόν της ίδιας κατηγορίας, το σύστημα θα μπορούσε να συνάγει ότι στο χρήστη δεν άρεσε το προϊόν που αγόρασε πρώτο (Sarwar et al.,2000).

Ένα άλλο τυπικό σενάριο της έμμεσης συλλογής δεδομένων στο διαδίκτυο είναι η καταγραφή των δεδομένων όταν ο χρήστης κάνει περιήγηση στις περιγραφές των προϊόντων, όπως στο Amazon.com, ή σαρώνει προϊόντα στο ηλεκτρονικό κατάστημα με ένα κινητό τηλέφωνο στα πλαίσια μιας άλλης εφαρμογής. Το βασικό

κίνητρο για τη χρήση έμμεσων αξιολογήσεων είναι ότι δεν υπάρχει κόστος για τον αξιολογητή όσον αφορά την εξέταση και την αξιολόγηση του στοιχείου. Ωστόσο, εάν η αξιολόγηση προκύπτει έμμεσα, μπορεί οι συστάσεις να μην είναι τόσο ακριβείς και αξιόπιστες. Για παράδειγμα, ο ίδιος υπολογιστής μπορεί να χρησιμοποιείται από περισσότερα του ενός άτομου. Όμως, το σύστημα μπορεί να θεωρεί ενδεχομένως ότι όλες οι διαφορετικές αλληλεπιδράσεις προέρχονται από τον ίδιο χρήστη και συνεπώς να καταλήγει σε λανθασμένες συστάσεις. Επιπλέον, ένα άλλο βασικό μειονέκτημα είναι η αδυναμία του συστήματος να εξάγει συμπεράσματα όταν δεν υπάρχουν ενέργειες από την πλευρά του χρήστη (Adomavicius & Tuzhilin, 2005).

2.2. Μορφές της διαδικασίας σύστασης

Αφού συλλεχθούν όλα τα απαραίτητα δεδομένα εισόδου, η διαδικασία της σύστασης προϋποθέτει τη μαζική επεξεργασία τους και τη διάθεση του αποτελέσματος είτε προς όλους τους χρήστες είτε προς έναν επιλεγμένο αριθμό εν δυνάμει καταναλωτών. Εν γένει, υπάρχουν τέσσερις επιλογές πάνω στις οποίες μπορεί να βασίζεται μια διαδικασία σύστασης: (Balabanović & Shoham, 1997).

2.2.1. Ένας προς έναν

Αυτές είναι προσεγγίσεις όπου η σύσταση ενός συγκεκριμένου χρήστη απευθύνεται σε έναν άλλο χρήστη, συμπεριλαμβανομένου του ίδιου. Ειδικότερα, ο χρήστης που κάνει τη σύσταση γνωρίζει πολλά για έναν ορισμένο τύπο του προϊόντος και / ή πολλά για ένα συγκεκριμένο ομότιμο (peer) χρήστη, οπότε και συνιστά ένα συγκεκριμένο προϊόν σε αυτόν το χρήστη. Αυτό το είδος σύστασης θα μπορούσε να παραλληλιστεί με το λεγόμενο word-of-mouth, όπου ένας καταναλωτής λέει προφορικά σε έναν άλλο καταναλωτή για την εμπειρία του σχετικά με ένα προϊόν. Μια προσέγγιση όπου τα στοιχεία του ιστορικού ενός χρήστη χρησιμοποιούνται για να προβλέψουν μια σύσταση για τον ίδιο τον χρήστη ονομάζεται αλγόριθμος φιλτραρίσματος με βάση το περιεχόμενο. Είναι πιθανό ένας συγκεκριμένος χρήστης να κάνει μια σύσταση στον εαυτό του, ώστε να μην ξεχνά την επιλογή του προϊόντος που έχει κάνει και με την οποία είναι ικανοποιημένος. Τυπικά παραδείγματα στο διαδίκτυο είναι ηλεκτρονικές λίστες αγοράς και αυτόματα αποθηκευμένα ηλεκτρονικά καλάθια όταν αποσυνδεθεί ο χρήστης πριν από την αγορά, που χρησιμοποιούνται στο Amazon.com για συστάσεις (Balabanović & Shoham, 1997).

2.2.2. Πολλοί προς πολλούς

Σε αυτές τις προσεγγίσεις, οι αξιολογήσεις πολλών απροσδιόριστων χρηστών μπορεί να συγκεντρώνονται και να υπόκεινται σε επεξεργασία (λόγου χάρη να υπολογίζεται ο μέσος όρος τους αν πρόκειται για αριθμητικές αξιολογήσεις) για ένα συγκεκριμένο προϊόν (το λεγόμενο φιλτράρισμα με βάση το αντικείμενο) και να διατίθενται σε πολλούς άλλους απροσδιόριστους χρήστες. Υπάρχει επίσης το ενδεχόμενο οι αξιολογήσεις των χρηστών να μην είναι αριθμητικά μετρήσιμες (π.χ. κριτικές για το προϊόν σε κείμενο) και να διατίθενται ως έχουν για πολλούς χρήστες.

Επειδή σε αυτήν την περίπτωση η σύσταση δεν έχει δημιουργηθεί ειδικά για ένα χρήστη, επαφίεται σε αυτόν να κρίνει αν αυτή ταιριάζει με τις ανάγκες του (Balabanović & Shoham, 1997).

