Πανεπιστήμιο Πειραιώς Τμήμα Ψηφιακών Συστημάτων

Π.Μ.Σ. Διδακτική της Τεχνολογίας & Ψηφιακά Συστήματα

Τίτλος Διπλωματικής Εργασίας:

Συστήματα Συστάσεων Βασισμένα στην εμπιστοσύνη

Φοιτήτρια: Μαριάννα Σκιαδά

ΑΜ: ΜΕ/10094

Επιβλέπουσα: Επίκουρη Καθηγήτρια Μαρία Χαλκίδη

ΟΚΤΩΒΡΙΟΣ 2013

Πίνακας περιεχομένων

Περίληψη

Παράλληλα, με τη διείσδυση του Ίντερνετ γεννήθηκαν και τα Συστήματα Συστάσεων (*Recommender Systems*) που άλλαξαν την προσέγγιση του μάρκετινγκ όσο αφορά το ηλεκτρονικό εμπόριο, εξυπηρετώντας την ευκολότερη πρόσβαση στην πληροφορία και την παροχή εξατομικευμένων υπηρεσιών. Αυτά τα συστήματα έχουν προφανή επιρροή σε περιβάλλοντα όπου το μέγεθος των δεδομένων ξεπερνά τις δυνατότητες οποιουδήποτε χρήστη να τα εξερευνήσει πλήρως και έχουν γίνει αναπόσπαστο κομμάτι πολλών ηλεκτρονικών καταστημάτων όπως το *Amazon*, το *PandoraRadio* και το *Netflix*. Τα Συστήματα Συστάσεων μπορούν να υλοποιούν διαφορετικές τεχνικές με πιο δημοφιλή αυτή της Συνεργατικής Μεθόδου (C*ollaboration Filtering*). Κατά τη Συνεργατική Μέθοδο δομούνται γειτονιές χρηστών -βάσει της ομοιότητάς τους με τον χρήστη για τον οποίο επιθυμούμε να παραχθούν οι υποδείξεις - ενώ συχνά χρησιμοποιείται και η εμπιστοσύνη στη διαδικασία.

Όμως επειδή η χρήση αυτών των συστημάτων συνδέεται με αύξηση των πωλήσεων (πχ η Amazon υπολογίζει ότι το 35% των πωλήσεών της προέρχονται από υποδείξεις που παράγει προς τους χρήστες της) συχνά αποτελούν στόχο κακόβουλων οντοτήτων που επιθυμούν να εμποδίσουν την εύρυθμη λειτουργία τους. Κατά συνέπεια σκοπός ενός Συστήματος Συστάσεων είναι όχι μόνο η παραγωγή καλών (ορθών) υποδείξεων αλλά και υποδείξεων που δεν έχουν επηρεαστεί από κακόβουλες οντότητες (εύρωστων). Οι έννοιες της ομοιότητας –πόσο μοιάζουν δύο χρήστες βάσει των παρελθουσών αξιολογήσεων τους σε κοινά βαθμολογημένα αντικείμενα – και της εμπιστοσύνης –ο βαθμός που ένας χρήστης θεωρείται αξιόπιστος- επισημαίνονται στη βιβλιογραφία για τη συνεισφορά τους στην ορθότητα των υποδείξεων. Η ομοιότητα αν και αποτελεί την πιο κλασική επιλογή για την υλοποίηση της Συνεργατικής Μεθόδου δεν θεωρείται εύρωστη , δηλαδή δεν μπορεί να αντιμετωπίσει τις αρνητικές συνέπειες μίας κακόβουλης επίθεσης, ενώ σύμφωνα με τους *O' Donovan & Smyth, (2005)* η εμπιστοσύνη δύναται.

Στην παρούσα εργασία γίνεται ανασκόπηση των διαφορετικών τεχνικών που χρησιμοποιούν τα Συστήματα Συστάσεων με έμφαση σε αυτή της Συνεργατικής Μεθόδου λόγω της εφαρμογής της στα κοινωνικά δίκτυα *(social network*). Επίσης συζητούνται οι κακόβουλες επιθέσεις ώστε να γίνουν κατανοητές οι στρατηγικές και τεχνικές υλοποίησής τους με σκοπό την ανάπτυξη πιο εύρωστων συστημάτων.

Στην προσπάθεια ανάπτυξης εύρωστων συστημάτων υπήρξε εστίαση στην έννοια της εμπιστοσύνης και στην πρόταση ότι είναι εύρωστη η οποία στηρίζεται μόνο σε μία εργασία των *O' Donovan & Smyth, (2005)* κατά την οποία χρησιμοποιείται επικουρικά η έννοια σε εκείνη της ομοιότητας και ως ορίζουν ως αξιόπιστο το χρήστη που έχει ιστορικό καλών υποδείξεων. Στα πειράματα που πραγματοποιήθηκαν, χρησιμοποιώντας το έτοιμο dataset του opinions, χρησιμοποιήθηκαν παραλλαγές της υπολογίσιμης εμπιστοσύνης όπως αυτή ορίζεται από τους *O' Donovan & Smyth, (2005)* σε επίπεδο τοπικότητας (ανάπτυξη καθολικών και τοπικών μεταβλητών εμπιστοσύνης) αλλά και προτάσσοντας μια νέα μετρική κατά την οποία δεν αρκεί ένας χρήστης να έχει ιστορικό θετικών υποδείξεων (δηλαδή να αποτελεί μέλος γειτονιών που κάνουν επιτυχημένες προτάσεις) αλλά και να έχει συνεισφέρει θετικά σε αυτές.

Από την αξιολόγηση των αποτελεσμάτων προκύπτει ότι η εμπιστοσύνη μπορεί να συντελέσει στην αύξηση της ευρωστίας του συστήματος και της ορθότητας των υποδείξεων ειδικά όταν χρησιμοποιείται ως κριτήριο επιλογής των χρηστών που θα δομήσουν κάθε γειτονιά για την παραγωγή υποδείξεων. Ενώ και η πρόταση που γίνεται ώστε ένας χρήστης να πρέπει να έχει συνεισφέρει θετικά στις παραγόμενες υποδείξεις για να θεωρηθεί αξιόπιστος φαίνεται να έχει θετικά αποτελέσματα έναντι της ευρωστίας όσο της ορθότητας.

Executive Summary

Along with the Internet penetration Recommender Systems were born and changed the marketing approach as regards electronic commerce, easier access to information and personalized services. These systems have obvious influence in environments where data size exceeds the capabilities of any user to fully explore them and have become an integral part of many online stores such as *Amazon*, *Pandora, Radio* and *Netflix*. Recommender Systems may implement different techniques while more popular is Collaboration Filtering. During Collaborative Filtering neighborhoods of users are generated for each user based on their similarity with him whilst trust is also often used in the process to produce recommendations.

Because their use is associated with an increase in sales (eg the Amazon calculates that 35 % of its sales come from the suggestions generated for users) are often targeted by malicious entities wishing to prevent their proper function. Accordingly, the purpose of Recommender Systems is not only to produce good (accurate) suggestions but also suggestions that have not been affected by malicious entities (robust suggestions). The concepts of similarity - how

much two users look alike based on past assessments in common rated items - and trust - the degree to which a user can be reliable – are both indicated in the literature for their contribution to the accuracy of the recommendations . Though similarity is the most classic choice for the implementation of collaboration filtering, is not robust -does not cope with the negative consequences of a malicious attack –while trust is *O' Donovan & Smyth, (2005)*

In this thesis are presented the different techniques that be implemented in a Recommender System with emphasis on Collaborative Filtering because of its implementation in social networks. Malicious attacks are also discussed to understand the implemented strategies and techniques in order to develop more robust systems.

In the effort to develop robust systems we focused on the concept of trust and the suggestion that trust is robust which is based only on one paper by O' *Donovan & Smyth, (2005)* where its use is ancillary to similarity. They also define as a trusted user someone who has a history of good suggestions to others. In the experiments carried out , using the *opinions* dataset, variations of computable trust ,as defined by *O 'Donovan & Smyth, (2005),* were used concerning the range of trust (global and local variables of trust), but also putting forward a new trust metric in which it is not enough for a user to have history of positive suggestions (ie being part of neighborhoods that make successful suggestions) but he also has to have contributed positively to these .

The evaluation results show that trust can help to increase the robustness and accuracy of the suggestions especially when used as a criterion for selecting the users that will construct each neighborhood in order to generate suggestions. Also our proposal that a user should have contributed positively to the produced suggestions to be considered reliable, seems to have positive effects on system's robustness and accuracy.

1 Εισαγωγή

Τα τελευταία χρόνια έχει υπάρξει ένας μετασχηματισμός του προβλήματος λήψης αποφάσεων. Δηλαδή ενώ τις προηγούμενες δεκαετίες το πρόβλημα που αντιμετώπιζαν τα άτομα ήταν από ποιες πηγές και πώς θα μπορέσουν να εξάγουν χρήσιμες πληροφορίες που θα τους οδηγήσουν στη σωστή επιλογή πλέον η ερώτηση που απασχολεί είναι ποιες πληροφορίες πρέπει να χρησιμοποιήσουν καθώς το διαθέσιμο μέγεθος είναι τεράστιο. Η συνεισφορά του Ίντερνετ είναι θεμελιώδης στην ύπαρξη υπερπληθώρας διαθέσιμων πληροφοριών καθώς δίνει τη δυνατότητα επικοινωνίας ατόμων που η γεωγραφική απόστασή τους απαγορεύει τη φυσική επαφή και προσφέρει φθηνή αποθήκευση δεδομένων σε σχέση με τις παραδοσιακές τεχνικές.

Η τεχνολογία του Ιντερνέτ αναπτύσσεται ραγδαίως προσφέροντας γρηγορότερες συνδέσεις και ολοένα περισσότερα άτομα γίνονται καθημερινοί χρήστες του. Έτσι παραμερίζοντας τα τεχνικά θέματα το Ίντερνετ κατάφερε να πραγματοποιήσει μια επανάσταση που δεν αφορά τόσο την χρησιμοποιούμενη τεχνολογία αλλά το περιεχόμενο. Αυτή η επανάσταση του παγκόσμιου ιστού θεωρήθηκε ότι φέρνει τη δεύτερη έκδοσή για αυτό και ονομάστηκε Web 2.0 η οποία δεν αναφέρεται σε μια εξέλιξη των τεχνικών προδιαγραφών αλλά μάλλον στις αλλαγές του τρόπου δόμησης και χρήσης του περιεχομένων των ιστοσελίδων. (Tim O'Reilly (2005-09-30). "What Is Web 2.0"[1])

Ένας ιστοχώρος που υλοποιεί τη λογική του Web 2.0 επιτρέπει και συχνά παροτρύνει τους χρήστες τους εκτός από παθητικοί αναγνώστες του περιεχομένου του να αποτελέσουν και δημιουργούς του. Δηλαδή παρέχει τις κατάλληλες συνθήκες για επικοινωνία και ανταλλαγή απόψεων μεταξύ των χρηστών σε μια εικονική κοινότητα , δίνοντάς του ενεργητικό ρόλο στην διαμόρφωση της ύλης. Παραδείγματα του Web 2.0 περιλαμβάνουν δικτυακούς τόπους κοινωνικής δικτύωσης, blogs, wikis τοποθεσίες διαμοιρασμού βίντεο (youTube), forum κά

Παρόλα αυτά η ανάγκη για γρήγορες και ορθές αποφάσεις παραμένει και εντείνεται εν μέσω του τεράστιου όγκου διαθέσιμων πληροφοριών και της δυσκολίας παράλληλης διαχείρισής του. Αυτό το πρόβλημα καλούνται να επιλύσουν τα Συστήματα Συστάσεων εκμεταλλευόμενα την επανάσταση του παγκόσμιου ιστού όπου ο χρήστης εκτός από καταναλωτής πληροφορίας αποτελεί και παραγωγό της. Δηλαδή αποτελούν εργαλεία που βοηθούν κάθε χρήστη να λάβει τις σωστές αποφάσεις αναγνωρίζοντας από τα τεράστιο

[1] http://oreilly.com/web2/archive/what-is-web-20.html

σύνολο των διαθέσιμων πληροφοριών εκείνες που θα οδηγήσουν στην καλύτερη απόφαση προσομοιώνοντας την ανθρώπινη συμπεριφορά.

Τα άτομα πραγματοποιούν αποφάσεις που βασίζονται σε παρελθούσες αποφάσεις που είχαν θετικό αποτέλεσμα, εμπιστεύονται άτομα που γνωρίζουν σε σχέση με κάποιο άγνωστο και αδιαφανή στον τρόπο λειτουργίας του περίπλοκο αλγόριθμο και άνθρωποι με παρόμοια γούστα είναι πιο πιθανό να κάνουν παρόμοιες μελλοντικές *επιλογές* ***(****Sinha & Swearingen, 2001, Resnick et , 1994)****.* Προσομοιάζοντας επιτυχώς τέτοιες ανθρώπινες συμπεριφορές τα Συστήματα Συστάσεων αποτελούν πια αναπόσπαστο κομμάτι πολλών εμπορικών και μη sites (amazon, youTube, lastFm) και προσφέρουν σημαντικά οφέλη σε χρήστες και ιδιοκτήτες. Συγκεκριμένα χαρακτηρίζονται ως οι περισσότερο υποσχόμενες εφαρμογές στο ηλ. Εμπόριο *(Spiekermann & Paraschiv, 2002)* και οι επιτυχημένες εφαρμογές Τεχνητής Νοημοσύνης από τους *(Felfernig et al., 2007).*

Όμως επειδή η προσομοίωση της πραγματικότητας δεν πρέπει να αφορά μόνο τη θετική πλευρά της ανθρώπινης συμπεριφοράς δεν θα πρέπει να θεωρείται αμελητέα ή άνευ αξίας η προσπάθεια κάποιων για εξαπάτηση ώστε να αυξήσουν τα κέρδη τους (χρήματα, φήμη, κτλ). Ένα «καλό» Σύστημα Συστάσεων δεν θα πρέπει να έχει σκοπό την ενίσχυση των χρηστών στην διαδικασία λήψης αποφάσεων ώστε να κάνουν επιλογές που ικανοποιούν αποτελεσματικά τις ανάγκες τους αλλά θα πρέπει και να τους βοηθά στην επιλογή αμερόληπτων πληροφοριών.

1.1 Περιγραφή του προβλήματος

Η μεταφορά εννοιών όπως η ομοιότητα *(Sinha & Swearingen, 2001, Resnick et al. 1994)* και η εμπιστοσύνη στα Συστήματα Συστάσεων έχουν αποτελέσει πολύ επιτυχημένες και διαδεδομένες πρακτικές. Σκοπός τους η ενίσχυση της ορθής διήθησης των πληροφοριών ώστε οι πιο χρήσιμες να παρουσιάζονται στο χρήστη. Τα επιτυχημένα Συστήματα Συστάσεων έχουν συνδεθεί με αύξηση πωλήσεων προκαλώντας έτσι το ενδιαφέρον παραγώγων/προμηθευτών που επιθυμούν να αυξήσουν τις πωλήσεις και τη φήμη των προϊόντων τους. Σύμφωνα με τους *Celma & Lamere (2007)* στο NetFlix, το 75% των ταινιών που βαθμολογούνται προέρχονται από υποδείξεις ενώ οι προτεινόμενες ειδήσεις του Google παράγουν 38% περισσότερα κλικ ενώ η Amazon υπολογίζει ότι το 35% των πωλήσεών της προέρχονται από τις υποδείξεις που παράγει προς τους χρήστες της. Αυτά τα θετικά αποτελέσματα οδηγούν συχνά τους παραγωγούς σε κακόβουλες πρακτικές ώστε να

αυξήσουν τις προβολές που πραγματοποιούν τα προϊόντα τους στους χρήστες του συστήματος. Οι πρακτικές αυτές προσπαθούν να παρακάμψουν την ορθή λειτουργία των Συστημάτων Συστάσεων και να αλλοιώσουν τα αποτελέσματα της διεργασίας τους προς όφελος τους. Τέτοιο παράδειγμα εταιρείας που παράγει ψευδείς συστάσεις προς τους καταναλωτές προέκυψε τον Ιούνιο του 2001[2], όταν η Sony Pictures παραδέχθηκε ότι είχε χρησιμοποιήσει πλαστά αποσπάσματα από ανύπαρκτες κριτικές για να προωθήσει μια σειρά ταινιών που είχε πρόσφατα κυκλοφορήσει.

Οι βασικότερες κατηγορίες Συστήματα Συστάσεων είναι οι εξής:

- Συστήματα Συστάσεων βάση περιεχομένου (Content-Based Recommender Systems)

Τα εν λόγω συστήματα φιλτράρουν τις διαθέσιμες πληροφορίες με βάση τα χαρακτηριστικά και τα στοιχεία από τα οποία συνιστάται. Δηλαδή, αυτοί οι αλγόριθμοι προσπαθούν να υποδείξουν αντικείμενα που να είναι παρόμοια με αυτά που ένας χρήστης άρεσε στο παρελθόν (ή εξετάζει στο παρόν).

- Συστήματα Συστάσεων Συνεργατικής Μεθόδου (Collaborative Filtering Recommender Systems).

Τα συστήματα που υλοποιούν τη Συνεργατική Μέθοδο βασίζονται στη συλλογή και ανάλυση μιας μεγάλης ποσότητας πληροφοριών σχετικά με τις συμπεριφορές, τις δραστηριότητες και τις προτιμήσεις των χρηστών . Στη συνέχεια παράγουν προβλέψεις χρησιμοποιώντας αντικείμενα που άρεσαν σε όμοιους χρήστες, δηλαδή χρήστες με παρόμοια συμπεριφορά.

Πιο διαδεδομένη και με καλύτερα αποτελέσματα έχει αξιολογηθεί η Συνεργατική Μέθοδος καθώς μπορεί να προσφέρει υποδείξεις για όλους τους τύπους χρηστών, ακόμα και σε όσους έχουν ιδιαίτερες και εξεζητημένες προτιμήσεις. Η ανοικτή της φύση , δηλαδή το γεγονός ότι επιτρέπει χωρίς περιορισμούς την αξιολόγηση αντικειμένων από τους χρήστες της επιτρέπει να σχηματίσει πολλές και διαφορετικές ομάδες όμοιων χρηστών αλλά ταυτόχρονα την καθιστά και ευάλωτη σε κακόβουλες επιθέσεις που θέλουν να επηρεάσουν τα αποτελέσματα των υποδείξεων.. Επίσης άλλα προβλήματα αφορούν το πρόβλημα των νέων χρηστών, δηλαδή επειδή οι νέοι χρήστες δεν έχουν βαθμολογήσει αρκετά προϊόντα δεν μπορεί να δημιουργηθεί ένα προφίλ βάσει του οποίου θα αναζητηθούν άλλοι

[2] http://news.bbc.co.uk/1/hi/entertainment/lm/1368666.stm

παρόμοιοι χρήστες και τέλος η σποραδικότητα των δεδομένων δηλαδή ότι λίγα αντικείμενα βαθμολογεί κάθε προφίλ σε σχέση με το σύνολο που παρατίθενται από το σύστημα *(Massa & Avesani, 2004).*

Προσπάθειες για την αντιμετώπιση των προβλημάτων γίνονται με την εισαγωγή νέων μετρικών πέραν της ομοιότητας ώστε να υποβοηθηθεί η δυναμική του αλγορίθμου. Μια τέτοια μετρική είναι η εμπιστοσύνη η οποία φιλοδοξεί να βελτιώσει τις αποδόσεις της Συνεργατικής Μεθόδου σε όρους ορθότητας, κάλυψης αλλά και ευρωστίας.

1.2 Στόχοι της εργασίας

Οι στόχοι οι οποίοι επιδιώκεται να καλυφθούν από την παρούσα εργασία είναι οι εξής:

- *Βιβλιογραφική επισκόπηση των Συστημάτων Συστάσεων*.

Μελέτη των Συστημάτων Συστάσεων ως προς τη λειτουργία τους, τις διαφορετικές τεχνικές υλοποίησης, τα θετικά και αρνητικά σημεία τους. Ιδιαίτερο βάρος δίνεται στις δύο βασικές τεχνικές που χρησιμοποιούνται κατά κύριο λόγω στα συστήματα δηλαδή τη Συνεργατική Μέθοδο και τις τεχνικές που παράγουν υποδείξεις οι οποίες στηρίζονται στο περιεχόμενο.

- *Βιβλιογραφική επισκόπηση της εμπιστοσύνης στα κοινωνικά δίκτυα*.

Μια έννοια που χρησιμοποιείται στα Συστήματα Συστάσεων και ιδιαίτερα σε αυτά που υλοποιούν τη Συνεργατική Μέθοδο είναι η εμπιστοσύνη. Ένας από τους κύριους στόχους της εργασίας είναι η περιγραφή της εμπιστοσύνης και η ανάλυση των χαρακτηριστικών και της συμπεριφοράς της στα κοινωνικά συστήματα Συστάσεων που την εμπλέκουν στη διαδικασία παραγωγής των υποδείξεών προς τους χρήστες τους.

- *Βιβλιογραφική επισκόπηση των κακόβουλων επιθέσεων*

Τα Συστήματα Συστάσεων προσφέρουν σημαντικά οφέλη στους εμπλεκόμενους (χρήστες και ιδιοκτήτες) ενώ συνδέονται και με αύξηση πωλήσεων. Κατά συνέπεια συχνά γίνονται στόχος επιθέσεων που σκοπό έχουν την εμπόδιση της εύρυθμης λειτουργίας του. Έτσι η περιγραφή και η εξέταση των χαρακτηριστικών, στρατηγικών και τρόπων αντιμετώπισης αποτελεί βασικό σκοπό της εργασίας με απώτερο στόχο την κατανόηση του τρόπου εκτέλεσής τους και εύρεση τρόπων αντιμετώπισής τους.

- *Προσπάθεια βελτίωσης των αλγορίθμων που χρησιμοποιούν την εμπιστοσύνη στη διαδικασία παραγωγής υποδείξεων*.

Βασικός στόχος της παρούσας διπλωματικής εργασίας είναι η ανάδειξη νέων μετρικών εμπιστοσύνης με σκοπό την βελτίωση της απόδοσης των αλγορίθμων σε ορθότητα και ευρωστία.