2.2.3. Πολλοί προς έναν

Σε αυτήν την περίπτωση, συστάσεις από πολλούς χρήστες αθροίζονται σε ένα συγκεκριμένο χρήστη. Η προσέγγιση αυτή βασίζεται στην ιδέα ότι οι άνθρωποι είναι ανεξάρτητα άτομα και τους αρέσουν διαφορετικά προϊόντα. Υπάρχουν αλγόριθμοι που εστιάζουν σε αυτό το γεγονός, λαμβάνοντας συστάσεις ειδικών χρηστών που έχουν ληφθεί έμμεσα και / ή άμεσα και απευθύνουν συστάσεις σε έναν άλλο χρήστη με παρόμοιο προφίλ. Η διαδικασία αυτή αναφέρεται συχνά ως συνεργατικό φιλτράρισμα (Malone et al., 1987).

2.2.4. Ένας προς πολλούς

Σε αυτήν την περίπτωση η σύσταση ενός χρήστη που έχει κάποια ιδιαίτερη σχέση με ένα προϊόν διατίθεται σε πολλούς ενδιαφερόμενους χρήστες. Ένα παράδειγμα είναι ένας εμπειρογνώμονας κρασιού που μοιράζεται την εμπειρία του για ένα ορισμένο κρασί με τους φίλους του μέσα από ένα κοινωνικό δίκτυο. Η εφαρμογή LivingSocial.com του Facebook είναι μέχρι στιγμής η πιο δημοφιλής αυτών των Recommender Systems που βασίζονται στα κοινωνικά δίκτυα. Επιτρέπει στους χρήστες να επιλέξουν ρητά τα προϊόντα που τους αρέσουν, να τα κριτικάρουν ρητά και να μοιράζονται αυτά τα σχόλια με φίλους στις πλατφόρμες κοινωνικών δικτύων όπως το Facebook. Αλλά υπάρχουν επίσης εφαρμογές του Facebook, που συλλέγουν συστάσεις προϊόντων έμμεσα και τις διαθέτουν στους φίλους, οι λεγόμενες κοινωνικές διαφημίσεις. Για παράδειγμα, το Facebook έμμεσα συλλέγει συναλλαγές κάποιου χρήστη του με το δικτυακό τόπο εισιτηρίων “Fandago” και τις γνωστοποιεί σε άλλους χρήστες. Αυτή η γνωστοποίηση ουσιαστικά μπορεί να αποτελέσει ταυτόχρονα σύσταση για μια ταινία ή ένα θεατρικό έργο. Σε αυτήν αλλά και σε παρόμοιες περιπτώσεις, η προστασία της ιδιωτικής ζωής είναι ένα σημαντικό ζήτημα. Η ιστορία δείχνει ότι οι καταναλωτές δε θέλουν να μοιραστούν συστάσεις προϊόντων που έχουν αποδοθεί στο όνομά τους με άλλους τυχαίους χρήστες (Sarwar et al., 2000).

3. Αλγόριθμοι και Τεχνικές Συστημάτων Συστάσεων

3.1. Αξιολόγηση Συστημάτων Συστάσεων

Σε κάθε τεχνική που αναλύθηκε παραπάνω τα αποτελέσματά της μπορούν να αξιολογηθούν και στην συνέχεια να βελτιώνονται κάθε φορά που υπάρχουν νέα δεδομένα. Υπάρχουν πολλοί τρόποι με τους οποίους ένα σύστημα μπορεί να αξιολογηθεί και για 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.

Η απόλυτη τιμή σφάλματος δίνεται από:

$$(\Delta x) = |x_i - x|, \text{ όπου}$$

x_i : η μέτρηση/βαθμολογία του χρήστη

x : η πραγματική τιμή/βαθμολογία του χρήστη

Το μέσο απόλυτο σφάλμα δίδεται από την ακόλουθη φόρμουλα:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x|$$

όπου n είναι το σύνολο των μετρήσεων.

- 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. 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 (Breese, 1998). Η ακρίβεια εκφράζεται είτε ως απόλυτο σφάλμα ή ως σχετικό σφάλμα. Το απόλυτο E_a σφάλμα της μέσης τιμής ενός μικρού αριθμού επαναλαμβανόμενων μετρήσεων δίνεται από τη σχέση:

$$E_a = x_\mu - x_t$$

Όπου x_t είναι μια αποδεκτή τιμή της ποσότητας που μετρείται. Συχνά όμως είναι περισσότερο χρήσιμο η ακρίβεια να εκφράζεται με το αντίστοιχο σχετικό σφάλμα δηλαδή τη ποσότητα:

$$\left[\frac{(x - x_t)}{x_t} \right] \times 100\%$$

Είναι προφανές ότι το απόλυτο αλλά και το σχετικό σφάλμα διαθέτουν πρόσημο που υποδηλώνει ότι το μετρούμενο μέγεθος είναι μεγαλύτερο από το πραγματικό (θετικό πρόσημο) ή όταν ισχύει το αντίθετο (αρνητικό).

- 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 (Breese, 1998):

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

3.2. Τεχνικές 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, 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 (Bridge, 2005).
- Τα συστήματα τα οποία είναι βασισμένα στην συνεργασία (Collaborative Systems). Τα συστήματα αυτά κρίνουν αν ένας χρήστης θα θεωρεί χρήσιμες κάποιες πληροφορίες ανάλογα με το ιστορικό που υπάρχει. Εξετάζουν δηλαδή αν οι χρήστες που έψαχναν 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 (Breese, 1998).
- Τα υβριδικά συστήματα συνεργασίας (Hybrid Systems). Τα συστήματα αυτά χρησιμοποιούν ένα συνδυασμό των δύο παραπάνω κατηγοριών για να παρουσιάσουν όσο τον δυνατόν καλύτερα αποτελέσματα.

3.2.1. Συστήματα βασισμένα στο Περιεχόμενο

Στην συγκεκριμένη τεχνική προτείνονται στο χρήστη αντικείμενα όμοια με αυτά που είχε προτιμήσει στο παρελθόν (Adomavicius & Tuzhilin, 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 (Breese, 1998).

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, 2005). Σε μια 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 (Balabanović & Shoham, 1997).

Σύγχρονες μέθοδοι φιλτραρίσματος με βάση το περιεχόμενο λειτουργούν ώστε να προτείνουν αντικείμενα που περιέχουν πληροφορίες όπως κείμενα γενικού ενδιαφέροντος, 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 (Breese, 1998).

Η πράξη και η θεωρία αποδεικνύει ότι οι πληροφορίες σχετικά με το προφίλ των χρηστών πηγάζουν ρητά από τον ίδιο το χρήστη. Αυτό επιτυγχάνεται μέσω ερωτηματολογίων που υποβλήθηκαν ή από την συναλλακτική συμπεριφορά τους στην πάροδο του χρόνου. 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 (Adomavicius & Tuzhilin, 2005). 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, ειδησεογραφικά άρθρα ή έγγραφα. Σε αυτήν την 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 Vector Space Model (VSM). Η Vector Space Model είναι μια αναπαράσταση εγγράφων κειμένου την οποία θα παρουσιάσουμε αναλυτικότερα σε επόμενο κεφάλαιο – ενότητα (Balabanović & Shoham, 1997).

Μειονεκτήματα των συστημάτων που βασίζονται στο περιεχόμενο.

Τα συστήματα που βασίζονται στο περιεχόμενο έχουν κάποιους περιορισμούς όπως (Adomavicius & Tuzhilin, 2005):

- Περιορισμοί οι οποίοι πηγάζουν 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 αλγόριθμων και το αποτέλεσμα τους δεν είναι σίγουρο πως μπορεί να είναι πάντα το ορθότερο. 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

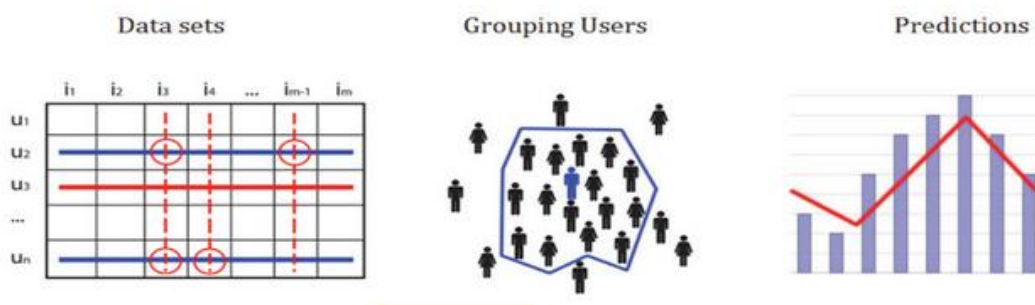
την αντιμετώπιση αυτού του προβλήματος έχουν δημιουργηθεί αλγόριθμοι οι οποίοι προσθέτουν τυχαιότητα στις προτάσεις του συστήματος.

- Τέλος δημιουργούνται πολλά προβλήματα με τους νέους χρήστες οι οποίοι δεν έχουν συγκεντρώσει πολλά στοιχεία στο προφίλ τους και έτσι λόγω ελλιπών δεδομένων τα αποτελέσματα που έχουν από τα συστήματα συστάσεων δεν μπορούν να είναι αξιόπιστα. (Bridge, 2005).

3.2.2. Συστήματα Βασισμένα στην Συνεργασία

Η τεχνική του φιλτραρίσματος βασισμένα στην συνεργασία στηρίζεται στις αξιολογήσεις των χρηστών που μοιράζονται παρόμοια ενδιαφέροντα μεταξύ τους. Είναι η δημοφιλέστερη και η πλέον 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 βασισμένα σε αντικείμενα που προηγουμένως είχαν βαθμολογηθεί από άλλους χρήστες (Adomavicius & Tuzhilin, 2005).

Η τεχνική του φιλτραρίσματος βασισμένα στην συνεργασία (ΣΒΣ) εξυπηρετεί τον 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 (Breese, 1998).