- *Εξέταση του βαθμού δυνατότητας χαρακτηρισμού της εμπιστοσύνης ως μετρική που μπορεί να βελτιώσει την ανθεκτικότητα του αλγορίθμου Συστάσεων στις επιθέσεις.*

Ένας από τους κυριότερους σκοπούς της συγκεκριμένης εργασίας είναι η αποτίμηση της εμπιστοσύνης ως προς την ανθεκτικότητά της στις κακόβουλες εξωγενείς επιθέσεις. Στη βιβλιογραφία υπάρχουν αναφορές για θετική συνεισφορά των αξιόπιστων χρηστών αλλά χωρίς να έχει γίνει εκτεταμένη έρευνα.

1.3 Δομή της εργασίας

Στη συνέχεια περιγράφεται η δομή της εργασίας βάσει των κεφαλαίων που την αποτελούν:

1. Κεφάλαιο 1: Εισαγωγή στην εργασία συζητώντας τις βασικές έννοιες που θα απασχολήσουν στη συνέχεια. Περιγραφή των στόχων που καλείται να επιτύχει η εργασία και περιγραφή της δομής της.

2. Κεφάλαιο 2: Ανασκόπηση της βιβλιογραφίας που αφορά τα Συστήματα Συστάσεων. Δηλαδή ορίζεται και περιγράφεται η λειτουργίας τους, αναφέρονται τα οφέλη που προσφέρουν στους χρήστες τους, αναλύονται προβλήματα που εμφανίζουν και τέλος περιγράφονται οι βασικές τεχνικές υλοποίησης που ακολουθούνται.

3. Κεφάλαιο 3: Ανασκόπηση της έννοιας της εμπιστοσύνης όπως αυτή περιγράφεται από τη βιβλιογραφία και σχετίζεται με τα κοινωνικά δίκτυα. Αναλύεται η εμπιστοσύνης βάσει των διαστάσεών της , περιγράφονται αναλυτικά οι διάφορες μετρικές της και τρόποι εξόρυξής της από το δίκτυο. Τέλος περιγράφονται κλασικές τεχνικές που τη συμπεριλαμβάνουν στην παραγωγή υποδείξεων και δημοφιλείς υλοποιήσεις.

4. Κεφάλαιο 4: Ανασκόπηση των επιθέσεων που εξαπολύονται στα Συστήματα Συστάσεων. Παρατίθενται ορισμοί από τη βιβλιογραφία και περιγράφονται οι διαστάσεις, οι στρατηγικές, οι τεχνικές των επιθέσεων καθώς και η μορφή των προφίλ επίθεσης. Το κεφάλαιο κλείνει με την αναφορά σε τεχνικές αντιμετώπισης των επιθέσεων.

5. Κεφάλαιο 5: Ανάλυση της πειραματικής αξιολόγησης που πραγματοποιήθηκε. Συγκεκριμένα γίνεται περιγραφή τω βημάτων που ακολουθήθηκαν για την δημιουργία των πειραμάτων που ελέγχουν την αποτελεσματικότητα της εμπιστοσύνης (μέσω χρήσης διαφορετικών μετρικών της) σε συστήματα που υλοποιούν τη Συνεργατική Μέθοδο και έχουν δεχθεί επίθεση. Στη συνέχεια περιγράφονται οι τρόποι αξιολόγησης των αποτελεσμάτων και σχολιασμός των αποτελεσμάτων.

6. Κεφάλαιο 6: Σύνοψη της εργασίας με παρουσίαση των συμπερασμάτων που προέκυψαν και την αναφορά παρατηρήσεων για μελλοντική έρευνα

7. Κεφάλαιο 7: Τέλος η εργασία κλείνει με την βιβλιογραφία πάνω στην οποία στηρίχθηκε η συγγραφή της εργασίας.

2 Συστήματα Συστάσεων

2.1 Εισαγωγή

Από τα μέσα της δεκαετίας του 1990 η χρήση του Ίντερνετ άρχισε να διαδίδεται με πολύ γρήγορους ρυθμούς. Όλο και περισσότεροι άνθρωποι το χρησιμοποιούσαν σε καθημερινή βάση και αποκτούσαν ευρυζωνικές συνδέσεις ώστε να πλοηγούνται με μεγαλύτερη ταχύτητα σε αυτό. Αντιλαμβανόμενοι την ευκαιρία που ανοιγόταν αυξανόμενος αριθμός καταστημάτων άρχισε να αποκτά ηλεκτρονική παρουσία και κατάφεραν όχι μόνο να ανταγωνιστούν τα καταστήματα με φυσική παρουσία αλλά και να τα ξεπεράσουν δημιουργώντας καινοτόμες επιχειρήσεις που έσπασαν το γεωγραφικό όριο των παραδοσιακών επιχειρήσεων πχ Amazon

Παράλληλά με την άνοδο της διείσδυσης του Ίντερνετ γεννήθηκαν και τα συστήματα υποδείξεων (recommender systems) που ουσιαστικά άλλαξαν την προσέγγιση του

μάρκετινγκ όσο αφορά το ηλεκτρονικό εμπόριο. Ενώ η ηλεκτρονική υπόσταση δίνει τη δυνατότητα προσφοράς πολύ μεγάλου όγκου πληροφορίας που σε ένα φυσικό κατάστημα θα επέφερε μεγάλο κόστος έχει το μειονέκτημα της έλλειψης προσωπικής επαφής. Δηλαδή δεν υπάρχει ο κλασικός υπάλληλος που θα προσφερθεί να εξυπηρετήσει και συνδυάζοντας το εμπόρευμα και τις απαιτήσεις του χρήστη θα προσφέρει προσωποποιημένες λύσεις. Αυτό το κενό έρχεται να λύσει η ύπαρξη των Συστημάτων Συστάσεων. Δηλαδή ο χρήστης δεν χρειάζεται να κάνει εξαντλητική αναζήτηση στον ηλεκτρονικό κατάλογο του καταστήματος , ούτε να χαθεί στην ογκώδη πληροφορία ψάχνοντας για κάποιο αντικείμενο με συγκεκριμένα χαρακτηριστικά. Τα Συστήματα Συστάσεων αναλαμβάνουν να προτείνουν επιλογές στο χρήστη φιλτράροντας ένα μεγάλο όγκο δεδομένων βάσει διαφορετικών κριτηρίων όπως δημογραφικά στοιχεία, παλαιότερες επιλογές , προτιμήσεις κα

Τα Συστήματα Συστάσεων ή Υπόδειξης (Recommender Systems) είναι εργαλεία λογισμικού που συμπεριλαμβάνουν τεχνικές οι οποίες προσφέρουν υποδείξεις αντικειμένων στους χρήστες του . Οι υποδείξεις στηρίζονται σε διαφορετικές τεχνικές –μοντέλα- που θα αναλυθούν στη συνέχεια της εργασίας. Τα αντικείμενα που προτάσσονται στους χρήστες μπορεί να είναι οποιασδήποτε φύσης όπως μουσικά κομμάτια, ταινίες, προϊόντα προς αγορά, ειδήσεις, άρθρα κτλ. Συνήθως κάθε Σύστημα Υπόδειξης εστιάζει σε μία κατηγορία προϊόντων χρησιμοποιώντας καταλλήλως τα ιδιαίτερα χαρακτηριστικά του (πχ ένα μουσικό κομμάτι μπορεί να ακουσθεί) να προσφέρει στο μέγιστο βαθμό τις υπηρεσίες του, δηλαδή χρήσιμες και αποδοτικές προτάσεις.

Αρχικά περιγράφηκαν από τους (*Resnick & Varian 1997*) ως συστήματα στα οποία οι χρήστες προσφέρουν προτάσεις ως δεδομένα εισόδου τα οποία αθροίζονται από το σύστημα και στη συνέχεια κατευθύνονται στους κατάλληλους χρήστες .

Νεώτερος και ευρύτερος ορισμός δόθηκε από τον *Burke (2002)* ο οποίος τα περιγράφει ως ένα σύστημα που παράγει προσωποποιημένες προτάσεις ως δεδομένα εξόδου ή έχει ως αποτέλεσμα την οδήγηση των χρηστών μέσω ενός προσωποποιημένου δρόμου σε ενδιαφέροντα ή χρήσιμα αντικείμενα ανάμεσα σε ένα μεγάλο χώρο πιθανών επιλογών.

Αυτά τα συστήματα έχουν προφανή επιρροή σε περιβάλλοντα όπου το μέγεθος των δεδομένων ξεπερνά τις δυνατότητες οποιουδήποτε χρήστη να το τις εξερευνήσει διεξοδικά**.** *(Schafer, Konstan & Riedl, 1999)* και έχουν γίνει αναπόσπαστο κομμάτι πολλών

ηλεκτρονικών καταστημάτων. Ενδεικτικά στη συνέχεια παρατίθενται κάποια ιδιαιτέρως γνωστά sites που στηρίζουν τη λειτουργία τους σε Συστήματα Υπόδειξης:

1. **Amazon** – βάσει του ιστορικού αγορών και προβολή αντικειμένων του χρήστη το σύστημα προτάσσει αντίστοιχα αντικείμενα
2. **Tripadvisor** – επιτρέπει το σχολιασμό των αντικειμένων από τους χρήστες και στη συνέχεια παράγει υποδείξεις βάσει της τρέχουσας αναζήτησης
3. **P&oraRadio** – παίζει μουσική με παρόμοια χαρακτηριστικά της αρχικής επιλογής του χρήστη
4. **Youtube –** προτάσσει παρόμοια αντικείμενα βάσει παλαιότερων αναζητήσεων και της τρέχουσας
5. **Netflix –** προτείνει ταινίες που πιθανώς να βρει ενδιαφέρουσες ο χρήστης βάσει παλαιότερων βαθμολογιών του και του ιστορικού επισκέψεών του στο σύστημα. Επίσης λαμβάνει υπόψιν και δημογραφικά χαρακτηριστικά όπως πχ το φύλο.
6. **IMDb –** προτείνει αντικείμενα βάσει των χαρακτηριστικών της τρέχοσας αναζήτησης

Το αποτέλεσμα της διεργασίας των Συστημάτων Υπόδειξης είναι τις περισσότερες φορές μια λίστα αντικειμένων που έχουν παραχθεί έτσι ώστε να ταιριάζουν στις ανάγκες του κάθε χρήστη. Δηλαδή η συγκεκριμένη λίστα έχει παραχθεί για συγκεκριμένο χρήστη κάθε φορά χρησιμοποιώντας πληροφορίες όπως παρελθούσες αγορές, ιστορικό πλοήγησης, βαθμολογίες, δημογραφικά στοιχεία κά.

Σύμφωνα με τους (*Resnick & Varian, 1997*) -οι οποίοι έδωσαν και τον πρώτο ορισμό των Συστημάτων Υπόδειξης- στην καθημερινότητά τους οι άνθρωποι πολύ συχνά στηρίζονται στις προτάσεις -υποδείξεις- άλλων ατόμων για να πάρουν αποφάσεις. Τέτοια παραδείγματα είναι οι συστατικές επιστολές που δίνονται σε υπάλληλους και φοιτητές, οι κριτικές εστιατορίων, βιβλίων , ταινιών κά. Έτσι τα Συστήματα Υπόδειξης αποτελούν μία μεταφορά της καθημερινότητας στον ψηφιακό κόσμο.

Στα πρώτα βήματά του το Ίντερνετ γνώρισε μεγάλη άνθηση λόγω της δυνατότητας που προσφέρει για αποθήκευση μεγάλου όγκου πληροφορίας με χαμηλό κόστος. Έτσι δίνεται η ευκαιρία στους χρήστες του λαμβάνουν αποφάσεις έχοντας όλες τις πληροφορίες που είναι απαραίτητες για τη σφαιρική και εις βάθος αξιολόγηση του προβλήματος. Όμως τα

τελευταία χρόνια αυτή η δυνατότητα ανάκτησης τόσο μεγάλου όγκου σχετικών και πιθανώς χρήσιμων πληροφοριών για κάθε ζήτημα έχει οδηγήσει στο λεγόμενο πρόβλημα της «Υπερχείλισης Πληροφορίας» (information overload). Οι χρήστες λόγω της υπερπληθώρας πληροφορίας που υπάρχει στα διάφορα sites δεν μπορούν να εντοπίσουν εκείνα τα αντικείμενα που ικανοποιούν καλύτερα τις ανάγκες τους. Δηλαδή ενώ το να έχει ο χρήστης εναλλακτικές λύσεις ανάμεσα από τις οποίες θα πρέπει επιλέξει την καλύτερη είναι χρήσιμο και επιθυμητό, η υπερπληθώρα εναλλακτικών μπορεί να δημιουργήσει σύγχυση στο χρήστη και να οδηγεί σε επιλογές που δεν ικανοποιούν αποτελεσματικά την ανάγκη τους.

Τα Συστήματα Υπόδειξης έχουν αποδειχθεί ένα χρήσιμο εργαλείο για την αντιμετώπιση του προβλήματος υπερχείλισης της πληροφορίας καθώς σκοπός τους είναι να διηθίσουν την προσφερόμενη από το σύστημα πληροφορία και αποπειράται να παρουσιάσει πληροφοριακά αντικείμενα (όπως ταινίες, μουσική, διαδικτυακοί τόπους, νέα) που πιθανόν να ενδιαφέρουν το χρήστη. *(Wen, 2008)*

Ένα ερώτημα που γεννάται είναι γιατί δεν θεωρούνται τα συστήματα υπόδειξης κομμάτι μίας μηχανής αναζήτησης ή ενός συστήματος εξόρυξης δεδομένων. Μία μηχανή αναζήτησης κάνει ταίριασμα, δηλαδή το σύστημα επιστρέφει ένα σύνολο αντικειμένων που ταιριάζουν στο περιεχόμενο της αναζήτησης βάσει του βαθμού ομοιότητας. Έτσι ως κριτήρια διαφοροποίησης των συστημάτων υποδείξεων από τα παραπάνω είναι η εξατομίκευση στις ανάγκες των χρηστών και η χρησιμότητα με το βαθμό ενδιαφέροντος του προτεινόμενου αντικειμένου από το χρήστη. (*Burke 2000)*

Η μελέτη των Συστημάτων Υπόδειξης είναι σχετικά νέα σε σύγκριση με άλλα εργαλεία και τεχνικές πληροφοριακών συστημάτων όπως οι βάσεις δεδομένων και οι μηχανές αναζήτησης. Τα συστήματα υπόδειξης άρχισαν να μελετώνται ως ανεξάρτητος τομέας από τα μέσα της δεκαετίας του 1990 και οι ρίζες τους μπορούν να αναχθούν στην Ανάκτηση Πληροφοριών (information retrieval*) (Salton 1989)*, στις Προβλεπτικές Θεωρίες (forecasting theories) (Armstrong 2001), καθώς επίσης και στο μοντέλο επιλογής των καταναλωτών όπως αναπτύχθηκε στο χώρο του μάρκετινγκ *(Lilien et al. 1992).* Επίσης βασίζονται σε διάφορες τεχνολογίες όπως η διήθηση πληροφοριών (μηχανές αναζήτησης), η μηχανική μάθηση και σε τεχνολογίες που προσαρμόζουν και εξατομικεύουν ένα σύστημα με σκοπό τη μοντελοποίηση κάθε χρήστη.

2.2 Οφέλη χρήσης Συστημάτων Συστάσεων

Όπως είπαμε τα συστήματα συστάσεων χρησιμοποιούνται από στο ηλεκτρονικό εμπόριο ώστε να προτείνουν στους χρήστες προϊόντα. ή/και υπηρεσίες. Τα προϊόντα αυτά μπορεί να προταθούν βάσει των συνολικών πωλήσεων του site , των δημογραφικών στοιχείων του χρήστη, της ανάλυσης των παλαιότερων αγορών του και της παρελθούσης συμπεριφοράς του. Αυτές οι τεχνικές τελικά είναι κομμάτι της προσπάθειας προσφοράς προσωποποιημένων υπηρεσιών από την ιστοσελίδα καθώς βοηθούν την προσαρμογή στα ιδιαίτερα χαρακτηριστικά του κάθε πελάτη. Τα συστήματα συστάσεων αυτοματοποιούν την προσωποποίηση στον ιστοχώρο επιτρέποντας την εξατομικευμένη προσωποποίηση σε κάθε πελάτη. Όπως πολύ στοχευμένα είπε ο CEO της Amazon , Jeff Benos «*Αν έχω τρία εκατομμύρια πελάτες στο διαδίκτυο θα πρέπει να έχω και τρία εκατομμύρια ηλεκτρονικά καταστήματα*».

Από τη χρήση των Συστημάτων συστάσεων προκύπτουν οφέλη και κέρδη τόσο για τον πάροχο του συστήματος όσο και για το χρήστη.

Συγκεκριμένα τα οφέλη για τους προμηθευτές τέτοιων συστημάτων σχετίζονται *με (Kantor, Ricci, Rokach, & B. Shapira, 2010)*:

- **Αύξηση των πωλήσεων**

Η κερδοφορία που προέρχεται από τις επιπλέον πωλήσεις σε σχέση με αυτές που θα πραγματοποιούνταν χωρίς την ύπαρξη του συστήματος συστάσεων προφανώς και αποτελεί το μεγαλύτερο όφελος. Όμως και για ένα μη εμπορικό site (πχ ηλεκτρονική εφημερίδα) παρόμοιο όφελος προκύπτει και για αυτό καθώς στοχεύει στην αύξηση του αριθμού των άρθρων-ειδήσεων που θα αναγνωσθούν από τους χρήστες του διαδικτυακού τόπου αυξάνοντας έτσι τις «πωλήσεις» του εν λόγω ιστοχώρου.

Γενικά θα μπορούσαμε να ισχυριστούμε ότι σκοπός των παρόχων είναι η αύξηση ότι από τη μεριά των πάροχων υπηρεσιών είναι η μετατροπή των επισκεπτών σε «αγοραστές». Η χρήση του όρου αγοραστής γίνεται με μεγαλύτερη ευρύτητα, δηλαδή δεν αφορά μόνο την αγορά προϊόντων με χρήση χρημάτων αλλά την ώθηση του χρήστη στην εκτέλεση της επιθυμητής ενέργειας πάνω στη σελίδα μας. Αυτή η ενέργεια μπορεί να αφορά την αγορά

προϊόντων ή την ανάγνωση ενός ακόμα άρθρου πέραν εκείνου που σε οδήγησε στη συγκεκριμένη ιστοσελίδα.

- **Αγορά διαφορετικών αντικειμένων**

Ένα άλλο σημαντικό όφελος που προκύπτει από τη χρήση των Συστημάτων Συστάσεων είναι ότι δίνουν την δυνατότητα εύρεσης στο χρήστη αντικειμένων που θα ήταν δύσκολο να το κάνει χωρίς την εξατομικευμένη υπόδειξη. Παραδείγματος χάριν σε ένα σύστημα συστάσεων βιβλίων στόχος του διαχειριστή είναι η πώληση όλων των βιβλίων κι όχι μόνο των δημοφιλών. Η τυχαία διαφήμιση των αντικειμένων θα ήταν άσκοπη και δαπανηρή ενώ η στοχευμένη διαφήμιση τέτοιων αντικειμένων σε χρήστες που βάσει του προφίλ τους μπορεί να ενδιαφέρονται μπορεί να αποτελέσει αποτελεσματική πρακτική.

- **Αύξηση ικανοποίησης των χρηστών**

Ένα Σύστημα Συστάσεων που έχει αναπτυχθεί ορθά μπορεί πέραν της αύξησης των αγορών να βελτιώσει συνολικά την εμπειρία του χρήστη. Δηλαδή θα εκτιμήσει το σύστημα ως ενδιαφέρον, αποτελεσματικό και εύχρηστο καθιστώντας το ευχάριστο στη χρήση . Ο συνδυασμός των καλών υποδείξεων και του ευχάριστου περιβάλλοντος θα κάνει θετική την εμπειρία χρήσης του συστήματος και θα οδηγήσει σε αύξηση της πιθανότητας χρησιμοποίησης του στο μέλλον και ευκολότερης αποδοχής των προτάσεών του.

- **Αύξηση πιστότητας των χρηστών**

Όταν ένας διαδικτυακός τόπος αναγνωρίζει έναν παλαιό χρήστη και του συμπεριφέρεται μοναδικά βάσει του προφίλ του τότε αυξάνει και την πιστότητα του προς αυτό. Η συγκεκριμένη ιδιότητα των Συστημάτων Συστάσεων, αποτελεί ένα από τα κύρια και σημαντικότερα χαρακτηριστικά τους, καθώς οι υποδείξεις προκύπτουν από την αξιοποίηση της πληροφορίας που αποκτήθηκε από τον χρήστη έμμεσα (ιστορικό πλοήγησης) ή / και άμεσα (βαθμολογίες αντικειμένων). Κατά συνέπεια όσο περισσότερο χρησιμοποιεί κάθε χρήστης το σύστημα και αλληλεπιδρά με αυτό τόσο βελτιώνεται η μοντελοποίηση του προφίλ του και κατ΄ επέκταση το αποτέλεσμα της διεργασίας συστάσεων πλησιάζει περισσότερο τις προσωποποιημένες ανάγκες του.

- **Καλύτερη κατανόησης των αναγκών των χρηστών**

Άλλη μια σημαντική λειτουργία των Συστημάτων Συστάσεων τα αποτελέσματα της οποίας μπορούν να χρησιμοποιηθούν από διάφορες άλλες εφαρμογές, είναι η περιγραφή των προτιμήσεων του χρήστη οι οποίες μπορεί να συλλέγονται άμεσα είτε να παράγονται από

το σύστημα Αυτή η γνώση μπορεί να φανεί ιδιαιτέρως χρήσιμη στη διαχείριση της εφοδιαστικής αλυσίδας δηλαδή στην καλύτερη διαχείριση των αποθεμάτων ή της παραγωγής προϊόντων.

Ένα επιτυχημένο Σύστημα Συστάσεων μπορεί σύμφωνα με τα παραπάνω να προσφέρει σημαντικά οφέλη στον πάροχό του αλλά η επιτυχία του κρίνεται από τη σχέση που αναπτύσσει με τον χρήστη του. Δηλαδή θα πρέπει να προσφέρει και στο χρήστη του σημαντικά οφέλη που θα οδηγήσουν στην αποτελεσματική ικανοποίηση των αναγκών του. Ουσιαστικά κάθε σύστημα καλείται να ικανοποιήσει και τα δύο ενδιαφερόμενα μέρη ισορροπώντας τη λειτουργία καταλλήλως.