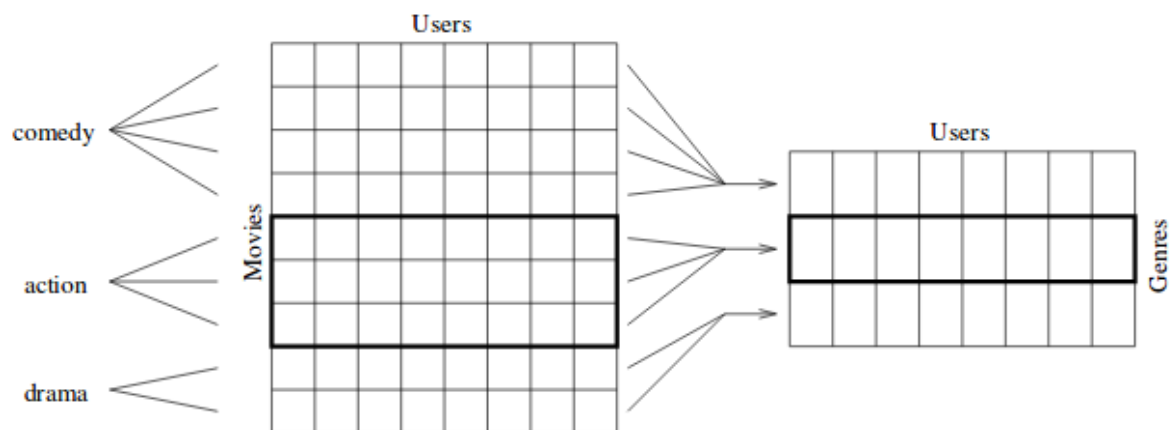


Εικόνα 3 1 Διαδικασία που ακολουθείται στα συστήματα βασισμένα στη συνεργασία

Γενικότερα τα (ΣΒΣ) αναλύουν τις απόψεις των άλλων χρηστών σε σχέση με τα στοιχεία, έτσι παρέχουν έναν βαθμό αρεσκείας όχι με βάση τη φύση του στοιχείου, αλλά βάσει της ανθρώπινης κρίσης. Οι παρακάτω εικόνες δείχνουν 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 (Balabanović & Shoham, 1997).

		ITEMS			
		I1	I2	...	Im
USERS	U1	?	7	...	?
	U2	5	?	...	10
	⋮	⋮	⋮	⋮	⋮
	Un	3	?	...	8

Εικόνα 3 2 Παράδειγμα μιας μήτρας από βαθμολογίες χρηστών για ένα αντικείμενο.



Εικόνα 3 3 Κατηγοριοποίηση και σύσταση των βαθμολογούμενων ταινιών βάση του φύλλου.

Εμπορικές εφαρμογές που υλοποιούν τέτοιο φιλτράρισμα είναι (Breese, 1998):

- 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 Street 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 γειννίας του χρήστη. Η πρώτη βασίζεται στη μνήμη (memory/heuristic - based) και η δεύτερη βασίζεται σε μοντέλο (model-based) τα οποία θα αναλύσουμε παρακάτω (Breese et al., 1998).

3.2.3. Συστήματα Προσανατολισμένα Στην Μνήμη (Memory Based)

Οι αλγόριθμοι με βάση τη μνήμη χαρακτηρίζονται ως «ευρετικοί»(heuristic) 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 Coefficient, PCC), Ευκλείδεια Μετρική (Euclidean distance) ή η ομοιότητα με βάση το συνημιτονικό διάνυσμα (cosine-based similarity) που θα αναλυθούν παρακάτω (Adomavicius & Tuzhilin, 2005).

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, 2005):

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

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

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

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

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

$$\bar{r}_c = (1/|S_c|) \sum_{s \in S_c} r_{c,s},$$

όπου

$$S_c = \{s \in S | r_{c,s} \neq \emptyset\}$$

3.2.4. 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 (Breese, 1998).

Η γενική ιδέα είναι να μοντελοποιηθούν οι αλληλεπιδράσεις χρηστών-προϊόντων με παράγοντες που εκπροσωπούν τα ελλιπή 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. Το μοντέλο αυτό εκπαιδεύεται χρησιμοποιώντας τα διαθέσιμα

δεδομένα και στη συνέχεια εφαρμόζεται για να προβλέψει τις βαθμολογήσεις των χρηστών σε καινούρια προϊόντα (Burke, 2000).

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. 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 (Burke, 2000).

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 (Bridg, 2005).

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) (Breese, 1998).

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 (Su & Khoshgoftaar, 2009). Συνήθως, οι αλγόριθμοι ταξινόμησης μπορούν να χρησιμοποιηθούν ως μοντέλα φιλτραρίσματος εάν οι αξιολογήσεις των χρηστών είναι κατηγοριοποιημένες, ενώ μοντέλα παλινδρόμησης και μέθοδοι SVD χρησιμοποιούνται για αριθμητικές βαθμολογίες.

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 (Bridge, 2005).

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

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 (Burke, 2000):

- **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 (Su & Khoshgoftaar, 2009).
- **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 (Breese, 1998):

$$f_i = \log(n/n_j)$$

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 et al., 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.

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 (Su & Khoshgoftaar, 2009):

- 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, 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 et al., 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 (Breese, 1998).
- 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 (Pazzani, 1999).
- 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. Για να αντιμετωπιστεί αυτό το πρόβλημα θα πρέπει να χρησιμοποιηθεί ένας συνδυασμός από τα συστήματα με βάση το περιεχόμενα και τα συστήματα με βάση την συνεργασία (Sarwar et al., 2000).

- 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 στη 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. 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 (Balabanović & Shoham, 1997). 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. 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 (Pazzani, 1999).