Οι *Herlocker et al., (2000)* καταγράφουν έντεκα τα σημεία που ένα Σύστημα Συστάσεων πρέπει να υλοποιεί ώστε να έχει θετικό αντίκτυπο η αλληλεπίδρασή τους με τους χρήστες. Κάποια σημεία μπορούν να χαρακτηρισθούν ως κύριες ή βασικές λειτουργίες που πρέπει να υλοποιεί όπως το να προσφέρει χρήσιμες υποδείξεις αντικειμένων προς τους χρήστες ενώ άλλοι μπορεί να θεωρηθούν ως «οπορτουνιστικοί». Κάτι παρόμοιο μπορεί να θεωρηθεί ότι πραγματοποιείται και από μία μηχανή αναζήτησης, όπου ο πρωταρχικός στόχος της μεν είναι να εντοπίσει αντικείμενα που να σχετίζονται με την αναζήτηση του χρήστη αλλά ταυτόχρονα μπορεί να ελέγξει την αξία μιας ιστοσελίδας (κοιτάζοντας την θέση της σελίδας στη λίστα αποτελεσμάτων αναζήτησης) ή για να ανακαλύψει διαφορετικές χρήσεις μίας λέξης σε ένα σύνολο εγγράφων. Συγκεκριμένα αναφέρονται στις παρακάτω λειτουργίες:

- **Εύρεση κάποιων καλών αντικειμένων**

Υπόδειξη υπό τη μορφή ταξινομημένης λίστας αντικειμένων που θα πιθανολογείται ότι θα αρέσουν στο χρήστη , συνοδευόμενη με την προβλεπόμενη βαθμολογία για κάθε στοιχείο της λίστας. Ουσιαστικά αυτή η λειτουργία αποτελεί και τη βασική υπηρεσία που προσφέρουν τα εν λόγω συστήματα όπου ενσωματώνονται.

- **Εύρεση όλων καλών αντικειμένων**

Υπόδειξη όλων των αντικειμένων που μπορούν να ικανοποιήσουν τις ανάγκες του χρήστη. Αυτή η λειτουργία έχει μεγαλύτερη αξία όταν ο αριθμός των αντικειμένων είναι σχετικά

περιορισμένος ή σε περιπτώσεις ιατρικών ή χρηματοοικονομικών συστημάτων που η λεπτομέρεια και πληρότητα των προτάσεων είναι κριτικής σημασίας.

- **Σχολιασμός πλαισίου**

Δεδομένου ενός υπάρχοντος πλαισίου παραδείγματος χάριν μίας λίστας αντικειμένων προς υπόδειξη, επισήμανση συγκεκριμένων αντικειμένων της λίστας η επιλογή των οποίων γίνεται βάσει μακροχρόνιων προτιμήσεων που έχει επιδείξει ο χρήστης κατά την αλληλεπίδρασή του με το σύστημα.

- **Υπόδειξη ακολουθίας**

Αντίθετα με την κλασική πρακτική παραγωγής ενός αντικειμένου προς υπόδειξη , παραγωγή μίας ακολουθίας αντικειμένων που θα ικανοποιήσουν την ανάγκη του χρήστη ως σύνολο. Παράδειγμα τέτοιας λειτουργίας είναι η υπόδειξη ολόκληρου playlist από το youtube μετά την αναζήτηση ενός τραγουδιού ή κάποιου βίντεο.

- **Υπόδειξη δέσμης**

Υπόδειξη ενός συνόλου αντικειμένων που συνδυάζονται μεταξύ τους επιτυχώς. Παραδείγματος χάριν ένα ταξίδι μπορεί να αποτελείται από διαφορετικούς προορισμούς, αξιοθέατα και καταλύματα που βρίσκονται σε μια συγκεκριμένη περιοχή. Από την πλευρά του χρήστη αυτές οι μεμονωμένες εναλλακτικές που του δίνονται μπορούν να θεωρηθούν και να επιλεχθούν ως ένας συνολικός προορισμός.

- **Απλή πλοήγηση**

Η συγκεκριμένη λειτουργία πραγματοποιείται όταν ο χρήστης περιηγείται στον κατάλογο με τα αντικείμενα χωρίς την άμεση πρόθεση να αγοράσει κάποιο από αυτά. Το Σύστημα Υπόδειξης πρέπει να βοηθήσει το χρήστη να περιηγηθεί σε αντικείμενα που είναι πιθανότερο να συμπίπτουν με τις προτιμήσεις και τα γούστα του κατά τη συγκεκριμένη πλοήγηση.

- **Εύρεση αξιόπιστης πηγής υπόδειξης**

Συχνά οι κάποιοι χρήστες δεν εμπιστεύονται τα Συστήματα Υπόδειξης και για αυτό πειραματίζονται με τη λειτουργία τους ώστε να ανακαλύψουν πόσο καλές προτάσεις παράγουν. Μερικά συστήματα μάλιστα δίνουν τη δυνατότητα στους χρήστες να ελέγξουν τον τρόπο λειτουργίας τους με σκοπό να μειώσουν τις αντιστάσεις και την πιθανή καχυποψία που μπορεί να αισθάνονται.

- **Χτίσιμο του προφίλ του χρήστη**

Αποτελεί μια βασική λειτουργία των Συστημάτων Υπόδειξης και συνδέεται με τη δυνατότητα του χρήστη να τροφοδοτεί το σύστημα με πληροφορίες για το τι του αρέσει και τι όχι. Κατά αυτόν τον τρόπο οι παραγόμενες υποδείξεις είναι εξατομικευμένες και προκύπτουν από τις επεξεργασία των εισερχόμενων δεδομένων. Αν λοιπόν το σύστημα δεν διαθέτει τέτοιου είδους γνώση για το άτομο δεν μπορεί να παράξει εξατομικευμένη πληροφορία και για αυτό κάνει προτάσεις που βασίζονται στις προτιμήσεις του μέσου χρήστη.

- **Αυτοέκφραση**

Αυτή η λειτουργία σχετίζεται με το γεγονός ότι κάποιο χρήστες δεν ενδιαφέρονται να λαμβάνουν υποδείξεις από ένα σύστημα αλλά τους αρέσει να μπορούν να διατυπώνουν και να εκφράζουν τις απόψεις τους για αντικείμενα μέσω βαθμολογιών και σχολίων.

- **Παροχή βοήθειας προς άλλους χρήστες**

Κάποιοι χρήστες νιώθουν χαρά να συνεισφέρουν με πληροφορίες για τα διάφορα αντικείμενα καθώς έτσι πιστεύουν ότι προσφέρουν στην κοινότητα. Αυτή η ανάγκη κάποιων χρηστών θα μπορούσε να αποτελέσει ένα σημαντικό κίνητρο για την εισαγωγή πληροφορίας σε συστήματα υπόδειξης που δεν χρησιμοποιούνται συστηματικά. Παραδείγματος χάριν έστω ένα Σύστημα Υποδείξεων αυτοκινήτων. Όταν ένας χρήστης αγοράσει ένα αυτοκίνητο και το βαθμολογήσει μέσω του εν λόγω συστήματος το κάνει γιατί πιστεύει ότι θα βοηθήσει άλλους κυρίως άλλους χρήστες χωρίς να διευκολύνει τον εαυτό του για την επόμενη μελλοντική αγορά του.

- **Άσκηση επιρροής προς άλλους χρήστες**

Τέλος υπάρχουν και χρήστες που συμμετάσχουν ενεργά στα Συστήματα Υπόδειξης καθώς έχουν στόχο να επηρεάσουν άλλους χρήστες για την αγορά συγκεκριμένων προϊόντων. Κάποιοι από αυτούς μάλιστα είναι κακόβουλοι καθώς προσπαθούν επί τούτου να αυξήσουν ή αντίστοιχα να μειώσουν τη δημοτικότητα επιλεγμένων προϊόντων προς δικό τους όφελος. Η κατηγορία των κακόβουλων χρηστών θα μας απασχολήσουν κατά κόρον στην εργασία. Ο ρόλος τους και οι πρακτικές τους θα αναλυθούν εις βάθος και θα αξιολογηθούν εναλλακτικοί τρόποι αντιμετώπισής τους.

Συμπληρώνοντας τα παραπάνω και οι (*Schafer et al, 1999)* ισχυρίζονται ότι τα Συστήματα Υπόδειξης ενισχύουν τις πωλήσεις στο ηλεκτρονικό εμπόριο με τρεις τρόπους:

- **Μετατροπή χρηστών σε αγοραστές:** Επισκέπτες μίας ιστοσελίδας συχνά περιηγούνται σε αυτή χωρίς ποτέ να αγοράζουν κάτι. Τα συστήματα υπόδειξης μπορούν να βοηθήσουν τους πελάτες ώστε να βρουν προϊόντα που επιθυμούν να αγοράσουν.

- **Cross-sell:** Τα συστήματα υπόδειξης βελτιώνουν τις σταυροειδής πωλήσεις προτείνοντας πρόσθετα προϊόντα προς αγορά στον πελάτη. Αν οι προτάσεις είναι καλές το μέγεθος της μέσης παραγγελίας αυξάνεται. Για παράδειγμα ένας ιστοχώρος μπορεί να προτείνει προϊόντα στον πελάτη κατά τον έλεγχο του καλαθιού του και λίγο πριν την πληρωμή βάσει των προϊόντων που ήδη έχει στο καλάθι του.

- **Πιστότητα:** Τα ηλεκτρονικά καταστήματα λειτουργούν σε ένα ιδιαίτερα ανταγωνιστικό περιβάλλον όπου η μετάβαση από το ένα κατάστημα στο άλλο είναι πολύ εύκολη. Για αυτό η ανάπτυξη πιστότητας από τους πελάτες είναι καθοριστικής σημασίας παράγοντας. *(Reichheld & Sesser, 1990, Reichheld, 1993).* Τα συστήματα υπόδειξης βελτιώνουν την πιστότητα δημιουργώντας μία σχέση προστιθέμενης αξίας μεταξύ του ηλεκτρονικού καταστήματος και του πελάτη. Οι ιστοσελίδες επενδύουν χρήματα ώστε να γνωρίσουν καλύτερα το προφίλ των πελατών τους. Χρησιμοποιούν τα συστήματα υπόδειξης για να εφαρμόσουν τη γνώση τους και υλοποιούν διαφοροποιημένες διεπαφές βάσει των αναγκών των εξατομικευμένων αναγκών του πελάτη. Οι πελάτες δείχνουν την θετικότητα τους προς τα καταστήματα που τους καλύπτουν και τους αντιπροσωπεύουν περισσότερο επισκέπτοντάς τα ξανά . Όσο περισσότερο ένας πελάτης χρησιμοποιεί ένα σύστημα υπόδειξης τόσο περισσότερο μαθαίνει πληροφορίες για το προφίλ του πελάτη και κατά επέκταση προτείνει και αντικείμενα καλύτερης ποιότητας προς τον πελάτη και συνεπώς αυξάνει την πιστότητα του πελάτη. Ακόμα και αν ένας ανταγωνιστής εξοπλιστεί με ένα όμοιο σύστημα υποδείξεων δεν θα μπορέσει να τραβήξει τον πιστό πελάτη καθώς εκείνος έχει επενδύσει χρόνο και ενέργεια στο αρχικό (Pine, et al. 1995).Τέλος πιστότητα δημιουργείται και μέσω ανάπτυξης σχέσεων μεταξύ πελατών. Δηλαδή ένας πελάτης θα επιστρέψει σε ένα ιστοχώρο που του προτείνει επικοινωνία με άτομα που επιθυμούν την αλληλεπίδραση με αυτόν.

Στη συνέχεια παρουσιάζονται τα οφέλη για τελικούς χρήστες και παρόχους από τη χρήση των συστημάτων υποδείξεων: (*Cremoneci et al 2012*)

- Οι καταναλωτές μπορεί να χαθούν μέσα στο μεγάλο όγκο πληροφοριών, προϊόντων και προσφερόμενων υπηρεσιών. Ένα τέτοιο σύστημα μπορεί να τους βοηθήσει στη μείωση της υπερπληροφόρησης, στη διευκόλυνση της αναζήτησης πληροφοριών, προϊόντων και υπηρεσιών, στην ευκολότερη και γρηγορότερη αναγνώριση των αντικειμένων και πληροφοριών που τους ενδιαφέρουν και στην αύξηση της ποιότητας της διαδικασίας λήψης αποφάσεων.

- Από την πλευρά του παρόχου τα οφέλη σχετίζονται με την έννοια της πειστικότητας (persuasivness), τη δυνατότητα δηλαδή να επηρεάσουν τη στάση, τις πεποιθήσεις, τις αποφάσεις και τις συμπεριφορές των χρηστών. Διευκολύνοντας την πρόσβαση στις πληροφορίες και συνδέοντας κομμάτια πληροφορίας που σχετίζονται ένα recommendation system μπορεί να επηρεάσει θετικά τη στάση του χρήστη έναντι της εφαρμογής και να δημιουργήσει μία σχέση με διάρκεια που στηρίζεται στην εμπιστοσύνη. Σε περιβάλλοντα ηλεκτρονικού εμπορίου ένα Σύστημα Υπόδειξης μπορεί να παροτρύνει έναν καταναλωτή να αγοράσει, να κατευθύνει τις αποφάσεις τους προς συγκεκριμένες κατευθύνσεις (πχ ευνοώντας ή αποφεύγοντας συγκεκριμένα αντικείμενα) και τελικά να αυξήσει τις πωλήσεις του.

2.3 Δομικά στοιχεία υπόδειξης

Στην πιο κοινή του σύνθεση, το πρόβλημα υπόδειξης ορίζεται ως το πρόβλημα εκτίμησης στοιχείων που δεν έχει δει ένας χρήστης. Η εκτίμηση αυτή συνήθως βασίζεται στις αξιολογήσεις που παρέχονται από τον χρήστη για άλλα στοιχεία. Αφού εκτιμήσουμε την πιθανή αξιολόγηση των αντικειμένων από το χρήστη ταξινομούμε τα στοιχεία με πιστοληπτική διαβάθμιση δημιουργώντας μια λίστα και προτείνουμε στο χρήστη το στοιχείο-α με την υψηλότερη εκτιμώμενη βαθμολογία-ες.

Σε κάθε περίπτωση και ανεξαρτήτως της τεχνικής που χρησιμοποιείται για την παραγωγή των υποδείξεων (θα μιλήσουμε αναλυτικά για αυτές παρακάτω) τα δεδομένα που χρησιμοποιούνται από τα Συστήματα Υπόδειξης περιλαμβάνουν τριών ειδών στοιχεία: τα

αντικείμενα, τους χρήστες και τις συναλλαγές, δηλαδή τις σχέσεις που αναπτύσσονται μεταξύ χρηστών και αντικειμένων.

2.3.1 Αντικείμενο

Τα αντικείμενα είναι τα στοιχεία που προτάσσονται από τα Συστήματα Υπόδειξης και μπορεί να χαρακτηρίζονται από την πολυπλοκότητα τους , την αξία τους ή το βαθμό χρησιμότητάς τους. Η αξία ενός αντικειμένου μπορεί να είναι θετική αν το αντικείμενο αποδείχτηκε χρήσιμο στο χρήστη ή αντίθετα να έχει αρνητική τιμή αν δεν ήταν χρήσιμο στο χρήστη ο οποίος υπέπεσε σε λανθασμένη επιλογή. Η απόκτηση ενός αντικειμένου πάντα συνοδεύεται από ένα κόστος το οποίο μπορεί να είναι ένα χρηματικό ποσό ή και ο χρόνος – κόπος για την αναζήτηση και εύρεση του συγκεκριμένου αντικειμένου.

Για παράδειγμα ο σχεδιαστής ενός Συστήματος Υπόδειξης ειδήσεων θα πρέπει να λάβει υπόψιν την πολυπλοκότητα ενός νέου αντικειμένου δηλαδή τη δομή του, την απεικόνιση του κειμένου, την αξία του που σχετίζεται με το χρόνο. Αλλά ταυτόχρονα ο σχεδιαστής θα πρέπει να αντιληφθεί ότι ακόμα κι αν ο χρήστης δεν πληρώνει για την ανάγνωση της είδησης καταναλώνει ένα κόστος. Το κόστος της αναζήτησης και ανάγνωσής του. Αν το συγκεκριμένο άρθρο είναι χρήσιμο στο χρήστη τότε το όφελος των πληροφοριών που εξέλαβε είναι υψηλότερο του κόστους και νιώθει ικανοποίηση. Αντίθετα αν το η είδηση που ανέγνωσε δεν σχετική με αυτό που έψαχνε τότε η αξία του αντικειμένου για αυτό το χρήστη είναι αρνητική.

Τα αντικείμενα που παρουσιάζουν χαμηλή πολυπλοκότητα και αξία είναι οι ειδήσεις, οι ιστοσελίδες, τα βιβλία, τα CDs και οι ταινίες, ενώ αντικείμενα με μεγαλύτερη πολυπλοκότητα και αξία είναι οι ψηφιακές κάμερες τα κινητά τηλέφωνα, οι ηλεκτρονικοί υπολογιστές κ.α. Τα αντικείμενα με την μεγαλύτερη πολυπλοκότητα και αξία είναι αυτά που σχετίζονται με πολιτικές ασφαλειών, ιατρικά και χρηματοοικονομικά ζητήματα, επενδύσεις, ταξίδια και θέσεις εργασίας. (*Montaner et al. 2003*).

Τα Συστήματα Υπόδειξης ανάλογα την τεχνολογία που χρησιμοποιούν μπορούν να χρησιμοποιηθούν διάφορες ιδιότητες και χαρακτηριστικά των αντικειμένων. Για παράδειγμα σε ένα Σύστημα Υπόδειξης ταινιών είδος της ταινίας, οι ηθοποιοί και ο σκηνοθέτης αποτελούν χαρακτηριστικά της κάθε ταινίας που μπορούν να χρησιμοποιηθούν για τη μοναδική περιγραφή της ταινίας και είναι δυνατή η απόκτηση γνώσης γύρω από το πόσο η χρησιμότητα ενός αντικειμένου εξαρτάται από τα χαρακτηριστικά του. Η

αναπαράσταση των αντικειμένων μπορεί να γίνει με τη χρήση διάφορων προσεγγίσεων για την αποτύπωση της πληροφορίας και την παρουσίασή της όπως ένας απλός κωδικός αλλά και μια λεπτομερέστερη αναπαράσταση χρησιμοποιώντας ένα σύνολο χαρακτηριστικών, καθώς επίσης και με τη μορφή μιας ιδέα σε μια οντολογική αναπαράσταση της περιοχής.

2.3.2 Χρήστης

Οι στόχοι και οι λόγοι χρήσης ενός Συστήματος Υπόδειξης μπορεί να διαφέρουν ανάμεσα στους χρήστες του σημαντικά. Για να επιτύχει την εξατομίκευση στις προτάσεις και στην αλληλεπίδρασή του με το χρήστη , τα Συστήματα Υπόδειξης χρησιμοποιούν ένα ευρύ σύνολο πληροφοριών για κάθε χρήστη.. Η πληροφορία αυτή μπορεί να εξαχθεί με πολλούς τρόπους, ενώ η μοντελοποίηση της συμπεριφοράς του χρήστη εξαρτάται από την τεχνική υπόδειξης που θα χρησιμοποιηθεί. Για παράδειγμα, στη συνεργατική μέθοδο (collaborative filtering), οι χρήστες μοντελοποιούνται ως μια απλή λίστα που περιέχουν τις βαθμολογίες του εν λόγω χρήστη για ορισμένα στοιχεία. Σε ένα σύστημα που αξιοποιεί δημογραφικά στοιχεία χρησιμοποιούνται τα κοινωνικά και δημογραφικά χαρακτηριστικά όπως η ηλικία, το φύλο, το επάγγελμα και η εκπαίδευση.

Το μοντέλο του χρήση δομείται από τα δεδομένα του κάθε χρήστη περιλαμβάνοντας τις ανάγκες και τις προτιμήσεις του. χρήστη που λέγεται για να αποτελέσει το πρότυπο των χρηστών . Έχουν προταθεί διάφορες προσεγγίσεις για τον τρόπο μοντελοποίησης των χρηστών και κατά κάποιο τρόπο τα Συστήματα Υποδείξεων μπορούν να θεωρηθούν ως εργαλεία που δημιουργούν υποδείξεις μέσω της κατασκευής και εκμετάλλευσης των μοντέλων αυτών . Δεδομένου ότι δεν είναι δυνατή η εξατομίκευση χωρίς τη χρήση ενός βολικού μοντέλου χρήστη (εκτός βέβαια αν οι προτάσεις δεν είναι εξατομικευμένες), το μοντέλο χρήστη θα παίζει πάντα κεντρικό ρόλο . Για παράδειγμα , θεωρώντας και πάλι τη συνεργατική μέθοδο φιλτραρίσματος , δημιουργείται το προφίλ του χρήστη είτε άμεσα από τις αξιολογήσεις του για διάφορα αντικείμενα είτε χρησιμοποιώντας αυτές τις αξιολογήσεις , το σύστημα παράγει ένα διάνυσμα συντελεστών τιμών , όπου οι χρήστες διαφέρουν στην τιμή που έχει κάθε συντελεστής για το μοντέλο τους.

Χρήσιμα δεδομένα που συχνά χρησιμοποιούνται είναι και το πρότυπο συμπεριφοράς του κάθε χρήστη στο σύστημα όπως το πρότυπο περιήγησης του στο χώρο , ή τα μοτίβα αναζήτησης ταξίδια (σε ένα σύστημα συστάσεων ταξιδιού) . Επιπλέον , δεδομένων των χρηστών το μοντέλο μπορεί να περιλαμβάνει και τις σχέσεις μεταξύ των χρηστών , όπως το

επίπεδο εμπιστοσύνης των σχέσεων αυτών μεταξύ των χρηστών . Ένα Σύστημα Υπόδειξης θα μπορούσε να χρησιμοποιήσει αυτές τις πληροφορίες για να παράξει υποδείξεις για τους χρήστες βάσει των προτιμήσεων άλλων χρηστών που μοιάζουν με τον τρέχοντα ή χαίρουν της εμπιστοσύνης τους,

2.3.3 Συναλλαγή

Ως συναλλαγή αναφέρεται η καταγεγραμμένη διαδικασία αλληλεπίδρασης μεταξύ του χρήστη και του Συστήματος Υπόδειξης. Οι συναλλαγές μοιάζουν με τα αρχεία καταγραφής δεδομένων στα οποία αποθηκεύονται σημαντικές πληροφορίες που παράγονται κατά τη διάρκεια αλληλεπίδρασης του χρήστη με το σύστημα και είναι χρήσιμες για τον αλγόριθμο ώστε να παρέχει υποδείξεις στους χρήστες του. Για παράδειγμα μια καταγραφή συναλλαγής θα μπορούσε να περιέχει αναφορά για ένα αντικείμενο που διάλεξε ο χρήστης και μια περιγραφή του πλαισίου (πχ πώς έγινε η αναζήτηση του αντικειμένου ή ποιος είναι ο στόχος του) που οδήγησε στη συγκεκριμένη υπόδειξη. Σε περίπτωση που είναι διαθέσιμη η αντίστοιχη πληροφορία στη συναλλαγή θα μπορούσε να συμπεριληφθεί και η βαθμολογία που έδωσε ο χρήστης για το επιλεγμένο αντικείμενο.