- 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 (Balabanović & Shoham, 1997).
- 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 (Breese, 1998).
- 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 (Balabanović & Shoham, 1997).
- 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 (Breese, 1998).

- 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 αρέσουν 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 προβλεφθεί (Sarwar et al., 2000).

3.2.5. Συστήματα βασισμένα σε Δημογραφικά δεδομένα (Demographic Based)

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 (Adomavicius & Tuzhilin, 2005).

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.

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 ωφελήσει σε τίποτα αν το συμπληρώσουν (Sarwar et al., 2000).

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.

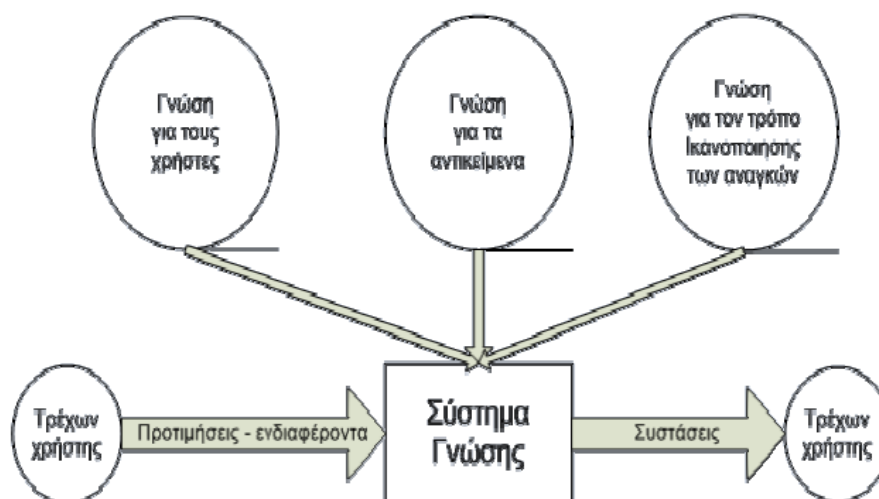
3.2.6. Συστήματα Βασισμένα Στην γνώση (Knowledge- Based Systems).

Η κατηγορία αυτή των συστημάτων είναι τα συστήματα τα οποία βασίζονται πάνω στην γνώση τους για κάποια χαρακτηρίστηκα 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, 1999):

- 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 (Balabanović & Shoham, 1997).

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 3 4 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.

3.2.7. 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, 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 χαρακτηρίστηκα 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.

3.2.8. 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 (Bridge et al., 2005):

- 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 τον στόχο 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, 1999).



I would like to eat at a restaurant that has:

Cuisine Price

Style Atmosphere Occasion

I would like to eat at a restaurant just like:

Chinois on Main Los Angeles

New Query Submit

Εικόνα 3 5 Αρχική σελίδα συστήματος Entree.

Entree Results

The Los Angeles restaurant you chose is:

Chinois On Main	
2709 Main St. (bet. Rose Ave. & Ocean Park Blvd.), Santa Monica, 310-392-9025	
Pacific New Wave	\$30-\$50
Extraordinary Decor, Extraordinary Service, Near-perfect Food, Hip Place To Be, On the Beach, Great for People Watching, Parties and Occasions, Weekend Brunch, Weekend Lunch, Fabulous Wine Lists	

We recommend:

Yoshi's Cafe	
3257 N. Halsted St. (Belmont Ave.), Chicago, 312-248-6160	
Asian, Japanese, French (New)	\$30-\$50
Extraordinary Decor, Extraordinary Service, Near-perfect Food, Need To Dress, Prix Fixe Menus, Quiet for Conversation, Very Busy - Reservations a Must, Romantic, Good Out of Town Business, Fabulous Wine Lists, Game, Parking/Valet	

less \$\$ nicer cuisine

traditional creative livelier quieter

For other suggestions, select:

Yoshi's Cafe	302 West	Lulu's
Penny's Noodle Shop	Ann's	Trio
Emilio's Tapas Bar & Restaurant	Nick's Fishmarket	Bossa Nova
Emilio's Granada		

Εικόνα 3.6 Knowledge-based αποτελέσματα συστήματος Entree.

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 (Adomavicius & Tuzhilin, 2005).

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 (Breese, 1998).

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 (Balabanović & Shoham, 1997).

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 (Adomavicius & Tuzhilin, 2005).

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 (Breese, 1998).

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 (Balabanović & Shoham, 1997).

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 (Adomavicius & Tuzhilin, 2005).

3.2.9. 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 (Burke, 2007).

4. Εφαρμοσμένος αλγόριθμος Συστήματος Συστάσεων

4.1. Παρουσίαση αλγορίθμου

Στην εφαρμογή η οποία ακολουθεί θα προσπαθήσουμε να παρουσιάσουμε και να αναλύσουμε έναν αλγόριθμο ενός συστήματος συστάσεων. Πιο συγκεκριμένα ο αλγόριθμος αφορά την υλοποίηση ενός συστήματος βασισμένο στη συνεργασία (collaborative filtering) όπως περιεγράφηκε στην ενότητα 3.2.2 και ο υπολογισμός της ομοιότητας θα γίνει βάσει της Ευκλείδειας απόστασης των δεδομένων του συστήματος.

Τα δεδομένα της εφαρμογής αποτελούνται ουσιαστικά από τις βαθμολογίες μιας ομάδας ατόμων. Τα άτομα αυτά θα μπορούσαν υποθετικά να αποτελούν πελάτες/χρήστες σε μια διαδικτυακή πλατφόρμα όπου προσφέρει ένα σύνολο από ταινίες και τηλεοπτικές σειρές προς προβολή. Οι εκάστοτε βαθμολογίες θα αφορούν τον βαθμό που έχουν δώσει οι users της πλατφόρμας σε διάφορες σειρές. Ως έξοδος, θα κατασκευάζεται μια λίστα από προτιμήσεις για έναν user όπου θα βασίζεται στις άλλες κοινές προτιμήσεις που έχουν οι χρήστες μεταξύ τους.

Έστω πως τα δεδομένα/αξιολογήσεις για κάθε πελάτη είναι της μορφής:

```
 {"user1",
  [{"Action", 1.46},
  {"Adventure", 5.0},
  {"Animation", 3.34},
  {"Biography", 2.32}]
},
 {"user2",
  [{"Action", 5.0},
  {"Adventure", 2.54},
  {"Documentary", 4.32},
  {"Family", 2.76}]
},
 {"user3",
  [{"Music", 4.89},
  {"Action", 4.72},
  {"Sport", 3.61},
  {"Thriller", 3.58},
  {"Talk Show", 3.23},
```

Αλγοριθμικές τεχνικές και εφαρμογές για Συστήματα Συστάσεων

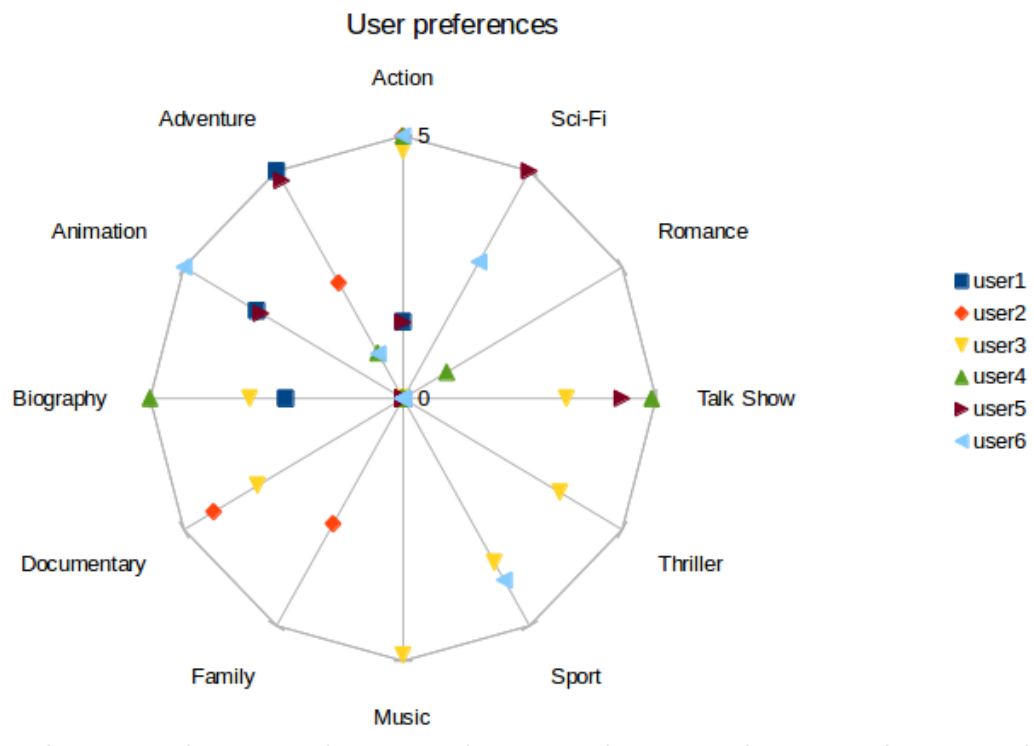
```
    {"Documentary", 3.32},
    {"Biography", 3.03}
  },
  {"user4",
    {"Action", 5},
    {"Adventure", 1},
    {"Biography", 5},
    {"Romance", 1},
    {"Talk Show", 4.92}}
  },
  {"user5",
    {"Talk Show", 4.33},
    {"Action", 1.45},
    {"Adventure", 4.79},
    {"Animation", 3.24},
    {"Sci-Fi", 5.0}}
  },
  {"user6",
    {"Action", 5},
    {"Adventure", 0.98},
    {"Animation", 5},
    {"Sci-Fi", 3},
    {"Biography", 5}}
  }
}
```

Συγκεντρωτικά σε έναν πίνακα:

Movies	user1	user2	user3	user4	user5	user6
Action	1.46	5.0	4.72	5.0	1.45	5.0
Adventure	5.0	2.54	0.0	1.0	4.79	0.98
Animation	3.34	0.0	0.0	0.0	3.24	5.0
Biography	2.32	0.0	3.03	5.0	0.0	0.0
Documentary	0.0	4.32	3.32	0.0	0.0	0.0
Family	0.0	2.76	0.0	0.0	0.0	0.0
Music	0.0	0.0	4.89	0.0	0.0	0.0
Sport	0.0	0.0	3.61	0.0	0.0	4.0
Thriller	0.0	0.0	3.58	0.0	0.0	0.0
Talk Show	0.0	0.0	3.23	4.92	4.33	0.0
Romance	0.0	0.0	0.0	1.0	0.0	0.0
Sci-Fi	0.0	0.0	0.0	0.0	5.0	3.0