Οι βαθμολογίες αποτελούν την πιο συχνή μορφή πληροφορίας που καταγράφεται στις συναλλαγές που συλλέγονται από ένα Σύστημα Υπόδειξης. Οι βαθμολογίες συλλέγονται μπορεί να συλλεχθούν άμεσα, ρωτώντας το χρήστη να δώσει τη γνώμη για ένα αντικείμενο σε μια βαθμολογική κλίμακα. Σύμφωνα με τους *(Schafer et al., (2007)***,** οι αξιολογήσεις μπορούν να πάρουν μία διάφορες μορφές:

- *Αριθμητικές Βαθμολογίες*: όπως η κλίμακα με 1-5 αστέρια που παρέχεται από το σύστημα υπόδειξης βιβλίων της *Amazon*
- *Τακτικές Βαθμολογίες*: όπως “συμφωνώ απολύτως, συμφωνών, ούτε συμφωνών ούτε διαφωνώ, διαφωνώ, διαφωνώ απολύτως» κατά την οποία ο χρήστης τίθεται να επιλέξει τον όρο που ταιριάζει καλύτερα στην άποψή του σχετικά με το αντικείμενο που αξιολογεί. Συνήθως χρησιμοποιείται στα πλαίσια ερωτηματολογίων,
- *Δυαδικές Βαθμολογίες*: που μοντελοποιούν επιλογές κατά τις οποίες ο χρήστης καλείται να απαντήσει αν ένα αντικείμενο του αρέσει ή όχι (δύο επιλογές μόνο).
- *Μοναδιαίες Βαθμολογίες*: που υποδεικνύουν ότι ο χρήστης είδε, αγόρασε ή αξιολόγησε θετικά ένα προϊόν. Σε αυτές τις περιπτώσεις η ανυπαρξία βαθμολογίας

σημαίνει ότι δεν υπάρχει καμιά πληροφορία που να συσχετίζει το χρήστη με το συγκεκριμένο αντικείμενο στα όρια του συστήματος *(*αυτό μπορεί να σημαίνει ότι ο χρήστης έχει αγοράσει το αντικείμενο από κάπου αλλού. **(***Schafer et al. 2007***)**).

Μια άλλη μορφή αξιολόγησης των αντικειμένων από τους χρήστες χρησιμοποιεί ετικέτες που συνδέονται από το χρήστη με αντικείμενα που παρουσιάζονται από το σύστημα. Οι ετικέτες είναι λέξεις κλειδιά που περιγράφουν χαρακτηριστικά των αντικειμένων στα οποία αναφέρονται και μπορεί να αποτελούνται από μία ή περισσότερες λέξεις. Οι χρήστες μπορούν να επιλέξουν τις ετικέτες από ένα προκαθορισμένο σύνολο όρων.

Τέλος κατά την καταγραφή συναλλαγών που αφορούν έμμεσες αξιολογήσεις , το σύστημα προσπαθεί να μαζέψει πληροφορίες για τη γνώμη των χρηστών με βάση τις ενέργειες τους στο σύστημα. Για παράδειγμα, αν ένας χρήστης εισάγει τη λέξη «γιόγκα» στο σύστημα υπόδειξης βιβλίων της Amazon θα πρέπει να εμφανισθεί μία λίστα με βιβλία σχετικά με το θέμα. Ο χρήστης μπορεί να κάνει κλικ σε ένα συγκεκριμένο βιβλίο της λίστας προκειμένου να λάβει πρόσθετες πληροφορίες . Σε αυτό το σημείο , το σύστημα μπορεί να συμπεράνει ότι ο χρήστης παρουσιάζει ενδιαφέρον για το εν λόγω βιβλίο .

2.4 Το πρόβλημα της υπόδειξης

Το πρόβλημα της υπόδειξης έγκειται στη πρόβλεψη βαθμολογιών αντικειμένων που δεν είναι γνωστά στους χρήστες. Οι προβλέψεις στηρίζονται είτε σε παλαιότερες βαθμολογίες του χρήστη είτε σε άλλα δεδομένα που θα αναφερθούν στη συνέχεια. Αφού υπολογιστεί η πρόβλεψη για τα αντικείμενα επιλέγονται ως προτάσεις προς το χρήστη εκείνα που με την υψηλότερη πρόβλεψη.

Πιο συγκεκριμένα το πρόβλημα ορίζεται ως εξής: Έστω A το σύνολο με τους χρήστες του συστήματος και I το σύνολο με όλα τα αντικείμενα που είναι πιθανόν να προταθούν όπως μουσική, βιβλία, ταινίες κτλ. Το σύνολο I με όλα τα αντικείμενα μπορεί να είναι πολύ μεγάλος. Μπορεί να περιλαμβάνει από εκατοντάδες μέχρι και εκατομμύρια στοιχεία (πχ YouTube). Έστω u η συνάρτηση που υπολογίζει την χρησιμότητα του αντικειμένου I για το χρήστη a . Έτσι για κάθε χρήστη $a \epsilon A$ θέλουμε να επιλέξουμε αντικείμενο $i \square \epsilon I$ που να μεγιστοποιεί την χρησιμότητά του για το χρήστη $\forall\, a \epsilon A, \qquad s'_a = arg_{s \epsilon S} \max u(i, a)$

Συνήθως η χρησιμότητα ενός αντικειμένου εκφράζεται με μία βαθμολογία που δείχνει κατά πόσο ικανοποιεί το χρήστη το συγκεκριμένο αντικείμενο.

Στα συστήματα υπόδειξης το προφίλ του χρήστη ορίζεται από ένα μοναδικό κωδικό για την αναγνώρισή του και συνοδεύεται από σύνολο δεδομένων που περιλαμβάνουν διάφορα στοιχεία του χρήστη όπως ηλικία, φύλο, μόρφωση κτλ καθώς και το σύνολο των βαθμολογιών που έχει δώσει για αντικείμενα του I . Αντίστοιχα και τα αντικείμενα του I περιγράφονται. Δηλαδή από ένα μοναδικό κωδικό και επιπρόσθετα μπορεί να συμπεριλαμβάνουν και άλλα δεδομένων όπως αν το σύνολο I αφορά ταινίες το $i \epsilon I$ περιγράφεται από τον τίτλο, τη διάρκεια, το είδος ταινίας κτλ.

Για την κατασκευή ενός μοντέλου από το προφίλ ενός χρήστη συχνά γίνεται διαχωρισμός μεταξύ των άμεσων και έμμεσων τρόπων συλλογής δεδομένων.

Παραδείγματα άμεσων τρόπων συλλογής δεδομένων αναφέρονται παρακάτω:

- Ζητώντας από του χρήστη να βαθμολογήσει ένα αντικείμενο με κλιμακωτή βαθμολόγηση
- Ζητώντας από του χρήστη να ιεραρχήσει μία ομάδα αντικειμένων από το πιο αγαπημένο προς το λιγότερο αγαπημένο
- Ζητώντας από του χρήστη να επιλέξει ανάμεσα σε δύο αντικείμενα που του παρουσιάζονται
- Ζητώντας από το χρήστη να δημιουργήσει μία λίστα με αντικείμενα που τον ενδιαφέρουν ή του αρέσουν

Παραδείγματα έμμεσων τρόπων συλλογής δεδομένων αναφέρονται παρακάτω:

- Παρατηρώντας τα αντικείμενα που ο χρήστης επισκέπτεται στο σύστημα
- Αναλύοντας πόσο χρόνο περνά ο χρήστης βλέποντας ένα αντικείμενο
- Κρατώντας αρχείο με τα αντικείμενα που αγόρασε
- Λαμβάνοντας λίστα με αντικείμενα που ο χρήστης άκουσε ή είδε στον υπολογιστή του
- Αναλύοντας τα κοινωνικά δίκτυ που ο χρήστης ανήκει και ανιχνεύοντας παρόμοιες προτιμήσεις

Το βασικό πρόβλημα των συστημάτων υπόδειξης έγκειται στο ότι η χρησιμότητα δεν ορίζεται σε όλο το χώρο IxU αλλά σε ένα υποχώρο αυτού. Αυτό σημαίνει ότι η χρησιμότητα πρέπει να extrapolated σε όλο το χώρο. Αρχικά η χρησιμότητα μπορεί να οριστεί μόνο στα αντικείμενα που έχει βαθμολογήσει ο χρήστης στο παρελθόν δηλαδή χρησιμοποιούμε ένα υποσύνολο του IxU .

Συχνά η απεικόνιση του χώρου IxU (αντικείμενα επί χρήστες) γίνεται με τη χρήση ενός πίνακα όπου φαίνεται ο κάθε χρήστης ποια αντικείμενα του συστήματος έχει βαθμολογήσει και αν ναι με τι βαθμό.

Για παράδειγμα στον πίνακα 1 έχουμε τον πίνακα χρηστών-αντικειμένων για ένα σύστημα υπόδειξης εστιατορίων. Στον οριζόντιο άξονα εμφανίζονται οι χρήστες του συστήματος και

στον κάθετο τα εστιατόρια (αντικείμενα προς βαθμολόγηση). Το σύμβολο Ø υποδηλώνει

ότι ο χρήστης δεν έχει αξιολογήσει το συγκεκριμένο αντικείμενο. Ενώ η κλίμακα βαθμολογιών είναι 1-5. Στόχος του συστήματος είναι να προβλέψει όσο το δυνατόν πιο επιτυχημένα τη βαθμολογία που θα έβαζε ο χρήστης στα αντικείμενα που δεν έχουν ακόμα αξιολογηθεί από τους χρήστες.

	Το μπαλκόνι του Αιγαίου	Πλακούρι	Βίντσι	Ηλιοτρόπιο
Χάρης	1	Ø	Ø	Ø
Ειρήνη	Ø	2	0	5

Σοφία	2	Ø	3	0
Κώστας	2	4	0	Ø

Πίνακας 1 Τμήμα του πίνακα χρηστών-αντικειμένων

Η μετάβαση από τις γνωστές βαθμολογίες στις προβλέψεις γίνεται ορίζοντας πρώτα τη συνάρτηση χρησιμότητας και στη συνέχεια υπολογίζοντάς την για κάθε αντικείμενο. Τα αντικείμενα με τις υψηλότερες προβλέψεις προτείνονται στους χρήστες.

Ο τρόπος με τον οποίο θα παραχθεί η πρόβλεψη είναι πολύ σημαντική για την επιτυχία ενός συστήματος υποδείξεων καθώς στόχος αποτελεί η πρόβλεψη να είναι όσο το δυνατόν πιο κοντά στη βαθμολογία που θα έβαζε ο ίδιος ο χρήστης. Έχουν υπάρξει πολλές προσεγγίσεις που προσπαθούν με διαφορετικά μονοπάτια να πετύχουν καλύτερα αποτελέσματα.

2.5 Ταξινομήσεις

Σύμφωνα με την ταξινόμηση των (*Balabanovic & Shoham, 1997*) οι διάφορες προσεγγίσεις συνοψίζονται στις εξής τρεις κατηγορίες:

- *Υποδείξεις που βασίζονται στο περιεχόμενο (Content-based recommendations):* τα αντικείμενα που προτείνονται στο χρήστη είμαι παρόμοια με αντικείμενα που ο ίδιος βαθμολόγησε θετικά στο παρελθόν.
- *Συνεργατική Μέθοδος Υπόδειξης (Collaborative recommendations):* τα αντικείμενα που προτείνονται στο χρήστη αποτελούν αντικείμενα που έχουν αξιολογηθεί θετικά στο παρελθόν από χρήστες με παρόμοιες προτιμήσεις.
- *Υβριδικές Προσεγγίσεις (Hybrid approaches):* αυτές οι προσεγγίσεις συνδυάζουν τεχνικές της συνεργατικής μεθόδου και της μεθόδου που βασίζεται στο περιεχόμενο.

Ενώ ο (*Burke 2002*) συμφωνεί με τις κατηγορίες συστημάτων που βασίζονται στο περιεχόμενο και τη συνεργατική μέθοδο και προσθέτει άλλες 3. Τις υβριδικές προσεγγίσεις δεν τις θεωρεί βασική κατηγορία αφού αποτελούνται από συνδυασμό δύο ή και περισσότερων βασικών προσεγγίσεων. Συγκεκριμένα αναφέρεται στις εξής:

Συνεργατική Μέθοδος (Collaborative recommendation) είναι πιθανόν η πιο δημοφιλής και η πιο διαδεδομένη τεχνική για την παραγωγή υποδείξεων. Η συγκεκριμένη μέθοδος αθροίζει τις βαθμολογίες και τις προτάσεις των αντικειμένων, αναγνωρίζει τα κοινά αντικείμενα μεταξύ χρηστών βάσει των βαθμολογιών του και παράγει νέες προτάσεις βάσει των συγκρίσεων ανάμεσα στους χρήστες. Ένα τυπικό προφίλ αποτελείται από ένα διάνυσμα αντικειμένων με τις βαθμολογίες τους το οποίο αυξάνεσαι όσο ο χρήστης αλληλεπιδρά με το σύστημα. Σημαντικά συστήματα που χρησιμοποιούν τη συνεργατική μέθοδο υπόδειξης είναι τα εξής: GroupLens/NetPerceptions , Ringo/Firefly, Tapestry (*Goldberg et al. 1992*) Τα εν λόγω συστήματα μπορεί να βασίζονται είτε στη μνήμη (memorybased) όπου οι χρήστες συγκρίνονται ως προς την ομοιότητά τους βάσει συσχέτισης ή άλλων μέτρων είτε σε μοντέλο όπου η σύγκριση των χρηστών προκύπτει από το ιστορικό των βαθμολογιών το οποίο χρησιμοποιείται για τη δημιουργία προβλέψεων (*Breese et al. 1998*).

Δημογραφική μέθοδος (Demographic recommender systems) στόχο έχουν να κατηγοριοποιήσουν το κάθε χρήστη βάσει προσωπικών ιδιοτήτων και παράγουν προτάσεις βάσει δημογραφικών κατηγοριών. Παράδειγμα αποτελεί η προσέγγιση του *Krulwich (1997*) όπου δημιουργεί δημογραφικές ομάδες όπως αυτές προέκυψαν από έρευνα μάρκετινγκ και προτείνει ανάλογα ένα εύρος προϊόντων και υπηρεσιών. Απαιτείται έρευνα για την συλλογή δεδομένων για την κατηγοριοποίηση. Οι δημογραφικές τεχνικές δημιουργούν συσχετίσεις μεταξύ ανθρώπων αλλά με χρήση διαφορετικών δεδομένων (δημογραφικά) από αυτή της συνεργατικής μεθόδου (προτιμήσεις και βαθμολογίες αντικειμένων). Το όφελος των δημογραφικών μεθόδων είναι ότι δεν απαιτούν για τη λειτουργία τους το ιστορικό των αξιολογήσεων του χρήστη σε αντίθεση με τις τεχνικές συνεργατικής υπόδειξης και τις τεχνικές που στηρίζονται στο περιεχόμενο.

Μέθοδος υποδείξεων που βασίζεται στο περιεχόμενο (Content-based recommendation) αποτελεί μία αναζήτηση που στηρίζεται σε συνεχή διήθηση των δεδομένων. Τα αντικείμενα ορίζονται βάσει των χαρακτηριστικών που τα περιγράφουν. Παραδείγματος χάριν σε ένα

σύστημα υποδείξεως κειμένου οι λέξεις που περιέχονται σε ένα κείμενο αποτελούν χαρακτηριστικά του. Σε ένα σύστημα υποδείξεως μουσικής χαρακτηριστικά ενός τραγουδιού είναι το είδος του, ο τραγουδιστής, ο συνθέτης κτλ. Το προφίλ των προτιμήσεων του κάθε χρήστη χτίζεται βάσει των χαρακτηριστικών των αντικειμένων που ο ίδιος έχει βαθμολογήσει.

Οι μέθοδοι υπόδειξης που βασίζονται στη χρησιμότητα (Utility-based και knowledge-based recommenders) δεν προσπαθούν να δημιουργήσουν μακροπρόθεσμες γενικεύσεις για τους γενικεύσεις αλλά βασίζουν τις προτάσεις τους στο βαθμό ταιριάσματος της ανάγκης του χρήστη και ενός συνόλου απόψεων. Τα συστήματα υπόδειξης που στηρίζονται στη χρησιμότητα δημιουργούν προτάσεις βάσει της υπολογιζόμενης χρησιμότητας για κάθε χρήστη. Το μείζον ζήτημα σε αυτή την περίπτωση είναι η δημιουργία της κατάλληλης συνάρτησης που θα μετρά αποτελεσματικά την χρησιμότητα του αντικειμένου για κάθε χρήστη. Το όφελος της μεθόδου είναι ότι προσμετρά παράγοντες στη συνάρτηση χρησιμότητας που δεν αφορούν καθαρά το προϊόν όπως η εμπιστοσύνη προς τον παραγωγό του προϊόντος και η διαθεσιμότητα του προϊόντος δίνοντας στο χρήστη σφαιρικότερη ενημέρωση και καλύτερη υποστήριξη των αναγκών του.

Οι μέθοδοι υπόδειξης που βασίζονται στη γνώση (*Knowledge-based recommendation)* προσπαθούν να προτείνουν αντικείμενα βάσει αναγκών και προτιμήσεων του χρήστη. Η διαφοροποίηση των συγκεκριμένων συστημάτων αφορά στο γεγονός ότι στηρίζονται στη λειτουργική γνώση. Δηλαδή έχουν γνώση για το πώς ένα συγκεκριμένο προϊόν μπορεί να ικανοποιήσει μία συγκριμένη ανάγκη και κατ΄ επέκταση μπορούν να δικαιολογήσουν πώς ένα προϊόν μπορεί να καλύψει μία ανάγκη. Το προφίλ του χρήστη μπορεί να είναι οποιαδήποτε γνωσιακή δομή που υποστηρίζει αυτή τη λειτουργία. Παραδείγματος χάριν για τη Google μπορεί να είναι η επερώτηση του χρήστη στην αναζήτηση .

Όμως όπως αναφέρθηκε και πρωτύτερα δύο είναι οι βασικές τεχνικές. Της συνεργατικής μεθόδου και της μεθόδου που στηρίζεται στο περιεχόμενο των αντικειμένων (*Massa 2004*). Για αυτό και κρίνεται σημαντική η περαιτέρω αναφορά στον τρόπο λειτουργίας τους, τα πλεονεκτήματα και τα μειονεκτήματα τους.

2.6 Προβλήματα Συστημάτων Υπόδειξης

Η προσφορά των συστημάτων υπόδειξης είναι ιδιαίτερα σημαντική καθώς βοηθά τους χρήστες να βρίσκουν ενδιαφέροντα αντικείμενα μέσα από μία τεράστια δεξαμενή δεδομένων προσφέροντας προσωποποιημένες υπηρεσίες βάσει των ατομικών αναγκών και προτιμήσεων. Παρόλα αυτά παρουσιάζουν συγκεκριμένες βασικές αδυναμίες που σύμφωνα με τους *Massa et al., (2004)* είναι:

- **Σποραδικότητα δεδομένων(data sparseness):**

Σε πραγματικά περιβάλλοντα πίνακας με τα αντικείμενα που μπορούν να προταθούν και τους χρήστες μπορεί να είναι από χιλιάδες μέχρι και εκατομμύρια. Όπως οι χρήστες βαθμολογούν ένα μικρό υποσύνολο των αντικειμένων και υπάρχουν πολλά αντικείμενα που δεν είναι δημοφιλή και δεν έχουν λάβει βαθμολογίες. Το πρόβλημα αυτό αφορά κυρίως τα συστήματα συνεργατικής υπόδειξης.

Ως σποραδικότητα ορίζεται ο αριθμός των κενών κελιών στον πίνακα χρηστών-αντικειμένων. Δύο δημόσιες βάσεις που χρησιμοποιούνται πολύ συχνά για την αξιολόγηση μοντέλων και αλγορίθμων στα συστήματα υπόδειξης είναι το EachMovie και το MovieLens με σποραδικότητα που φτάνει το 97,6% και το 95,8% αντίστοιχα. Η σποραδικότητα αποτελεί μεγάλο πρόβλημα και για νέα συστήματα καθώς δεν υπάρχουν αποθηκευμένες βαθμολογίες ώστε να μπορούν να γίνουν προτάσεις στους χρήστες.

Το αποτέλεσμα της σποραδικότητας είναι ότι δύο τυχαίοι χρήστες του συστήματος έχουν μικρό αριθμό κοινών αντικειμένων που έχουν βαθμολογήσει και οι δύο. Έτσι το μέτρο ομοιότητα των χρηστών στηρίζεται σε δεδομένα θορύβου και δεν είναι αξιόπιστα. Στην πράξη πολλές φορές δεν υπάρχουν καθόλου κοινά αντικείμενα που να έχουν βαθμολογηθεί και από τους δύο ενεργούς χρήστες με αποτέλεσμα να είναι αδύνατος ο υπολογισμός της ομοιότητάς τους.
Για το πρόβλημα αυτό έχουν δημιουργηθεί τεχνικές που σκοπό έχουν να μειώσουν τις πολλές διαστάσεις της βάσης και κατά επέκταση και τη σποραδικότητα. Όμως για βάσεις με πολύ μεγάλο βαθμό κενών κελιών ακόμα και τέτοιες τεχνικές δεν είναι αποδοτικές.