Η διαγραμματική απεικόνιση για το σύνολο των ταινιών παραπάνω των φαίνεται στη κάτωθι εικόνα όπου μπορούμε να διακρίνουμε πως ο «user1» έχει αρκετά κοινά

στοιχεία με τον χρήστη «user5». Σε αυτό το αποτέλεσμα θα οδηγηθούμε όταν εκτελέσουμε την εφαρμογή και θα είναι το αποτέλεσμα του αλγορίθμου.



Εικόνα 4.1 Αξιολογήσεις των χρηστών για κάθε ένα είδος ταινίας

Για τον υπολογισμό της Ευκλείδειας απόστασης μεταξύ δυο ατόμων όσον αφορά τις βαθμολογίες τους θα χρησιμοποιήσουμε τον παρακάτω μαθηματικό τύπο:

$$d(User_i, User_j) = \sqrt{(X1 - X2)^2 + (Y1 - Y2)^2}$$

Όπου όσο μικρότερη είναι η μεταξύ τους απόσταση τόσο ταιριάζουν οι προτιμήσεις τους. Αυτό έχει σαν αποτέλεσμα εάν μια ταινία είναι αρεστή (ή και όχι) στον έναν από τους 2, τότε είναι πολύ πιθανό και ο 2^{ος} χρήστης να προτιμήσει να αγοράσει (ή όχι) αυτήν την ταινία οπότε και θα τους προταθεί προς αγορά.

Επιπρόσθετα, αξίζει να αναφερθεί πως εφόσον ενδιαφερόμαστε για μια κλίμακα σύγκρισης θα χρησιμοποιήσουμε τον ακόλουθο τύπο που γίνεται κανονικοποίηση στο διάστημα [0,1] της παραπάνω τιμής:

$$\frac{1}{[1 + d(User_i, User_j)]} \in [0,1]$$

4.2. Δομή Εφαρμογής

Ο αλγόριθμος είναι υλοποιημένος στην γλώσσα C++11 και αποτελείται από 3 φάσεις. Η πρώτη φάση αποτελεί την εισαγωγή των ήδη γνωστών τιμών που έχει αποθηκευμένο το σύστημα.

```
01. // Vector of all Users.
02. std::vector<User> userVec;
03. fill_user_data(userVec);
```

Στην 2^η φάση βρίσκουμε για έναν συγκεκριμένο χρήστη που μας ενδιαφέρει με ποιους άλλους μοιράζεται κοινά ενδιαφέροντα. Χρησιμοποιώντας την παραπάνω μαθηματική παράσταση, αποθηκεύουμε σε έναν vector κατά φθίνουσα σειρά τους χρήστες με τους οποίους βρίσκεται πιο κοντά στις προτιμήσεις.

```
01. /*
02.  * Check similarities for a single user.
03.  * At first we need to extract user's info from the vector.
04.  */
05. std::string recomName = argv[1];
06. User recommendUser = extract_user_metadata(recomName, userVec);
07.
08. // Find Similar users.
09. std::map<std::string, double> results;
10. euclidean_similarity(recommendUser, userVec, results);
11.
12. std::cout << "Similarity to other users:" << std::endl;
13. print_results(results);
```

Η εκάστοτε απόσταση βρίσκεται από το παρακάτω snippet κώδικα.

```
01. // The Euclidean algorithm.
02. void euclidean_similarity(const User recomUser, std::vector<User>& userVec, std::map<std::string, double>& results) {
03.
04.     double sum = 0;
05.     for ( std::vector<User>::iterator pIt = userVec.begin(); pIt != userVec.end(); ++pIt ) {
06.         sum = 0;
07.         if ( pIt->name == recomUser.name ) {
08.             //std::cout << "[DEBUG] User : " << pIt->name << " is the same. Skipping."<< std::endl;
09.             continue;
10.         }
11.
12.         for ( std::map<std::string, double>::const_iterator recomIt = recomUser.preferences.begin(); recomIt != recomUser.preferences.end(); ++ recomIt ) {
13.             double dif = 0;
14.             if ( pIt->preferences.find(recomIt->first) != pIt->preferences.end() ) {
15.                 dif = recomIt->second - pIt->preferences[recomIt->first];
16.             }
17.             sum = sum + dif*dif;;
18.         }
19.         results[pIt->name] = 1 / (1 + std::sqrt(sum));;
20.     }
21. }
```

Στην 3^η φάση της εφαρμογής, πολλαπλασιάζουμε τα στοιχεία για τα οποία δεν έχει ακόμη δηλώσει ενδιαφέρον ο χρήστης προς μελέτη, με το εκάστοτε ειδικό βάρος που υπολογίστηκε από το προηγούμενο βήμα για όλους τους υπόλοιπους.

```
01. // Calculate the preferences according to the Recommender System.
02. std::map<std::string, double> itemsScore;
03. std::map<std::string, double> weight;
04. preference_calculation(userVec, recommendUser, results, itemsScore, weight);
```

```
01. void preference_calculation(std::vector<User>& userVec, User& recommendUser,  
02. std::map<std::string, double>& results, std::map<std::string, double>& itemsScore, std::map<std::string, double>& weight) {  
03.     for (std::map<std::string, double>::iterator it = results.begin(); it != results.end(); ++it) {  
04.         User usr = extract_user_metadata(it->first, userVec);  
05.         for ( std::map<std::string, double>::const_iterator prIt = usr.preferences.begin(); prIt != usr.preferences.end(); ++ prIt ) {  
06.             if ( !recommendUser.preferences[prIt->first] ) {  
07.                 itemsScore[prIt->first] += it->second * prIt->second;  
08.                 weight[prIt->first] += it->second;  
09.             }  
10.         }  
11.     }  
12.  
13.     for ( std::map<std::string, double>::const_iterator scIt = itemsScore.begin(); scIt != itemsScore.end(); ++ scIt ) {  
14.         itemsScore[scIt->first] = itemsScore[scIt->first] / weight[scIt->first];  
15.     }  
16. }
```

4.3. Παρουσίαση αποτελεσμάτων

Η εφαρμογή παρουσιάζει κατά φθίνουσα σειρά τους χρήστες με τους οποίους μοιράζεται κοινά ενδιαφέροντα ο προς μελέτη «user1».

Similarity to other users:

user5 (=0.811156)

user3 (=0.230605)

user2 (=0.188295)

user6 (=0.151336)

user4 (=0.143346)

Από τα παραπάνω προκύπτει πως είναι αρκετά κοντά στον «user5» καθώς η μεταξύ τους απόσταση είναι πολύ μικρή. Το αποτέλεσμα ήταν αυτό που αναμέναμε από την οπτική απεικόνιση των αρχικών δεδομένων.

Στο τέλος αναμένουμε πως οι επιλογές του «user5», όπου δεν είναι κοινές με τον «user1» θα είναι αρκετά υψηλά στη λίστα που θα προταθούν/διαφημιστούν στον αρχικό χρήστη.

Recommender System results:

Sci-Fi (=5)

Music (=4.89)

Talk Show (=4.18732)

Documentary (=3.7695)

Sport (=3.76453)

Thriller (=3.58)

Family (=2.76)

Romance (=1)

Από τα παραπάνω προκύπτει για παράδειγμα πως είναι αρκετά πιθανό οι «Sci-Fi» είδους ταινιών να προτιμώνται από τον πελάτη για αυτό και θα είναι οι πρώτες που θα προβάλλονται προς κατανάλωση.

5. Συμπεράσματα και προοπτικές

Η εφαρμογή μπορεί να βελτιωθεί υλοποιώντας και άλλους αλγορίθμους που αφορούν τα συστήματα προσανατολισμένα στη Μνήμη (Memory based). Με αυτόν τον τρόπο μπορούμε να συγκρίνουμε τα αποτελέσματα και να ελεγχθεί ο λόγος για τον οποίο παρατηρούνται αποκλίσεις (εάν υπάρχουν σημαντικές διαφορές).

Επίσης πρόσφορο έδαφος σε τέτοιου είδους εφαρμογές βρίσκει και η χρησιμοποίηση των data mining και machine learning αλγορίθμων γιατί μπορούμε πολύ εύκολα να οπτικοποιήσουμε και να συγκρίνουμε τα αποτελέσματα.

Βιβλιογραφία

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering*, (6), 734-749.
- Balabanović, M., & Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66-72.
- Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (pp. 43-52). Morgan Kaufmann Publishers Inc.
- Bridge, D., Göker, M. H., McGinty, L., & Smyth, B. (2005). Case-based recommender systems. *The Knowledge Engineering Review*, 20(3), 315-320.
- Burke, R. (1999). Integrating knowledge-based and collaborative-filtering recommender systems. In *Proceedings of the Workshop on AI and Electronic Commerce* (pp. 69-72).
- Burke, R. (2000). A case-based reasoning approach to collaborative filtering. In *European Workshop on Advances in Case-Based Reasoning* (pp. 370-379). Springer, Berlin, Heidelberg.
- Burke, R. (2007). Hybrid web recommender systems. In *The adaptive web* (pp. 377-408). Springer, Berlin, Heidelberg.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5-53.
- Malone, T. W., Yates, J., & Benjamin, R. I. (1987). Electronic markets and electronic hierarchies. *Communications of the ACM*, 30(6), 484-497.
- Papagelis, M., Plexousakis, D., & Kutsuras, T. (2005). Alleviating the sparsity problem of collaborative filtering using trust inferences. In *Trust management* (pp. 224-239). Springer, Berlin, Heidelberg.

- Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13(5-6), 393-408.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). *Application of dimensionality reduction in recommender system-a case study*. Minnesota Univ Minneapolis Dept of Computer Science.
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence, 2009*.
- Reischach, F., Guinard, D., Michahelles, F., & Fleisch, E. (2009). A mobile product recommendation system interacting with tagged products. In *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on* (pp. 1-6). IEEE.
- Αραβαντινός Αθανάσιος. *Σφάλματα και στατιστική επεξεργασία πειραματικών μετρήσεων. Γραφικές παραστάσεις, κλίση καμπύλης, μέθοδος των ελαχίστων τετραγώνων.*