- **Κρύα έναρξη (Cold start):**

Δεύτερη αδυναμία των συστημάτων υπόδειξης είναι το πρόβλημα της κρύας εκκίνησης Το πρόβλημα αυτό αφορά τους νέους χρήστες ενός συστήματος που εγγράφονται αλλά δεν έχουν δώσει ακόμα βαθμολογίες σε αντικείμενα ώστε να μπορούν να παραχθούν προτάσεις για αυτούς ή η αξιοπιστία των προτάσεων είναι χαμηλή. Θεωρείται ότι το πρόβλημα υπάρχει για κάποιο χρήστη αν έχει αξιολογήσει λιγότερα από πέντε αντικείμενα. Το πρόβλημα αυτό αφορά τόσο τα συστήματα που βασίζονται στο περιεχόμενο όσο και σε αυτά που χρησιμοποιούν τη συνεργατική μέθοδο.

- **Επιθέσεις (attacks):**

Ένα άλλο ζήτημα που προκύπτει είναι αυτό των κακόβουλων χρηστών που επιθυμούν να επηρεάσουν τη λειτουργία του συστήματος. Αν ο τρόπος παραγωγής των προτάσεων είναι γνωστός ένα αποτελεσματικός τρόπος επίθεσης είναι ο εξής: Δημιουργία προφίλ που θα έχουν βαθμολογήσει αντικείμενο κοντά στο μέσο όρο και το αντικείμενο που επιθυμούν να προβάλλουν υψηλότερα. Αυτός ο τρόπος επίθεσης εκμεταλλεύεται τεχνικές όπου λαμβάνεται υπόψιν για την παραγωγή προτάσεων για ένα χρήστη προφίλ άλλων χρηστών. Στόχος των σύγχρονων συστημάτων υπόδειξης είναι να αυξήσουν όσο το δυνατόν περισσότερο το χρόνο και το κόστος που χρειάζεται ένα κακόβουλος χρήστης να επιτεθεί αποθαρρύνοντας τον με αυτό τον τρόπο.

Τα συστήματα υπόδειξης είναι δύσκολο να γίνουν κατανοητά και να ελεγχθούν: Έχει αναφερθεί στη βιβλιογραφία ότι συχνά τα εν λόγω συστήματα θεωρούνται «μαύρα κουτιά» που ο χρήστης δεν μπορεί να κατανοήσει τον τρόπο λειτουργία τους. Επίσης είναι αδύνατο να ελέγξουν τις προτάσεις που τους γίνονται οπότε αν η ποιότητα των προτάσεων πέσει τότε ο χρήστης αναγκάζεται να εγκαταλείψει τη χρήστη του συστήματος.

2.7 Επεξήγηση βασικών Μεθόδων Υπόδειξης

2.7.1 Υποδείξεις Βασισμένες στο Περιεχόμενο

Τα συστήματα υπόδειξης που στηρίζονται στο περιεχόμενο μπορούν να χρησιμοποιηθούν σε ένα ευρύ φάσμα τομέων όπως μουσική, εστιατόρια, ταινίες, ιστοσελίδες κτλ. Αν και τα επιμέρους χαρακτηριστικά των διαφορετικών τομέων διαφέρουν, τα συγκεκριμένα συστήματα διαθέτουν κοινό τρόπο να περιγραφεί το προφίλ του χρήστη, τα αντικείμενα

που θα προταθούν, τα αντικείμενα που αρέσουν στο χρήστη και τα αντικείμενα μεταξύ τους για να βρεθεί ο βαθμός ομοιότητάς τους. Το προφίλ του χρήστη ανανεώνεται αυτόματα ώστε να είναι πάντα ενημερωμένο με τις τελευταίες προτιμήσεις του.

Τα συστήματα υπόδειξης που στηρίζονται στο περιεχόμενο έχουν τις ρίζες τους στο πεδίο της Ανάκτησης Πληροφοριών (Information Retrieval) (*Balabanovic et Shoham 1997*). Η λειτουργία τους είναι να αναλύουν την περιγραφή των αντικειμένων με σκοπό την ανίχνευση αντικειμένων που να είναι επιθυμητά στο χρήστη,

Θεμελιώδους σημασίας για τα συγκεκριμένα συστήματα είναι τα αντικείμενα και ο τρόπος αναπαράστσής τους. Για να μπορεί ένα αντικείμενο να αποτελεί πρόταση προς ένα χρήστη θα πρέπει να είμαι αποθηκευμένο στη βάση του συστήματος στον πίνακα των αντικειμένων. Οι στήλες του πίνακα αποτελούν τις ιδιότητες του αντικειμένου που επίσης μπορεί να ονομάζονται στη βιβλιογραφία μεταβλητές, πεδία ή χαρακτηριστικά. Κάθε αντικείμενο έχει τιμές για αυτές τις ιδιότητες και μέσω αυτών περιγράφεται. Επίσης κάθε αντικείμενο λαμβάνει και ένα μοναδικό κωδικό ώστε να μπορεί να ανακληθεί μέσω ατού. Παράδειγμα τμήματος ενός τέτοιου πίνακα φαίνεται στον πίνακα 2 που ακολουθεί:

	Κωδικός	Τύπος κουζίνας	Περιοχή	Τιμή κατά άτομο €
Το μπαλκόνι του Αιγαίου	15	Παραδοσιακή ελληνική	Άνδρος	
Πλακούρι	46	Παραδοσιακή ελληνική	Τήνος	12
Βίντσι	48	Ψησταριά	Άνδρος	17
Ηλιοτρόπιο	79	Χορτοφάγων	Αμοργός	20

Πίνακας 2 Τμήμα του πίνακα αντικειμένων

Ο αντίστοιχος πραγματικός πίνακας ενός συστήματος θα είχε εκατοντάδες εγγραφές αντικειμένων και περισσότερες ιδιότητες να τα περιγράφουν.

Το παραπάνω παράδειγμα αφορά την αναπαράσταση δομημένων δεδομένων καθώς υπάρχει και η περίπτωση να είναι αδόμητα. Τέτοιο παράδειγμα αποτελεί ένα άρθρο του οποίου το περιεχόμενο δεν μπορεί να σπάσει σε ιδιότητες και αυτού του είδους τα δεδομένα εμφανίζουν πολλαπλάσια πολυπλοκότητα επεξεργασίας καθώς κάποιες λέξεις έχουν πολλαπλές σημασίες.(*Pazzani & Billsus 2007*).

Το προφίλ του χρήστη είναι μια έννοια που χρησιμοποιείται από τα περισσότερα συστήματα υπόδειξης. Το προφίλ αποτελείται από διαφορετικούς τύπους πληροφορίας. Δύο πολύ βασικές διαστάσεις του προφίλ είναι α) το μοντέλο των προτιμήσεων του χρήστη και β) το ιστορικό της αλληλεπίδρασης του χρήστη με το σύστημα.

Το μοντέλο των προτιμήσεων περιγράφει το είδος των αντικειμένων που ενδιαφέρουν το χρήστη. Υπάρχουν πολλές εναλλακτικές αναπαραστάσεις αυτής της περιγραφής με πιο συχνή της συνάρτησης που για κάθε αντικείμενο προβλέπει την πιθανότητα του να ενδιαφέρεται για αυτό ο χρήστης. Για λόγους αποτελεσματικότητας αντί να εφαρμόζεται η συγκεκριμένη συνάρτηση για όλα τα αντικείμενα χρησιμοποιείται για την ανάκτηση των ν αντικειμένων που είναι πιο πιθανόν να ενδιαφέρουν το χρήστη,

Το ιστορικό των αλληλεπιδράσεων με το σύστημα μπορεί να περιλαμβάνει την αποθήκευση των αντικειμένων που ο χρήστης έχει δει μαζί με άλλες πληροφορίες (όπως αν αγόρασε ή όχι το προϊόν, πόση ώρα διάβασε το εν λόγω άρθρο κα). Άλλος τύπος ιστορικού μπορεί να αφορά τις επερωτήσεις που τέθηκαν από το χρήστη στο σύστημα (πχ ότι ο χρήστης έψαξε βιβλία του Ιουλίου Βερν με τιμή κάτω των 20 ευρώ)

Οι αλληλεπιδράσεις χρήστη-συστήματος μπορούν να έχουν πολλές χρήσεις. Αφενός το σύστημα μπορεί να κρατήσει τις σελίδες που επισκέφτηκε πρόσφατα ο χρήστης ώστε να τον διευκολύνει σε επόμενες επισκέψεις του. Αφετέρου μπορεί από το σύνολο των αντικειμένων που έχει σκοπό να προτείνει να αφαιρέσει αντικείμενα που ήδη έχει αγοράσει ή δει. Βέβαια αυτό δεν ισχύει πάντα καθώς υπάρχουν περιπτώσεις που επειδή ένα αντικείμενο το έχουμε αγοράσει δεν σημαίνει ότι δεν μπορεί να ικανοποιήσει ξανά και το ίδιο αποτελεσματικά τις ανάγκες μας. Πχ μια κρέμα προσώπου. Τέλος μία ακόμα σημαντική χρήστη του ιστορικού είναι ότι μπορεί να χρησιμοποιηθεί για την προπόνηση από αλγορίθμους μηχανικής μάθησης για την αξιολόγηση της επίδοσής τους.

Τα συστήματα υπόδειξης που στηρίζονται στο περιεχόμενο συχνά χρησιμοποιούν κανόνες που εφαρμόζονται στο ιστορικό των χρηστών για να προκύψουν από εκεί οι προτάσεις. Παράδειγμα η Amazon που περισσότερο γνωστή για τη συνεργατική μέθοδο που ακολουθεί υπάρχουν κομμάτια του συστήματος που στηρίζονται στο περιεχόμενο, όπως είναι τα «αγαπημένα». Αυτή η κατηγορία αντικειμένων προκύπτει από τις απευθείας απαντήσεις του χρήστη για το ποια θεωρεί αντικείμενα είτε έμμεσα βάσει των επιλογών του και του μοτίβου συμπεριφοράς του στο site (ουσιαστικά μέσω του ιστορικού του συγκεκριμένου χρήστη).

Στις μεθόδους που στηρίζονται στο περιεχόμενο η χρησιμότητα ενός αντικειμένου βασίζεται στις χρησιμότητες των αντικειμένων που έχει ορίσει ο χρήστης και είναι παρόμοιο με αυτό. Παραδείγματος χάριν σε ένα σύστημα που κάνει προτάσεις για βιβλία για να κάνει πρόταση σε ένα χρήστη θα επιλέξει τα αντικείμενα που μοιάζουν περισσότερο σε αντικείμενα που στο παρελθόν έχει αξιολογήσει θετικά ο χρήστης βασιζόμενο στα χαρακτηριστικά των ταινιών δηλαδή ηθοποιούς, σκηνοθέτη, είδος ταινίας κτλ.

Η συγκεκριμένη μέθοδος έχει τις ρίζες της στην ανάκτηση και διήθηση πληροφοριών αλλά διαφοροποιείται κυρίως λόγω της χρήστης του προφίλ του χρήστη που περιλαμβάνει πληροφορίες για το γούστο, τις προτιμήσεις και τις ανάγκες του χρήστη. Η λήψη αυτών των δεδομένων μπορεί να πραγματοποιηθεί άμεσα ή έμμεσα. Άμεσα μέσω ερωτηματολογίων κά και έμμεσα από τις συναλλαγές του χρήστη με το σύστημα με την πάροδο του χρόνου.
Όπως αναφέρεται και από το όνομα της μεθόδου στηρίζεται στο περιεχόμενο. Όπου ως περιεχόμενο ορίζεται το προφίλ του αντικειμένου δηλαδή μια ομάδα ιδιοτήτων και χρησιμοποιείται για να υπολογισθεί κατά πόσο το συγκεκριμένο αντικείμενο είναι κατάλληλο για το σκοπό της υπόδειξης. Όπως έχει αναφερθεί τα συστήματα υπόδειξης που στηρίζονται στο περιεχόμενο σχεδιάζονται κυρίως για αντικείμενα που περιγράφονται με κείμενο και συνήθως αυτά τα συστήματα περιγράφονται με λέξεις-κλειδιά.

Όπως ειπώθηκε ξανά τα συστήματα που βασίζονται στα περιεχόμενα υποδεικνύουν αντικείμενα παρόμοια με όσα είχε βαθμολογήσει ο χρήστης στο παρελθόν. Συγκεκριμένα διάφορα αντικείμενα είναι υποψήφια και συγκρίνονται με τα αντικείμενα που βαθμολογήθηκαν στο παρελθόν και τα αντικείμενα που ταιριάζουν καλύτερα επιλέγονται να προταθούν.

Θεωρούμε ως προφίλ που βασίζεται στο περιεχόμενο (*ContentBasedProfile*(*c*)) το προφίλ του χρήστη το οποίο περιέχει προτιμήσεις και επιλογές του. Το προφίλ χτίζεται αναλύοντας το περιεχόμενο των αντικειμένων που σε προηγούμενο χρόνο αξιολογήθηκαν από τον χρήστη μέσω τεχνικών ανάκτησης πληροφοριών. Δηλαδή μπορεί να αναπαρασταθεί ως ένα διάνυσμα με βάρη όπου κάθε βάρος υποδηλώνει την σημασία της κάθε λέξης-κλειδί όπως αυτή προκύπτει από τη χρήση διάφορες τεχνικές.

Πριν την δημιουργία του προφίλ των χρηστών είναι απαραίτητο για κάθε αντικείμενο του συστήματος να έχουν ορισθεί οι ιδιότητές του. Η διαδικασία αυτή γίνεται αυτόματα ή χειροκίνητα χωρίς πάντα να είναι εύκολο. Στη συνέχεια υπολογίζεται και αποθηκεύεται από το σύστημα η ομοιότητα μεταξύ των αντικειμένων χρησιμοποιώντας μέτρα συσχέτισης σαν αυτά της συνεργατικής μεθόδου. Όταν πρέπει να παραχθούν προτάσεις τότε λαμβάνονται ως βάσει τα θετικά βαθμολογημένα αντικείμενα που έχει αξιολογήσει ο χρήστης και βρίσκουμε με ποια άγνωστα για το χρήστη αντικείμενα της βάσεις παρουσιάζουν υψηλή ομοιότητα συγκρίνοντας τις τιμές που υπολογίστηκαν πριν την έναρξη της διαδικασίας,

Σύμφωνα με τους *Pazzani και Billsus (1997)* τα συστήματα υπόδειξης που στηρίζονται στο περιεχόμενο των αντικειμένων μπορούν να λειτουργήσουνε με τη χρήση αλγορίθμων μηχανικής μάθησης για την εκμάθηση των προφίλ των χρηστών όπως τα δένδρα απόφασης, near neighbor, Roccio's algorithm, γραμμικούς κατηγοριοποιητές, πιθανοτικές και μπαϊεσινές μεθόδους (Bayes). Η επιλογή εξαρτάται από τον τρόπο αναπαράστασης του περιεχομένου του αντικειμένου.

2.7.1.1 Πλεονεκτήματα υπόδειξης βάσει περιεχομένου

Η μέθοδος υποδείξεων βασισμένη στο περιεχόμενο παρουσιάζει σημαντικά πλεονεκτήματα σε σχέση με την συνεργατική (*Adomavicius & Tuzhilin, 2005*).

- *Ανεξαρτησία χρήστη*

Τα συστήματα υπόδειξης που στηρίζονται στο περιεχόμενο χρησιμοποιούν αποκλειστικά σε προηγούμενες αξιολογήσεις του κάθε χρήστη χωρίς να συμπεριλαμβάνουν καθόλου τις προτιμήσεις άλλων χρηστών. Έτσι δεν χρειάζεται να δημιουργηθεί γειτονιά χρηστών για την παραγωγή υποδείξεων, αρκούν οι βαθμολογίες του τρέχοντα χρήστη. Επέκταση της «ανεξαρτησίας» στην παραγωγή προβλέψεων για κάθε χρήστη αποτελεί η αντοχή στις επιθέσεις. Αφού η διαδικασία της υπόδειξης λαμβάνει ως εισερχόμενα δεδομένα μόνο τις

αξιολογήσεις του τρέχοντα χρήστη δεν είναι δυνατή η κακόβουλη παρέμβαση με σκοπό την τροποποίηση της λίστας των αντικειμένων προς υπόδειξη.

Κάνοντας την παραδοχή ότι ο κακόβουλος χρήστης δεν έχει απευθείας πρόσβαση στα δεδομένα ο πιο συχνός τρόπος επίθεσης είναι μέσω της δημιουργίας ψεύτικων χρηστών και ακολουθώντας την κατάλληλη στρατηγική βαθμολόγησης προσπαθούν να επηρεάσουν το σύστημα. Αυτού του είδους οι επιθέσεις δεν έχουν κανένα κέρδος στα συστήματα που στηρίζονται στο περιεχόμενο καθώς η επιλογή των προτάσεων προκύπτει αποκλειστικά από το προφίλ του ενεργού χρήστη.

- *Διαφάνεια*

Μια ενδιαφέρουσα πλευρά των Συστημάτων Υπόδειξης αφορά τη διαφάνεια, δηλαδή το κατά πόσο είναι ορατή είναι στους χρήστες η διαδικασία παραγωγής υποδείξεων. Η διάσταση αυτή δεν σχετίζεται με το πόσο καλά είναι τα αποτελέσματα που παράγονται αλλά έχει άμεση σχέση με το βαθμό εμπιστοσύνης των χρηστών προς αυτό καθώς τα άτομα έχουν την τάση να είναι καχύποπτα απέναντι σε αποτελέσματα που προέρχονται από διαδικασίες «μαύρα κουτιά» δηλαδή που δεν μπορούν να αντιληφθούν τη λογική λειτουργίας τους. Όσο σημαντικό είναι για τους χρήστες να λαμβάνουν υποδείξεις καλής ποιότητας άλλο τόσο σημαντικό είναι να αντιλαμβάνονται πώς αυτά λειτουργούν. Δηλαδή να έχουν λάβει επαρκείς πληροφορίες για το πώς προέκυψε η επιλογή προς σύσταση συγκεκριμένων αντικειμένων προς αυτούς. Γνωρίζοντας τον τρόπο λειτουργίας τους αποκτούν εμπιστοσύνη προς το σύστημα και είναι περισσότερο δεκτικοί και συγκαταβατικοί στις συστάσεις τους.

- *Νέο αντικείμενο*:

Τα συστήματα υπόδειξης που στηρίζονται στο περιεχόμενο έχουν τη δυνατότητα να εντάξουν στις υποδείξεις τους αμέσως ένα αντικείμενο που προστίθεται στη βάση δεδομένων τους χωρίς να έχει λάβει απολύτως καμία βαθμολογία αρκεί να έχει δομηθεί σωστά η περιγραφή του. Έτσι δεν χρειάζεται να περάσει αρκετό χρονικό διάστημα μέχρι να λάβει αρκετές βαθμολογίες και να μπορέσει να χρησιμοποιηθεί. Οποιοδήποτε αντικείμενο ανεξαρτήτως χρόνου ύπαρξης στο σύστημα ή αριθμού βαθμολογιών που έχει λάβει μπορεί να προταθεί άμεσα βάσει της περιγραφής του και μόνο.

2.7.1.2 Μειονεκτήματα υπόδειξης βάσει περιεχομένου

- *Περιορισμένη Ανάλυση περιεχομένου*

Τα αντικείμενα στα συστήματα υποδείξεων που βασίζονται στο περιεχόμενο περιγράφονται από ιδιότητες. Αν ένα αντικείμενο δεν έχει περιγραφεί επαρκώς ή ακόμα και αν δεν περιγραφεί σωστά –δηλαδή χρησιμοποιώντας λέξεις κλειδιά που να έχουν σημασία- τότε δημιουργεί πρόβλημα στο σύστημα παράγοντας υποδείξεις αμφιβόλου ποιότητας και μπορεί να οδηγήσει αντικείμενα που αν είχαν περιγραφθεί επαρκώς κι με σωστή επιλογή λέξεων να είναι ιδιαίτερα δημοφιλή να προτάσσονται σπανίως στους χρήστες.

- *Υπερεξειδείκευση*

Τα συστήματα υπόδειξης που είναι βασισμένα στο περιεχόμενο όπως αναφέρθηκε προτείνουν βάσει ομοιότητας αντικειμένων που βαθμολογήθηκαν στο παρελθόν και των αντικειμένων της βάσης και μάλιστα η ομοιότητα θα πρέπει να είναι υψηλή. Τα συστήματα αυτά δεν διαθέτουν άλλο τρόπο έρευνας για την ανίχνευση χρήσιμων προτάσεων.. Αυτό έχει σαν αποτέλεσμα η αναζήτηση να γίνεται κάθε φορά σε συγκεκριμένο υποσύνολο αντικειμένων της βάσης χωρίς να δίνεται δυνατότητα μεταπήδησης σε άλλο υποσύνολο αν ο χρήστης δεν βαθμολογήσει κάποιο αντικείμενο από αυτό. Το πρόβλημα αυτό ονομάζεται *serendipity problem και υποδηλώνει ότι μία πρόταση που απλά είναι ακριβής δεν σημαίνει αυτόματα ότι είναι και χρήσιμη για το χρήστη*. Δηλαδή δεν δίνεται η δυνατότητα στο χρήστη να δει αντικείμενα που πιθανόν να θεωρήσει ενδιαφέροντα και που αν δεν του προτάσσονταν διαφορετικά δεν θα τα ανακάλυπτε.

- *Νέος χρήστης*

Το σύστημα για να μπορέσει να χτίσει το προφίλ ενός χρήστη χρειάζεται ο χρήστης να έχει αλληλεπιδράσει αρκετά με το σύστημα και να έχει βαθμολογήσει ένα ικανό αριθμό αντικειμένων ώστε να μπορεί να κρίνει τις επιλογές και τις ανάγκες του. Αυτό όμως απαιτεί χρόνο στον οποίο το σύστημα θα αδυνατεί να παράγει ποιοτικές προτάσεις. Το πρόβλημα αυτό ονομάζεται και cold start.

2.7.2 Υποδείξεις Βασισμένες στη Συνεργατική Μέθοδο

Η συνεργατική μέθοδος είναι μία τεχνική των συστημάτων υπόδειξης που βασίζεται στην απλή σκέψη ότι όταν δύο άτομα μοιάζουν θα μοιάζουν και τα αντικείμενα που εκτιμούν (*Massa, 2004*).

Σε αντίθεση με τη μέθοδο που στηρίζεται στο περιεχόμενο και χρειάζεται την περιγραφή του αντικειμένου η συνεργατική μέθοδος στηρίζεται στην γνώμη των άλλων χρηστών όπως αυτή εκφράζεται μέσω βαθμολογίας. Μέσω των βαθμολογιών η μέθοδος μπορεί να αναγνωρίζει χρήστες που βαθμολογούν με παρόμοιο τρόπο και να προτείνει στο χρήστη αντικείμενα που έχουν εκτιμηθεί θετικά από τους παρόμοιους χρήστες. Δηλαδή δεν έχει σημασία τι είδους είναι τα αντικείμενα (μουσικό κομμάτι, ταινία, βιβλίο, άρθρο κτλ) καθώς η τεχνική λαμβάνει υπόψιν μόνο τις βαθμολογίες των χρηστών και μπορεί να εφαρμοστεί σε οποιοδήποτε τομέα χωρίς να χρειάζεται η περιγραφή των αντικειμένων (αυτόματη ή χειροκίνητη) δίνοντας της δυνατότητα δημιουργίας μεγάλης βάσεις με αντικείμενα χωρίς να δημιουργείται πρόβλημα επεκτασιμότητας. Τα πρώτα συστήματα που υλοποίησαν την ιδέα της Συνεργατικής Μεθόδου είναι το Tapestry *(Goldberg et al., 1992)* το GroupLens *(Resnick et al.,1994)* και το Ringo *(Shardan& & Maes, 1995)* όπου σκοπός τους ήταν να αυτοματοποιήσουν τις προβλέψεις αντικειμένων βάσει της ομοιότητας των χρηστών στο σύστημα που υπολογίζεται χρησιμοποιώντας τις παρελθούσες βαθμολογίες τους.

Τυπικό δεδομένο εισόδου για την τεχνική είναι ο πίνακας βαθμολογιών στον οποίο οι γραμμές αφορούν τους χρήστες, οι στήλες αφορούν τα αντικείμενα και το περιεχόμενο των κελιών είναι οι δοθείσες βαθμολογίες των χρηστών για κάθε αντικείμενο. Η συνεργατική μέθοδος χωρίζεται σε τρία βήματα:

- Το πρώτο βήμα αφορά τον υπολογισμό της ομοιότητας του ενεργού χρήστη με καθέναν από τους υπόλοιπους μέσω σύγκρισης της βαθμολογίας που έχουν δώσει οι χρήστες για τα κοινά αξιολογημένα αντικείμενα. Ο πιο συνήθης και διαδεδομένος τρόπος υπολογισμού της ομοιότητας είναι μέσω του συντελεστή συσχέτισης Pearson στον οποίο θα αναφερθούμε εκτενέστερα στη συνέχεια.
- Το δεύτερο βήμα είναι η πρόβλεψη της βαθμολογίας που αφορά την προσπάθεια πρόβλεψης της πραγματικής βαθμολογίας που θα έδινε ο ενεργός χρήστης στο συγκεκριμένο αντικείμενο. Το δεύτερο βήμα επαναλαμβάνεται ώστε να υπάρξει πρόβλεψη για όλα τα αντικείμενα που δεν έχει ακόμα βαθμολογήσει ο ενεργός χρήσης
- Το τρίτο βήμα ολοκληρώνει την μέθοδο προτείνοντας στον ενεργό χρήστη τα αντικείμενα του βήματος δύο με τις υψηλότερες προβλέψεις. Δηλαδή τα αντικείμενα που βάσει της συνεργατικής μεθόδου προβλέφθηκε ότι θα έπαιρναν τις υψηλότερες βαθμολογίες από τον ενεργό χρήστη,

Οι ερευνητές έχουν κατηγοριοποιήσει τους διάφορους αλγορίθμους που υλοποιούν τη ιδέα της Συνεργατικής Μεθόδου σε δύο βασικές κατηγορίες: αυτούς που στηρίζονται στη μνήμη Memory-based (user-based) και σε αυτούς που στηρίζονται σε μοντέλο Model-based (item-based).

2.7.2.1 Αλγόριθμοι συνεργατικής μεθόδου που βασίζονται στη μνήμη

Οι αλγόριθμοι που στηρίζονται στη μνήμη χρησιμοποιούν όλα τα δεδομένων χρηστών και αντικειμένων για να παράγουν πρόβλεψη *(breese at al., 1998, Sarwar et al., 2001).* Σκοπός τους είναι να υπολογίσουν το μέτρο της ομοιότητας μεταξύ του τρέχοντα και των ενεργών χρηστών. Για τον υπολογισμό λαμβάνονται υπόψιν οι αξιολογήσεις των χρηστών (έμμεσες ή άμεσες) Αυτά τα συστήματα χρησιμοποιούν τεχνικές ώστε να δομήσουν ομάδες χρηστών που ονομάζονται γειτονιές γύρω από κάθε χρήστη. Οι γειτονιές δομούνται βάσει της ομοιότητας. Δηλαδή για τη γειτονιά κάθε χρήστη επιλέγονται άλλη χρήστες που παρουσιάζουν παρόμοια συμπεριφορά, δηλαδή βαθμολογούν αντικείμενα με παρόμοιο τρόπο ή έχουν την τάση να αγοράζουν παρόμοια αντικείμενα. Όταν φορμαριστεί η γειτονιά του χρήστη τότε χρησιμοποιούνται διαφορετικές τεχνικές που συνδυάζουν τις επιλογές των χρηστών της γειτονιάς ώστε να παράγουν προβλέψεις για τον τρέχοντα χρήστη. Οι τεχνικές αυτές αναφέρονται ως τεχνικές κοντινού γείτονα (nearest-neighbor) ή τεχνικές που βασίζονται στο χρήστη (user-based) είναι πιο δημοφιλείς και ευρύτερα χρησιμοποιούμενες από τους αλγορίθμους που βασίζονται σε μοντέλο.

Κάθε χρήστης έχει στη διάθεσή του το μέσο όρο των βαθμολογιών για τα αντικείμενα που έχει αξιολογήσει. Στη συνέχεια οι βαθμολογίες που θα προβλεφθούν, υπολογίζονται εισάγοντας το σταθμισμένο άθροισμα των βαθμολογιών των υπολοίπων χρηστών. Τα βάρη μπορούν να καθοριστούν μέσω της ομοιότητας που έχει παρουσιάσει ο νέος χρήστης με τους άλλους χρήστες. Έτσι όσο πιο όμοιοι είναι μεταξύ τους, τόσο μεγαλύτερη συνεισφορά έχουν στο άθροισμα και επομένως τα αντίστοιχα βάρη θα είναι μεγαλύτερα. Στη συνέχεια παρουσιάζονται τρεις βασικές μέθοδοι υπολογισμού των βαρών.

- **Mean Square Differences:**

  Ο συντελεστής βαρύτητας (το βάρος) υπολογίζεται ως το αντίστροφο της μέσης τιμής του τετραγώνου της διαφοράς όπως εμφανίζεται στη συνέχεια:

$$w(a,i) = \frac{1}{(U_i - U_a)^2}$$

Εξίσωση 1

- **Vector Similarity:**

Ο συντελεστής βαρύτητας (βάρος) υπολογίζεται χρησιμοποιώντας το μέγεθος της γωνίας των διανυσμάτων του ενεργού και του τρέχοντας χρήση για κάθε συνδυασμό χρηστών.

$$w(a,i) = \sum_j \left[v_1(a,j) \quad v_1(i,j) \right] / \left(\sqrt{\sum_1 (k \in I)} v_1(a,k)^2 \right) \sqrt{\sum_1 (k \in I)} v_1(i,k$$

Εξίσωση 2

- **Pearson Correlation:**

Εξίσωση 3

Η ομοιότητα **sim(u,u')** μεταξύ των χρηστών u(τρέχοντα χρήστη) και u' (ενεργού χρήστη), καθορίζει την "απόσταση" μεταξύ τους και χρησιμοποιείται ως βάρος για τις βαθμολογίες των προβλέψεων . Έτσι όσο μεγαλύτερη είναι η τιμή της ομοιότητας του κάθε γείτονα δηλαδή όσο περισσότερο όμοιοι είναι οι χρήστες u και u' τόσο περισσότερο θα συνεισφέρει το βάρος στη βαθμολογία ru',i.

Οι *Breese et al. (1998)* και *Sarwar et al. (2001*), δημιούργησαν μια νέα προσέγγιση για τον υπολογισμό του μέτρου της ομοιότητας και την ονόμασαν προσέγγιση βάση συνημίτονου.

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

Εξίσωση 4

Όπου rx,s και ry,s είναι οι βαθμολογίες του αντικειμένου s που έχουν δοθεί από τους χρήστες x και y αντίστοιχα, το Sxy είναι το σύνολο απ' όλα τα αντικείμενα που βαθμολογήθηκαν από κοινού από τους χρήστες, ενώ X·Y είναι το εσωτερικό

2.7.2.2 Αλγόριθμοι συνεργατικής μεθόδου που βασίζονται σε μοντέλο

Οι αλγόριθμοι που στηρίζονται σε μοντέλο παρέχουν προβλέψεις δημιουργώντας πρωτύτερα μοντέλο για τις βαθμολογίες του χρήστη. Οι αλγόριθμοι αυτής της κατηγορίας χρησιμοποιούν πιθανοτική προσέγγιση και προσπαθούν να υπολογίσουν την εκτιμώμενη βαθμολογία για τα προτεινόμενα αντικείμενα βασιζόμενοι στις παλαιότερες αξιολογήσεις του χρήστη. Για τη δημιουργία του μοντέλου μπορεί να δημιουργηθεί με διάφορους αλγορίθμους μηχανικής μάθησης όπως τα Μπαϊεσιανά δίκτυα (Bayesian Network) , η συσταδοποίηση (clustering**)** και προσεγγίσεις που βασίζονται σε κανόνες (rule-based approaches).

Τα Μπαϊεσιανά δίκτυα δημιουργούν ένα πιθανοτικό μοντέλο για τον υπολογισμό της εκτιμώμενης βαθμολογίας των χρηστών ενώ οι αλγόριθμοι συσταδοποίησης αντιμετωπίζουν το πρόβλημα ως κατηγοριοποίηση. Δηλαδή ομαδοποιεί παρόμοιους χρήστες στην ίδια ομάδα και εκτιμούν την πιθανότητα ενός χρήστη να ανήκει σε μία συγκεκριμένη κατηγορία και μετά υπολογίζει την εκτιμώμενη βαθμολογία. Τέλος οι προσεγγίσεις με κανόνες ψάχνουν συσχετίσεις μεταξύ των αντικειμένων που έχουν αξιολογηθεί ή αγορασθεί από κάθε χρήστη ώστε να παράγουν προτάσεις ανάλογα με τη δύναμη της συσχέτισης.

Σύμφωνα με τους *Chirita et al (2005)* δύο είναι οι πιο δημοφιλείς τύπο συνεργατικών μεθόδων.

- Αλγόριθμοι που βασίζονται στο χρήστη

Στους αλγορίθμους συνεργασίας που στηρίζονται στην ομοιότητα μεταξύ χρηστών αρχικά πρέπει να δημιουργηθεί γειτονιά χρηστών για κάθε έναν. Στη συνέχεια για να υπολογισθεί η πρόβλεψη για ένα αντικείμενο (το αντικείμενο αυτό δεν έχει βαθμολογηθεί ακόμα από τον τρέχοντα χρήστη) υπολογίζεται ένας σταθμισμένος μέσος όρος των βαθμών που έχουν δώσει οι κ πιο κοντινοί χρήστες. Για τον υπολογισμό της ομοιότητας μεταξύ των χρηστών χρησιμοποιούμε τον συσχετισμό Pearson-r που υπολογίζεται ως εξής:

$$sim_{u,v} = \frac{\sum_{i \in I}(r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I}(r_{u,i} - \overline{r_u})^2} \sqrt{\sum_{i \in I}(r_{v,i} - \overline{r_v})^2}}$$

Εξίσωση 5

Όπου $r_{u,i}$ δηλώνει τη βαθμολογία του χρήστη u για το αντικείμενο i και $\overline{r_u}$ είναι ο μέσος όρος των βαθμολογιών που έχει δώσει ο χρήστης u για όλα τα αντικείμενα που έχει βαθμολογήσει. Αντίστοιχα $r_{v,i}$ είναι η βαθμολογία του χρήστη v για το αντικείμενο i και $\overline{r_v}$ είναι ο μέσος όρος των βαθμολογιών που έχει δώσει ο χρήστης v για όλα τα αντικείμενα που έχει βαθμολογήσει.

Για να υπολογίσουμε την πρόβλεψη για ένα αντικείμενο i για τον χρήστη u χρησιμοποιούμε την παρακάτω έκφραση:

$$p_{u,i} = \overline{r_u} + \frac{\sum_{v \in V} sim_{u,v} (r_{u,i} - \overline{r_u})}{\sum_{v \in V} |sim_{u,v}|}$$

Εξίσωση 6

Όπου V είναι το σύνολο των k όμοιων χρηστών. Στους υπολογισμούς των προβλέψεων για κάθε αντικείμενο συμπεριλαμβάνονται μόνο όσοι έχουν βαθμολογήσει το αντικείμενο αυτό.

Τα συστήματα που στηρίζονται στο χρήστη έχουν τα παρακάτω προβλήματα βάσει των (Sarwar et al. 2001):

- *Σποραδικότητα Δεδομένων*: Στην πραγματικότητα πολλά εμπορικά Συστήματα Υπόδειξης χρησιμοποιούνται για να αξιολογήσουν πολύ μεγάλα σύνολα προϊόντων (όπως το Amazon που πουλάει βιβλία, cd, μουσική κτλ). Σε τέτοιου είδους συστήματα ακόμα και οι πιο ενεργοί χρήστες δεν θα έχουν αξιολογήσει περισσότερο από το 1% των προϊόντων προς αξιολόγηση. Έτσι ο πίνακας αντικειμένου- χρήστη θα είναι πολύ αραιός κι επομένως η ποιότητα των προβλέψεων ίσως να μην είναι στα αναμενόμενα επίπεδα. είναι τόσο ικανοποιητική. Εκτός των χρηστών με λίγες βαθμολογίες το συγκεκριμένο πρόβλημα αφορά και όλους τους νέους χρήστες (cold start problem) καθώς δεν

υπάρχει καμία καταχωρημένη βαθμολογία. Κατά συνέπεια θα είναι δύσκολο να δομηθεί προσωποποιημένη γειτονιά όμοιων χρηστών για να προκύψουν προσωποποιημένες προτάσεις.

- *Επεκτασιμότητα*: Οι αλγόριθμοι που στηρίζονται στο χρήστη απαιτούν υψηλό αριθμό υπολογιστικής ισχύος όσο αυξάνει ο αριθμός των νέων χρηστών και αντικειμένων που καταχωρούνται στο σύστημα. Για ένα web-based σύστημα υπόδειξης με χιλιάδες χρήστες και αντικείμενα, είναι πολύ πιθανόν να προκύψουν σοβαρά προβλήματα επεκτασιμότητας.

- Αλγόριθμοι που βασίζονται στο αντικείμενο

Σε αυτή την κατηγορία αλγορίθμων υπολογίζονται ομοιότητες μεταξύ αντικειμένων. Από το σύνολο των αντικειμένων που έχουν βαθμολογηθεί από το χρήστη επιλέγονται τα $\mathbf{k}$ αντικείμενα που είναι περισσότερο όμοια με το αντικείμενο για το οποίο ζητείται πρόβλεψη. Από αυτά τα $\mathbf{k}$ αντικείμενα παράγεται ένας σταθμισμένος μέσος όρος που χρησιμοποιείται για να υπολογισθεί η συσχέτιση μεταξύ του αντικειμένου στόχου και των $\mathbf{k}$ αντικειμένων.

Για τον υπολογισμό της ομοιότητας αντικειμένων χρησιμοποιείται το προσαρμοσμένο συνημίτονο ομοιότητας (adjusted cosine similarity). Έστω $\mathbf{U}$ το σύνολο των χρηστών που βαθμολόγησαν τα αντικείμενα $\mathbf{i}$ και $\mathbf{j}$. Το προσαρμοσμένο συνημίτονο ομοιότητας (adjusted cosine similarity) υπολογίζεται ως εξής:

$$sim_{i,j} = \frac{\sum_{u \in U}(r_{u,i} - \overline{r_u})(r_{u,j} - \overline{r_u})}{\sqrt{\sum_{u \in U}(r_{u,i} - \overline{r_u})^2}\sqrt{\sum_{u \in U}(r_{u,j} - \overline{r_u})^2}}$$

Εξίσωση 7

Όπου $\mathbf{r_{u,i}}$ είναι η βαθμολογία του χρήστη $\mathbf{u}$ για το αντικείμενο $\mathbf{i}$ και $\overline{\mathbf{r_u}}$ είναι ο μέσος όρος των βαθμολογιών που έχει δώσει ο χρήστης $\mathbf{u}$. Αντίστοιχα $\mathbf{r_{u,j}}$ είναι η βαθμολογία του χρήστη $\mathbf{u}$ για το αντικείμενο $\mathbf{j}$.

Για τον υπολογισμό της πρόβλεψης για το αντικείμενο στόχος χρησιμοποιείται η παρακάτω έκφραση:

$$p_{u,i} = \frac{\sum_{j \in l} sim_{i,j} \, r_{u,j}}{\sum_{j \in l} |sim_{i,j}|}$$

Εξίσωση 8

Η παραπάνω έκφραση αντιπροσωπεύει την ομάδα των $\mathbf{l}$ πιο όμοιων αντικειμένων με το αντικείμενο στόχος $\mathbf{i}$ που έχουν ήδη βαθμολογηθεί από το χρήστη $\mathbf{u}$ και $\mathbf{r_{u,i}}$ είναι η βαθμολογία του χρήστη $\mathbf{u}$ για το αντικείμενο $\mathbf{i}$.

2.7.3 Διαφοροποίηση- Σύγκριση μεθόδων

Όπως και στην περίπτωση των τεχνικών που στηρίζονται στο περιεχόμενο η κύρια διαφοροποίηση μεταξύ τεχνικών που στηρίζονται σε μνήμη και μοντέλο για την διαδικασία παραγωγής υποδείξεων είναι ότι οι τεχνικές που χρησιμοποιούν μοντέλο υπολογίζουν τη χρησιμότητα (βαθμολογία) προβλέψεις όχι σε ad-hoc ευρετικούς κανόνες αλλά σε ένα μοντέλο το οποίο δημιουργήθηκε από τη χρήση στατιστικών μεθόδων και μεθόδων μηχανικής μάθησης.

Οι τεχνικές που στηρίζονται σε μοντέλο έχουν τη δυνατότητα καλύτερης αντιμετώπισης των προβλημάτων σποραδικότητας που απαντώνται κατά κανόνα σε όλα τα συστήματα (λόγω του μεγάλου αριθμού αντικειμένων και χρηστών που συμπεριλαμβάνουν). Επίσης προσφέρουν στο χρήστη τη δυνατότητα να αντιληφθεί πώς προέκυψαν οι προτάσεις που του υποδεικνύονται και τέλος αυξάνουν την ορθότητα των προβλέψεων, δηλαδή υποδεικνύουν αντικείμενα που αρέσουν στους χρήστες (βαθμολογούνται θετικά).

Όμως αντιμετωπίζουν και προβλήματα που σχετίζονται με την αντίστροφη πορεία της απόδοσης των προβλέψεων και της επεκτασιμότητας, αναγκάζοντας την ύπαρξης ισορροπίας μεταξύ των δύο εννοιών. Επίσης η υλοποίηση των μοντέλων υποδείξεων είναι αρκετά δαπανηρή και τέλος όταν γίνεται προσπάθεια για μείωση της σποραδικότητας των δεδομένων αυτή ακολουθείται με απώλεια πληροφοριών.

Αντίθετα οι τεχνικές που στηρίζονται στη μνήμη έχουν να αντιμετωπίσουν το πρόβλημα μη ύπαρξης στατιστικού μοντέλου που να τροφοδοτείται από τις αξιολογήσεις των χρηστών

και να είναι υπεύθυνο για την παραγωγή προτάσεων. Έτσι οι υποδείξεις είναι δύσκολο να τεκμηριωθεί μέσω ποιας λογικής παρήχθησαν και έχει σχετίζεται με προβλήματα αξιοπιστίας έναντι της τεχνικής. Η σποραδικότητα των βαθμολογιών κι εδώ δημιουργεί προβλήματα καθώς δεν είναι εύκολη η εύρεση όμοιων χρηστών για την δημιουργία γειτονιάς και κατ΄ επέκταση δεν μπορούν να παραχθούν προσωποποιημένες προτάσεις στον ενεργό χρήστη. Τέλος αυτή η τεχνική έχει να αντιμετωπίσει το πρόβλημα της κλιμάκωσης που αφορά τη διαθέσιμη μνήμη του υπολογιστικού συστήματος και την υπολογιστική ισχύ του. Οι υπολογισμοί που πρέπει να γίνονται σε πραγματικό χρόνο είναι μεγάλου φόρτου και είναι σχεδόν αδύνατη η ανανέωση της ομοιότητας ή της γειτονιάς κάθε χρήση σε πραγματικό χρόνο.

Λόγω του ότι και οι δύο τεχνικές (στήριξη σε μοντέλο και στήριξη σε περιεχόμενο) έχουν αδυναμίες συχνά γίνεται προσπάθεια συγκερασμού τους ώστε να υπάρξει συνεργία θετικών χαρακτηριστικών και αποδυνάμωση των μειονεκτημάτων.

Μία μέθοδος που συνδυάζει τεχνικές μνήμης και μοντέλου προτάθηκε από τους *Polat & Du, (2003)*, όπου εμπειρικά αποδείχθηκε ότι η χρήση αυτής της συνδυασμένης προσέγγισης μπορεί να παρέχει καλύτερης ποιότητας συστάσεις από ότι η μεμονωμένη χρήση των τεχνικών. Μια διαφορετική προσέγγιση για τη βελτίωση των επιδόσεων των υφιστάμενων αλγορίθμων συνεργατικής μεθόδου (*Takacs et al., 2009)* , όπου το σύνολο των εισερχόμενων δεδομένων (αξιολογήσεις αντικειμένων) που δίδονται από τον χρήστη είναι προσεκτικά επιλεγμένα με διάφορες τεχνικές που αποκλείουν το θόρυβο και να αξιοποιούν τη σποραδικότητα των δεδομένων των αξιολογήσεων .

3 Εμπιστοσύνη

3.1 Εισαγωγή

Οι επιχειρήσεις, το εμπόριο και γενικά ο κόσμος λειτουργεί πλέον χωρίς γεωγραφικά ή πολιτιστικά όρια , ενώ ταυτόχρονα διακινούνται υπέρογκα ποσά πληροφορίας , καθιστώντας την επικοινωνία και την αλληλεπίδραση ατόμων που δεν γνωρίζονται με τον παραδοσιακό -φυσικό τρόπο- ή και καθόλου σύνηθες και πολύ πιθανό σενάριο. Παραδείγματα τέτοια αλληλεπίδρασης μπορεί να είναι η επαφή μέσω κάποιου ιστότοπου γνωριμιών, μέσω ηλεκτρονικού ταχυδρομείου ή η συζήτηση για κάποιο θέμα σε φόρουμ. Η

επικοινωνία έχει αλλάξει μορφή και βαδίζει στο δρόμο της ψηφιοποίησης όπως και πολλές άλλες λειτουργίες που παραδοσιακά γινόταν μόνο μέσω φυσικής παρουσίας όπως πχ οι πληρωμές.

Όμως σε ένα κόσμο που η αβεβαιότητα και η καχυποψία απέναντι στους άλλους ανθρώπους , στην κοινωνία και γενικά στο σύστημα μεγαλώνει μπορεί να υπάρξει εμπιστοσύνη στο ίντερνετ; Όπως στις φυσικές κοινωνίες η αλληλεπίδραση των ατόμων εξαρτάται από την εμπιστοσύνη που έχουν αναπτύξει μεταξύ τους τα μέλη, το ίδιο ακριβώς ισχύει και στις διαδικτυακές κοινότητες που αποτελούν ουσιαστικά μεταφορά των φυσικών κοινοτήτων.

Στηρίζοντας την παραπάνω μεταφορά αναφέρεται ότι οι *(Sinha & Swearingen, 2001)* ανίχνευσαν ότι οι χρήστες που ανήκουν σε κοινωνικά δίκτυα θεωρούν πιο αξιόπιστες -και κατά συνέπεια προτιμούν σε μεγαλύτερο βαθμό- προτάσεις που προέρχονται από φίλους τους –δηλαδή άτομα για τα οποία αισθάνονται εμπιστοσύνη- και όχι από τον άγνωστο αλγόριθμο του συστήματος. Για αυτό το λόγο θα πρέπει να δίνεται η δυνατότητα στους χρήστες των κοινωνικών δικτύων να δηλώσουν εμπιστοσύνη προς τους άλλους . Για τους διαχειριστές ενός online συστήματος, υπάρχουν πολλές στρατηγικές για την αύξηση της εμπιστοσύνη των χρηστών του προς το σύστημα και προς τους άλλους χρήστες του. Στα web-based κοινωνικά δίκτυα όπου οι χρήστες κάνουν σαφείς δηλώσεις σχετικά με τα επίπεδα εμπιστοσύνης προς τους άλλους επιδιωκόμενος στόχος δεν είναι απαραίτητα η ανάπτυξη εμπιστοσύνης προς τη συγκεκριμένη ιστοσελίδα ή τα άλλα μέλη του αλλά η εξαγωγή χρήσιμων πληροφοριών για αυτούς μέσω των υπαρχόντων δεδομένων που έχουν αποθηκευθεί. Η εμπιστοσύνη μεταξύ δύο ατόμων και οι παράγοντες που συντελούν στην ανάπτυξή της δεν είναι πάντα εύκολο να ορισθούν και να απεικονισθούν σε ένα μοντέλο. Η απόφαση στηρίζεται στις προσωπικές ανάγκες, προτιμήσεις, παρελθούσες εμπειρίες από το συγκεκριμένο άτομο κά. Δηλαδή είναι μία απόφαση που στηρίζεται σε λογικά κριτήρια αλλά και ψυχολογικούς παράγοντες που επηρεάζονται από την καθημερινότητά μας και αλλάζουν με την πάροδο του χρόνου και τις εμπειρίες που βιώνουμε.

Στην πραγματικότητα, το Διαδίκτυο παρουσιάζει ένα τεράστιο πλεονέκτημα για τη διάδοση πληροφοριών πέραν των παραδοσιακών τρόπων επικοινωνίας, όλες οι δηλώσεις εμπιστοσύνης μπορούν να γίνουν δημόσια και να είναι μόνιμα ορατές και να είναι ανακτήσιμες ανά πάσα στιγμή από όλους ή μόνο από κάποιους συγκεκριμένους χρήστες.

Αξίζει να σημειωθεί ότι το θέμα της εμπιστοσύνης, οι διαστάσεις της και οι πιθανές χρήσεις της ήταν ένα ενδιαφέρον θέμα για πολλούς επιστημονικούς τομείς για αιώνες. Αυτό δεν πρέπει να εκπλήσσει δεδομένου ότι η εμπιστοσύνη είναι μια πολύ ανθρώπινη και κοινωνική αντίληψη που απασχόλησε τους φιλοσόφους πολύ πριν από την έλευση των πρώτων ηλεκτρονικών υπολογιστών. Η εμπιστοσύνη έχει μελετηθεί από ερευνητές και στοχαστές σε διάφορα πεδία, εκτός από επιστήμη των υπολογιστών, όπως την οικονομία, την πολιτική, την εξελικτική βιολογία, την ανθρωπολογία, την κοινωνιολογία και τη φιλοσοφία.

Σημαντικά συστήματα που παρέχουν στους χρήστες τη δυνατότητα δημιουργίας σχέσεων εμπιστοσύνης αποτελούν τα κοινωνικά συστήματα υπόδειξης (Social Recommender Systems) που αποτελούν υποσύνολο των Web 2.0 εφαρμογών. Τέτοιου είδους συστήματα χρησιμοποιούν το περιεχόμενο που έχει ανεβάσει κάθε χρήστης και τις σχέσεις που έχει δημιουργήσει σε αυτό για να παράξει προσωποποιημένες προτάσεις προς αυτόν. Παραδείγματα κοινωνικών συστημάτων υπόδειξης που έχουν ενσωματωθεί σε δημοφιλέστατες εφαρμογές κοινωνικής δικτύωσης και όχι μόνο είναι το Facebook και το LinkedIn.

3.2 Ορισμός εμπιστοσύνης

Όπως αναφέρθηκε ήδη η έννοια της εμπιστοσύνης πριν χρησιμοποιηθεί από την επιστήμη υπολογιστών και τα συστήματα υπόδειξης αποτέλεσε ερευνητικό θέμα για άλλα πεδία όπως η ψυχολογία, η κοινωνιολογία , η φιλοσοφία κά . Κοινό συμπέρασμα αποτελεί ότι η εμπιστοσύνη είναι υποκειμενική και διαφέρει από άτομο σε άτομο αφού καθένας είναι και διαφορετικός.

Σύμφωνα με τον *Sztompka (1999)* «Η εμπιστοσύνη είναι ένα στοίχημα για τις μελλοντικές ενδεχόμενες ενέργειες των άλλων.» Υπάρχουν δύο βασικές διαστάσεις κατά τον ορισμό: η πίστη και η δέσμευση. Κατ 'αρχάς, ένα άτομο πιστεύει ότι το έμπιστο πρόσωπο θα ενεργήσει με έναν ορισμένο τρόπο. Η πεποίθηση μόνο, ωστόσο, δεν είναι αρκετή για να οδηγήσει σε εμπιστοσύνη. Εμπιστοσύνη αναπτύσσεται όταν η πεποίθηση χρησιμοποιείται ως βάση για την ανοικοδόμηση δέσμευσης για μια συγκεκριμένη δράση.

Στο πεδίο της επιστήμης υπολογιστών έχουν δοθεί οι παρακάτω ορισμοί για την εμπιστοσύνη: Σύμφωνα με τους *Massa & Avesani (2007)* δήλωση εμπιστοσύνης ορίζεται η

ρητή έκφραση άποψης από ένα χρήστη για έναν άλλο λαμβάνοντας υπόψιν συγκεκριμένα χαρακτηριστικά αυτού του χρήστη. Για παράδειγμα, σε ένα ιστοχώρο όπου οι συμμετέχοντες παραθέτουν σχόλια σχετικά με τα προϊόντα, οι χρήστες θα μπορούσαν να κληθούν να εκφράσουν μια θετική δήλωση εμπιστοσύνης για το χρήστη "του οποίου κριτικές και οι βαθμολογίες έχουν αξιολογηθεί σταθερά ως πολύτιμες " και μια αρνητική δήλωση εμπιστοσύνης σχετικά με χρήστες « των οποίων τα σχόλια θεωρούνται συνεχώς προσβλητικά, ανακριβή, ή σε γενικές γραμμές χωρίς αξία» –οι συγκεκριμένες φράσεις αυτές χρησιμοποιούνται από το epinions.com όταν ζητείται από τους χρήστες να δηλώσουν την εμπιστοσύνης τους . Η Golbeck στη διατριβή της βασίζει τους ορισμούς της κυρίως στην κοινωνιολογία και ορίζει ότι ο χρήστης Α εμπιστεύεται το χρήστη Β αν δεσμεύεται σε μια ενέργεια με βάση την πεποίθηση ότι οι μελλοντικές δράσεις του Β θα οδηγήσουν σε ένα καλό αποτέλεσμα.

Μια σημαντική σημείωση για την εμπιστοσύνη είναι ότι δεν είναι στην πραγματικότητα μια συμπαγής και ενιαία τιμή. Παραδείγματος χάριν ένα συγκεκριμένο πεδίο της διασκέδασης η μουσική. Μπορεί κάποιος χρήστης να δημιουργήσει μια γενική άποψη για το ποιους χρήστες εμπιστεύεται περισσότερο για το γούστο και τις επιλογές του αλλά θα ήταν πιο ακριβής και σίγουρος για την εμπιστοσύνη του αν μπορούσε να δηλώσει τιμές για κάθε είδος μουσικής. Δηλαδή μπορεί να εμπιστεύεται το χρήστη Α για ροκ κομμάτια αλλά καθόλου για ποπ. Αυτός ο κατακερματισμός των μουσικών κομματιών θα μπορούσε να συνεχιστεί σε μεγαλύτερο βάθος αλλά τότε θα αυξηθεί σημαντικά η πολυπλοκότητα χωρίς να υπάρχει αντίστοιχο κέρδος.

Η χρήστη της έννοιας της εμπιστοσύνης έχει αποδειχθεί ότι αυξάνει την αποτελεσματικότητα των συστημάτων υπόδειξης και κυρίως αυτών που χρησιμοποιούν συνεργατικές μεθόδους για την παραγωγή προτάσεων στους χρήστες. Σύμφωνα με τους *O'Donovan & Smyth (2005*) ένα χρήστης για να συμπεριληφθεί από το σύστημα για τη διαδικασία παραγωγής υποδείξεων δεν αρκεί να έχει υψηλή ομοιότητα με τον τρέχοντα χρήστη αλλά θα πρέπει να είναι και αξιόπιστος, δηλαδή να έχει ιστορικό αξιόλογων υποδείξεων προς αυτόν. Τα κοινωνικά αυτά συστήματα υπόδειξης που επικροτούν τη αλληλεπίδραση των χρηστών και χρησιμοποιούν τις δηλώσεις εμπιστοσύνης ονομάζονται ως *trust-based enhanced recommender systems*. Σε αυτές τις περιπτώσεις δομείται δίκτυο όπου έχει ως κόμβους τους χρήστες του οι οποίοι συνδέονται μεταξύ τους με ακμές, η τιμή των οποίων δηλώνει το βαθμό εμπιστοσύνης του εξερχόμενου κόμβου προς τον

εισερχόμενο. Παραδείγματα τέτοιων δίκτυα, τα δεδομένα των οποίων χρησιμοποιούνται ευρέως και για αξιολόγηση αλγορίθμων συστημάτων υπόδειξης είναι:

- *FilmTrust* (*Golbeck, 2006)*: ένα online κοινωνικό δίκτυο που οι χρήστες έχουν τη δυνατότητα να αξιολογήσουν ταινίες μέσω μίας δεκάβαθμης κλίμακας και επίσης μπορούν να δηλώσουν εμπιστοσύνη σε χρήστες. Η δήλωση σημαίνει εμπιστοσύνη ίση με 1 ενώ αν δεν υπάρχει δήλωση τότε η εμπιστοσύνη αξιολογείται ως 0.
- *Epinions*: δικτυακός τόπος όπου οι χρήστες μπορούν να αξιολογήσουν και να παραθέσουν την άποψή τους για αντικείμενα (ταινίες, αυτοκίνητα, μουσική κά) μέσω κλίμακας από 1-5. Οι χρήστες μπορούν επίσης να δηλώσουν την εμπιστοσύνης τους στους άλλους χρήστες του συστήματος βαθμολογώντας τους με «1» ενώ όσους δεν εμπιστεύονται με «-1» τοποθετώντας τους στη λεγόμενη «μαύρη λίστα» (Blacklist)

3.3 Διαστάσεις εμπιστοσύνης

Σύμφωνα με την διδακτορική διατριβή της Golbeck η εμπιστοσύνη δομείται από τρεις βασικές διαστάσεις:

- **Μεταβατικότητα (Transitivity)**

Βασική ιδιότητα της εμπιστοσύνης είναι η μεταβατικότητα (Transitivity). Η εμπιστοσύνη δεν είναι απολύτως μεταβατική με τη μαθηματική έννοια του όρου, Δηλαδή, αν η Alice εμπιστεύεται ιδιαίτερα τον Bob και ο Bob εμπιστεύεται πολύ τον Chuck, δεν είναι πάντα επαγωγικό ότι η Alice θα εμπιστεύονται ιδιαίτερα Chuck. Υπάρχει, ωστόσο, μια αντίληψη ότι η εμπιστοσύνη μπορεί να διαδοθεί μεταξύ των ανθρώπων. Σε γενικές γραμμές, όταν συναντάται έναν άγνωστο πρόσωπο, είναι κοινό για τους ανθρώπους να ρωτήσουν έμπιστους φίλους τις απόψεις τους σχετικά με το πόσο εμπιστεύονται αυτό το νέο πρόσωπο.

Υπάρχουν δύο τύποι της εμπιστοσύνης που μπορεί κανείς να εκφράσει: την εμπιστοσύνη σε ένα άτομο και την εμπιστοσύνη στις συστάσεις του ατόμου για άλλα άτομα. Η Alice μπορεί να εμπιστεύεται τον Bob να της συστήσει έναν υδραυλικό, αλλά δεν τον εμπιστεύεται καθόλου να της συστήσει άλλα άτομα για να λάβει τη γνώμη τους για υδραυλικούς. Παρά τη διχοτόμηση, σε κοινωνικά δίκτυα είναι προτιμότερο να χρησιμοποιείται η ενιαία αξία της εμπιστοσύνης για μείωση της πολυπλοκότητας της έκφρασης. Ένα ενιαίο σύστημα

βαθμολόγησης είναι καλή πρακτική καθώς οι χρήστες σπάνια αν όχι ποτέ, καλούνται να εκφράσουν τόσο λεπτομερείς απόψεις για άλλους.

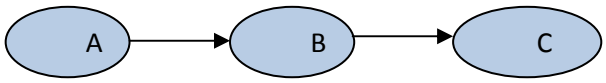
Ο ορισμός της εμπιστοσύνης υποστηρίζει την ιδέα της μεταβατικότητας. Υπενθυμίζεται ότι η εμπιστοσύνη περιλαμβάνει την πεποίθηση ότι το έμπιστο πρόσωπο θα διαπράξει μια ενέργεια που θα παράγει ένα καλό αποτέλεσμα. Συνεχίζοντας το παράδειγμα με τον υδραυλικό όταν η Alice ρωτήσει τον Bob αν ο Chuck είναι ή όχι ένας καλός υδραυλικός, θα χρησιμοποιήσει την απάντηση του Bob για να στηρίξει τη δράση της καθώς πιστεύει ότι ο Bob θα της δώσει πληροφορίες που θα οδηγήσουν σε ένα καλό αποτέλεσμα. Έτσι, αν ο Bob πει η Alice θα πρέπει να εμπιστεύεται τον Chuck, τότε η Alice στηριζόμενη στην εμπιστοσύνη της για τον Bob θα αναπτύξει κάποια εμπιστοσύνη προς τον Chuck. Θα έχουμε κάποια εμπιστοσύνη στο πρόσωπό του, διότι, με βάση τις πληροφορίες του Bob, είναι αξιόπιστος. Το ίδιο επιχείρημα μπορεί να επεκταθεί σε μακρύτερες αλυσίδες εμπιστοσύνης (μονοπάτια).

Επειδή η εμπιστοσύνη δεν είναι απολύτως μεταβατική, θα αναμένουμε ότι υποβαθμίζεται η αξιοπιστία της κατά μήκος μιας αλυσίδας των γνωριμιών. Η Alice είναι πιθανόν να έχει περισσότερη εμπιστοσύνη στον Chuck αν ξέρει τον Bob άμεσα από τα να δεχθεί την αξιοπιστία του μέσω μίας αλυσίδας ανθρώπων που περνούν πληροφορίες μέσω των σχέσεών τους. Υπολογιστικά, αυτή η ιδέα της διάδοσης της εμπιστοσύνη στις αλυσίδες συνδέσεων των κοινωνικών δικτύων (αξιοποιώντας έτσι κάποια μορφή της μεταβατικότητας) έχει ευρέως μελετηθεί και υλοποιηθεί *(Gray, et al., 2003, Guha, Kumar, 2004, Jøsang, 1996, Jøsang et al., 2003, Richardson et al., 2003, Ziegler & Lausen, 2004).*

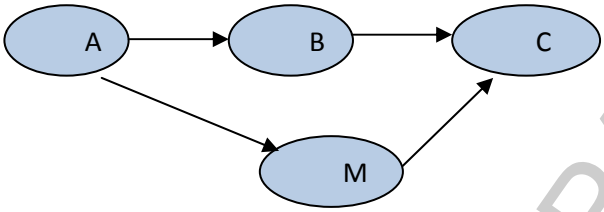
- **Δυνατότητα σύνθεσης (Composability)**

Η μεταβατικότητα όπως ήδη αναφέρθηκε περιγράφει πώς μια τιμή εμπιστοσύνης μπορεί να κυκλοφορήσει μέσω μίας αλυσίδας ανθρώπων. Αυτό απεικονίζεται στο μέρος (α) της Εικόνας 1. Οι συστάσεις σχετικά με την αξιοπιστία ενός άγνωστου προσώπου χρησιμοποιούνται για να υποστηρίξουν την πεποίθηση σχετικά με τις δράσεις του ατόμου αυτού και κατά συνέπεια να οδηγήσει σε κάποιο επίπεδο εμπιστοσύνης. Όμως πολλές φορές ζητείται η άποψη πολλών ατόμων και για αυτό η Alice πρέπει να συνθέσει τις πληροφορίες και να αποφασίσει ανάλογα αν πρέπει ή όχι να εμπιστευτεί τον Chuck. Κατά συνέπεια η δυνατότητα σύνθεσης της εμπιστοσύνης είναι ένα άλλο σημαντικό

χαρακτηριστικό για τη δόμηση των υπολογισμών που θα οδηγήσουν σε συγκεκριμένη τιμή εμπιστοσύνης από την Alice στον Chuck.



Εικόνα 1α Η Alice αποκτά άποψη για τον Chuck μέσω της εμπιστοσύνης της προς τον Bob



Εικόνα 1b Η Alice αποκτά άποψη για τον Chuck μέσω της εμπιστοσύνης της προς τον Bob και τη Mary οι οποίοι έχουν σχέση εμπιστοσύνης με τον Chuck

Η έννοια της συνθεσιμότητας αποκτά νόημα θεωρώντας ότι οι συστάσεις των φίλων χρησιμοποιούνται για την ανοικοδόμηση πίστης προς τον άγνωστο που θα οδηγήσει σε δήλωση εμπιστοσύνης. Ο τρόπος που θα γίνει η σύνθεση της εμπιστοσύνης και η ροή της μπορεί να ποικίλει από κατάσταση σε κατάσταση και από πρόσωπο σε πρόσωπο. *Richardson et al. (2003).*

- **Εξατομίκευση και ασυμμετρία (Personalization & Asymmetry)**

Μια διάσταση της εμπιστοσύνης που είναι σημαντική σε κοινωνικά δίκτυα και που κατά το παρελθόν παραβλεπόταν είναι η εξατομικευμένη εμπιστοσύνη. Η εμπιστοσύνη είναι εγγενώς μια προσωπική άποψη. Δύο άνθρωποι συχνά έχουν πολύ διαφορετικές απόψεις σχετικά με την αξιοπιστία του ίδιου προσώπου. Για παράδειγμα, ας εξετασθεί το πεδίο της πολιτική. Στις Ηνωμένες Πολιτείες, όταν ρωτήσει κάποιος «εμπιστεύεστε τον τρέχοντα Πρόεδρο;" ο πληθυσμός θα χωριστεί - κάποιοι θα τον εμπιστεύονται πολύ μεγάλο βαθμό, και οι άλλοι θα έχουν πολύ λίγη εμπιστοσύνη στις ικανότητές του.

Ο ορισμός της εμπιστοσύνης περιλαμβάνει την πεποίθηση ότι οι ενέργειες ενός έμπιστου ατόμου θα οδηγήσει σε θετικό αποτέλεσμα. Τι μπορεί να θεωρηθεί ως θετικό αποτέλεσμα διαφέρει από το ένα άτομο στο άλλο. Επειδή όλοι έχουν συμφέροντα, οι προτεραιότητες και οι απόψεις τους μπορεί να έρχονται σε σύγκρουση με τα συμφέροντα, τις προτεραιότητες και απόψεις των άλλων. Κατά συνέπεια πότε και πόσο εμπιστεύονται οι άνθρωποι θα ποικίλλει ανάλογα τα δεδομένα της κατάστασης. Λόγω της υποκειμενικότητας

της αλήθειας και ότι σπάνια υπάρχει μία μόνο οπτική για ένα γεγονός, ένα καθολική μέτρο για την αξιοπιστία ενός ατόμου είναι επίσης σπάνιο. Οι υπολογισμοί για την εμπιστοσύνη πρέπει να γίνουν από την πλευρά του ατόμου και να αντικατοπτρίζει τα συμφέροντα και τις επιθυμίες του

Η ασυμμετρία της εμπιστοσύνης είναι επίσης σημαντική, διάσταση καθώς αντανακλά ένα συγκεκριμένο τύπο εξατομίκευσης. Για δύο άτομα που εμπλέκονται σε μια σχέση, η εμπιστοσύνη δεν είναι απαραιτήτως αμφίδρομη και ισότιμη. Επειδή τα άτομα έχουν διαφορετικές εμπειρίες, ψυχολογικά υπόβαθρα και απόψεις είναι κατανοητό γιατί δύο άνθρωποι μπορούν να εμπιστεύονται ο ένας τον άλλον με διαφορετικό επίπεδο. Για παράδειγμα, οι γονείς και τα παιδιά εμπιστεύονται σαφώς οι μεν τους δε σε διαφορετικό επίπεδο, καθώς τα παιδιά δεν είναι ικανά να αναλάβουν πολλά καθήκοντα. Αυτή η έντονη ασυμμετρία μπορεί να συμβεί και σε άλλες κοινωνικές σχέσεις των ανθρώπων. Αυτό μπορεί να πραγματοποιηθεί πλήρως φθάνοντας στο άκρο της «μονόδρομης εμπιστοσύνης» όπου οι περιστάσεις αναγκάζουν ένα άτομο να εμπιστευτεί τον άλλο χωρίς να υπάρχει αμοιβαία εμπιστοσύνη *(Hardin, 2002, Cook, 2001).*

Ωστόσο, οι περισσότεροι ασυμμετρίες στην εμπιστοσύνη δεν είναι τόσο ακραίες όσο αυτή που μόλις αναφέρθηκε. Στις περισσότερες περιπτώσεις υπάρχει αμοιβαία εμπιστοσύνη *(Hardin, 2002)* κατά την οποία κάθε μέρος έχει κάποια εμπιστοσύνη για το άλλο, αλλά εξακολουθούν να υπάρχουν διαφορές στο πόσο εμπιστεύονται ο ένας τον άλλον. Για παράδειγμα, οι εργαζόμενοι συνήθως λένε ότι εμπιστεύονται τους επόπτες περισσότερο από ό, τι οι επόπτες εμπιστεύονται τους εργαζόμενους. Ασύμμετρη εμπιστοσύνη μπορεί να προκύψει σε οποιαδήποτε σχέση, και αναπαραστάσεις των σχέσεων εμπιστοσύνης σε μοντέλα κοινωνικών δικτύων πρέπει να επιτρέπουν τέτοιες ακραίες διαφοροποιήσεις.

3.4 Μετρικές εμπιστοσύνης

Οι μετρικές εμπιστοσύνης είναι αλγόριθμοι που σκοπό έχουν να προβλέψουν κατά πόσο πρέπει να θεωρείται αξιόπιστος προς κάθε χρήστη ένας άγνωστος προς αυτόν που όμως ανήκει στο δίκτυο εμπιστοσύνης. Η τιμή της αξιοπιστίας προκύπτει από τις αξιολογήσεις των υπολοίπων χρηστών του συστήματος (*Massa & Avesani 2007 , Ziegler & Lausen, 2004*) προς τον άγνωστο. Επίσης μπορούν να χρησιμοποιηθούν για τη διάδοση της εμπιστοσύνης ώστε να αυξηθεί το ποσοστό κάλυψης των χρηστών με υποδείξεις.

Η ανάγκη μέτρησης της εμπιστοσύνης και διαχείρισής της είναι πολύ διαδομένη και δεν περιορίζεται στο Σημασιολογικό Ιστό (Semantic Web) (*Blaze et al.,1996)*. Οι πρώτες προτάσεις για μέτρηση ανάγονται στις αρχές της δεκαετίας του ενενήντα, όταν η μέτρηση εμπιστοσύνης χρησιμοποιήθηκε για την υποστήριξη της Υποδομής Δημόσιου Κλειδιού (PKI) (*Zimmermann 1995*). Νέοι τομείς και πεδία έρευνας εκτός του PKI ασχολούνται δυναμικά με το θέμα από τότε όπως P2P δίκτυα, κινητές τεχνολογίες, και συστήματα βαθμολογιών σε online κοινότητες, όπου η συντήρηση των ρητών αρχών πιστοποίησης δεν είναι εφικτές πλέον και συνεπώς έστρεψαν το ερευνητικό ενδιαφέρον προς την εμπιστοσύνη. Η πληθώρα των διαθέσιμων μετρικών μπορεί ταξινομηθεί σε πολλούς άξονες. Ενδεικτικά αναφέρουμε τις εξής κατηγοριοποιήσεις:

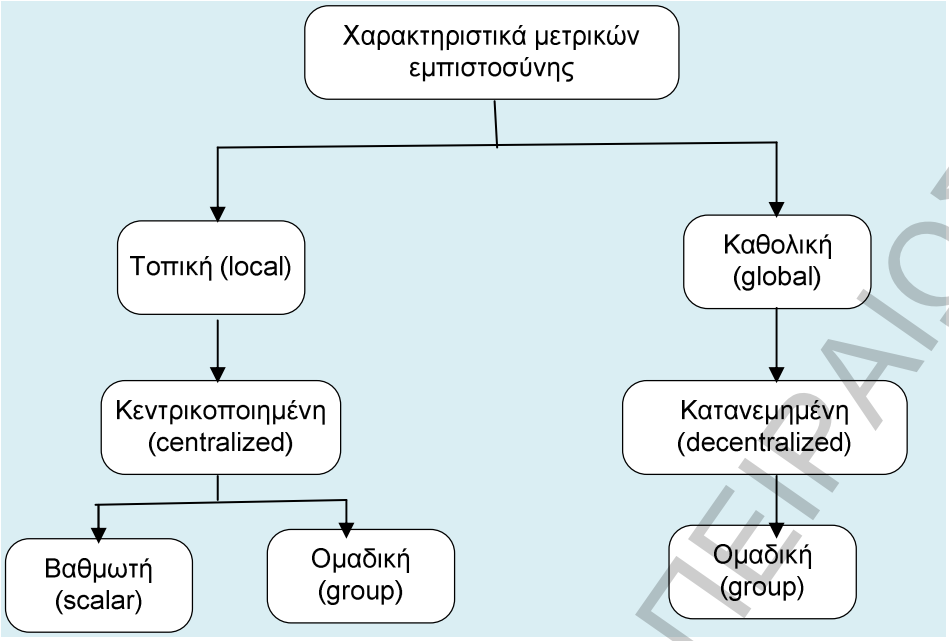
- Σύμφωνα με τους *Massa & Avesani 2007, Ziegler & Lausen, 2004*,*Massa & Avesani 2007*) η εμπιστοσύνη διαχωρίζεται σε τοπική και καθολική. Οι τοπικές μετρικές εμπιστοσύνης λαμβάνουν υπόψιν τους πολύ προσωπικές και υποκειμενικές απόψεις των χρηστών και καταλήγουν σε διαφορετικές τιμές για κάθε χρήστη. Ενώ οι καθολικές μετρικές προβλέπουν καθολικές τιμές δηλαδή για το πώς η κοινότητα συνολικά βλέπει το συγκεκριμένο χρήστη. Έτσι δεν στέκεται στην υποκειμενική άποψη των χρηστών αλλά χρησιμοποιεί τον κανονικοποιημένο μέσο όρο των απόψεων τους βάσει των δοθέντων τιμών. Οι τοπικές μετρικές υπολογιστικά είναι περισσότερο κοστοβόρες καθώς είναι απαραίτητος ο υπολογισμός για κάθε χρήστη του συστήματος τις τιμές εμπιστοσύνης τους για όλους τους υπόλοιπους χρήστες. Στο *Massa & Avesani (2007)* αναφέρεται ότι η διαφορά μεταξύ καθολικών και τοπικών μετρικών γίνεται ιδιαίτερα προφανής στο ζήτημα των αμφιλεγόμενων χρηστών (controversial users). Ως αμφιλεγόμενοι ορίζονται οι χρήστες που αξιολογούνται από τους υπόλοιπους χρήστες με πολύ διαφορετικούς τρόπους. Για παράδειγμα τους κρίνουν κάποιοι πολύ θετικά ενώ κάποιοι άλλοι πολύ αρνητικά. Αυτό σαν γεγονός δεν είναι περίεργο ή αδύνατο καθώς όπως έχει ήδη προαναφερθεί η εμπιστοσύνη είναι υποκειμενική έννοια και τα κριτήρια που την παράγουν διαφέρουν από χρήστη σε χρήστη. Κατά συνέπεια μία καθολική μετρική δεν μπορεί να δώσει σωστή πληροφορία για το επίπεδο εμπιστοσύνης του συγκεκριμένου χρήστη, θα πρέπει να μπουν πιο προσωπικά κριτήρια με τη χρήση τοπικών μετρικών για να μπορέσει το σύστημα να αποδώσει ορθή τιμή εμπιστοσύνης.

- Η εμπιστοσύνη χωρίζεται σε δύο κατηγορίες, τη διαπροσωπική εμπιστοσύνη που σχετίζεται με συγκεκριμένο πλαίσιο (context-specific interpersonal trust) όπου γίνεται επικέντρωση στο επίπεδο εμπιστοσύνης που δηλώνει ένας χρήστης για έναν άλλο για μία συγκεκριμένη κατάσταση και τη συστημική/απρόσωπη εμπιστοσύνη (system/impersonal trust) όπου ορίζεται η συνολική εμπιστοσύνη του συστήματος προς κάθε χρήστη. Η έννοια της εμπιστοσύνης αποκτά μία αυξανόμενη προσοχή ανάμεσα στα ερευνητικά πεδία και υπάρχουν αρκετές διαφορετικές οπτικές της που έχουν προταθεί για τον τρόπο μέτρησης και χρήστης της.

- Η εμπιστοσύνη του χρήστη (user trust) χωρίζεται σε άμεση εμπιστοσύνη (direct trust) και εμπιστοσύνη υπόδειξης (recommendation trust). Η άμεση εμπιστοσύνη αφορά την εμπιστοσύνη που δείχνει ο τρέχων χρήστης προς τον ενεργό ενώ η εμπιστοσύνη υπόδειξης σχετίζεται με την εμπιστοσύνη που δείχνει ένα σύνολο έμπιστων χρηστών του τρέχοντας για τον ενεργό. Το σταθμισμένο άθροισμα των δύο αυτών τιμών αποτελεί την εμπιστοσύνη του χρήστη.

- Επίσης βάσει του *Massa (2003)* γίνεται αναφορά ότι η εμπιστοσύνη μπορεί να είναι άμεση (direct) ή διαδεδομένη (propagated). Άμεση εμπιστοσύνη έχουμε όταν ο χρήστης Α έχει δηλώσει ρητώς πόσο αξιόλογο θεωρεί το χρήστη Β ενώ διαδεδομένη όταν ο χρήστης Α δεν έχει κάνει δήλωση για το χρήστη Β αλλά μέσω του δικτύου δημιουργείται αλυσίδα εμπιστοσύνης μεταξύ των δύο χρηστών πάνω στην οποία η εμπιστοσύνη μεταφέρεται ώστε να προκύψει πρόβλεψη για το πόσο ο Α θεωρεί αξιόπιστο τον Β

- Τέλος οι *Ziegler & Lausen, (2004)* εντοπίζει τρεις βασικές διαστάσεις με ιδιαίτερα χαρακτηριστικά. Οι άξονες αυτοί δεν είναι ορθογώνιοι, αν και διάφορα χαρακτηριστικά γνωρίσματα επιβάλλουν περιορισμούς στο εύρος των χαρακτηριστικών άλλων διαστάσεις. Κάποιες από τις κατηγορίες που αναφέρονται έχουν ήδη ορισθεί ξανά. Για παράδειγμα διακρίνονται σε τοπική (local) και καθολική (global) εμπιστοσύνη ενώ άλλη διάκριση αφορά το διαχωρισμό σε βαθμωτή (scalar) και ομαδική εμπιστοσύνη (group). Ωστόσο σύμφωνα με τους *Ziegler & Lausen, (2004)* δεν υπάρχει ρητή κατηγοριοποίηση των μετρικών εμπιστοσύνης κατά μήκος των διαφόρων αξόνων που να συμπληρώνεται με ανάλυση της αλληλεπίδρασης των αξόνων εικόνα 3



Εικόνα 2 κατηγοριοποίηση μετρικών εμπιστοσύνης

- **Προοπτική δικτύου (Network Perspective)**

Η διάσταση της προοπτικής του δικτύου επηρεάζει τη σημασιολογία που αποδίδεται στις υπολογιζόμενες τιμές της εμπιστοσύνης. Οι μετρικές εμπιστοσύνης μπορούν να διαχωρισθούν σε αυτές με καθολική (global) και αυτές με τοπική (local) εμβέλεια. Οι καθολικές μετρικές εμπιστοσύνης λαμβάνουν υπόψη όλες τις συνδέσεις και τους ισότιμους χρήστες που συνδέονται μεταξύ τους με σχέσεις εμπιστοσύνης. Οι καθολικές τιμές εμπιστοσύνης ανατίθενται για κάθε άτομο με βάση τις πλήρεις πληροφορίες από το γράφημα εμπιστοσύνης. Πολλές καθολικές μετρικές εμπιστοσύνης, δανείζεται τις ιδέες τους από το φημισμένο αλγόριθμο PageRank *(Page et al., 1998)* για τον υπολογισμό της φήμη μίας ιστοσελίδας. Ο PageRank αποτελεί έναν από τους βασικούς αλγορίθμους αναζήτησης της google σύμφωνα με τον οποίο ο ιστοχώρος είναι ένα δίκτυο περιεχομένου χωρίς κεντρικοποιημένο έλεγχο ποιότητας. Αυτή την εργασία προσπαθεί να ασκήσει PageRank ακολουθώντας την εξής απλή λογική: αν ένας σύνδεσμος από τη σελίδα προς τη σελίδα Β αναπαριστά μία θετική ψήφο της Α στη Β τότε η καθολική κατάταξη της σελίδας εξαρτάται από τον αριθμό και την ποιότητα των εξερχόμενων συνδέσμων της. Συνοψίζοντας, η βασική σκέψη πίσω από την προσέγγιση αυτή είναι ότι οι κόμβοι θα πρέπει να κατατάσσονται ψηλότερα όσο καλύτερη είναι η κατάταξη των κόμβων που δείχνουν σε αυτούς. Προφανώς, αυτή η προσέγγιση λειτουργεί για την εμπιστοσύνη και τη

φήμη της σελίδας επίσης. Από την άλλη πλευρά οι μετρικές εμπιστοσύνης με τοπική εμβέλεια λαμβάνουν υπόψιν την προσωπική προκατάληψη. Ενδιαφέρον παρουσιάζει ότι ορισμένοι ερευνητές υποστηρίζουν ότι μόνο τοπικές μετρικές εμπιστοσύνης είναι αληθινές , δεδομένου ότι οι παγκόσμιες μετρικές υπολογίζουν τη συνολική φήμη και όχι εξατομικευμένη εμπιστοσύνη *(Mui et al., 2002)*. Το σκεπτικό πίσω από τις τοπικές μετρικές εμπιστοσύνης είναι ότι τα άτομα που κάποιος εμπιστεύεται μπορεί να διαφέρουν πλήρως από εκείνα που εμπιστεύεται και θεωρεί αξιόπιστα κάποιος άλλος. Οι τοπικές μετρικές εμπιστοσύνης εκμεταλλεύονται τις πληροφορίες που ανταλλάσσονται ή μοιράζονται τα άτομα του συστήματος μέσω της δήλωσης εμπιστοσύνης.

- **Θέση που γίνονται οι υπολογισμοί (Computation Locus)**

Ο δεύτερος άξονας αφορά το χώρο όπου οι σχέσεις εμπιστοσύνης μεταξύ των ατόμων αξιολογούνται ποσοτικά. Τοπικές (με την έννοια του χώρου) ή κεντρικές προσεγγίσεις εκτελούν όλους τους υπολογισμούς σε ένα μόνο μηχάνημα και ως εκ τούτου θα πρέπει να έχουν πλήρη πρόσβαση στις σχετικές πληροφορίες εμπιστοσύνης. Τα δεδομένα εμπιστοσύνης μπορούν να διανεμηθούν στη συνέχεια μέσω του δικτύου.

Οι κατανεμημένες μετρικές για τον υπολογισμό της εμπιστοσύνης και της φήμης, επιφορτίζουν κάθε κόμβο του δίκτυο. Με τη λήψη πληροφοριών εμπιστοσύνης από προηγούμενους κόμβους στο γράφημα εμπιστοσύνης, ο τρέχων κόμβος συγχωνεύει τα δεδομένα που έλαβε με τους δικούς του ισχυρισμούς εμπιστοσύνης και διαδίδει τη σύνθεση των τιμών της στους κόμβους που τον διαδέχονται στο γράφο. Η διαδικασία υπολογισμού της εμπιστοσύνης είναι αναγκαστικά ασύγχρονη και η σύγκλισή της εξαρτάται από την οκνηρία ή αντίθετα την διάθεση των κόμβων για τη διάδοση της πληροφορίας. Ένα άλλο χαρακτηριστικό των κατανεμημένων μετρικών εμπιστοσύνης αναφέρεται στο γεγονός ότι είναι εγγενώς καθολικές. Αν και το επιμέρους υπολογιστικό φορτίο μειώνεται σε σχέση με τις κεντρικοποιημένες προσεγγίσεις υπολογισμού της εμπιστοσύνης, παρόλα αυτά στους κόμβους πρέπει να αποθηκεύονται πληροφορίες που σχετίζονται με την εμπιστοσύνη οποιουδήποτε άλλου κόμβο του συστήματος.

- **Αξιολόγηση σύνδεσης (Link Evaluation)**

Ο τελευταίος άξονας διακρίνει τις μονοδιάστατες (scalar) από τις ομαδικές (group) μετρικές εμπιστοσύνης. Οι μονοδιάστατες μετρικές αναλύουν τους ισχυρισμούς εμπιστοσύνης ανεξάρτητα, ενώ οι ομαδικές αξιολογούν σύνολα ισχυρισμών παράλληλα. Ο αλγόριθμος PageRank [29] και άλλες σχετικές προσεγγίσεις ανήκουν στις καθολικές και ομαδικές

μετρικές εμπιστοσύνης καθώς η φήμη μιας σελίδας εξαρτάται από την κατάταξη των αναφερόμενων σε αυτή σελίδων, πράγμα που συνεπάγεται παράλληλη αξιολόγηση αυτών των κόμβων λόγω της αμοιβαίας εξάρτησης. Ο αλγόριθμος Advogato αποτελεί ένα παράδειγμα τοπικής και ομαδικής μετρικής εμπιστοσύνης . Οι περισσότερες άλλες μετρικές ανήκουν στην κατηγορία των μονοδιάστατων, ακολουθώντας πορείες εμπιστοσύνης από τις πηγές στους στόχους , χωρίς να εκτελούν παράλληλες αξιολογήσεις των ομάδων δηλώνουν εμπιστοσύνη στο στόχο. Μία άλλη βασική διαφορά μεταξύ μονοδιάστατων μετρικών εμπιστοσύνης και των ομαδικών αφορά το λειτουργικό σχεδιασμό τους. Σε γενικές γραμμές οι μονοδιάστατες μετρικές υπολογίζουν την εμπιστοσύνης μεταξύ των δύο συγκεκριμένων προσώπων α και β τα οποία λαμβάνονται από το σύνολο των παραγόντων V. Από την άλλη πλευρά, οι ομαδικές μετρικές εμπιστοσύνης υπολογίζουν κατατάξεις εμπιστοσύνης για ομάδες ατόμων από το σύνολο V. Το συνολικό γράφημα εμπιστοσύνης είναι μόνο σημαντικό για τις καθολικές και ομαδικές μετρικές αλλά όχι για τις τοπικές. Ανεπίσημα, οι τοπικές ομαδικές μετρικές εμπιστοσύνης μπορούν να οριστούν ως μετρικές για τον υπολογισμό γειτονιών αξιόπιστων χρηστών ενός ατόμου α. Τέλος ότι οι μονοδιάστατες μετρικές είναι εγγενώς τοπικές, ενώ οι ομαδικές μετρικές εμπιστοσύνης δεν επιβάλλουν περιορισμούς σχετικά με τα χαρακτηριστικά που σχετίζονται με άλλους άξονες.

3.5 Εμπιστοσύνη και Συνεργατικές μέθοδοι υπόδειξης

3.5.1 Παραδοσιακή μορφή της συνεργατικής μεθόδου

Ο ανοικτός χαρακτήρας των συστημάτων υπόδειξης που χρησιμοποιούν τεχνικές συνεργατικής μεθόδου δίνει τη δυνατότητα σε κακόβουλους χρήστες να αποκτήσουν πρόσβαση στο σύστημα δημιουργώντας ψεύτικους πολλαπλούς λογαριασμούς με σκοπό να εισάγουν μεροληπτικά δεδομένα και να επηρεάσουν κατά το δοκούν το σύστημα. Οι παραδοσιακές τεχνικές συνεργατικών μεθόδων δεν μπορούν να προσφέρουν προστασία. Κατά συνέπεια η θωράκιση των συστημάτων υπόδειξης ώστε να προσφέρουν αμερόληπτες υπηρεσίες έχει αναδειχθεί σε σημαντικό ζήτημα.

Πρόσφατες έρευνες έχουν πραγματοποιηθεί πάνω σε θέματα ασφάλειας αυτών των συστημάτων με σκοπό να διασφαλισθεί η ακεραιότητα τους από κακόβουλα εγχειρήματα επηρεασμού των αποτελεσμάτων τους. Οι προσπάθειες μπορούν να κατηγοριοποιηθούν σε τεχνικές που αυξάνουν την ευρωστία των συστημάτων και σε τεχνικές που αποσκοπούν

στην ανεύρεση και την αναχαίτιση της επιρροής των μεροληπτικών ή/και κακόβουλων προφίλ.

Υπενθυμίζοντας τη συνεργατική μέθοδο (collaborative filtering) αναφέρεται ότι για κάθε χρήστη σχηματίζεται βάσει της ομοιότητα του τρέχοντος προφίλ και των υπολοίπων, μια γειτονιά από ομότιμους χρήστες που έχουν δηλώσει παρόμοιες προτιμήσεις. Στη συνέχεια προβλέπονται βαθμολογίες αγνώστων αντικειμένων για τον τρέχοντα χρήστη βάσει των βαθμολογιών που έχουν δώσει οι γείτονές του. Ο πυρήνας αυτού του αλγόριθμου είναι να υπολογιστεί η ομοιότητα μεταξύ των χρηστών. Υπάρχουν διάφορες μέθοδοι μπορούν να χρησιμοποιηθούν για να υπολογιστεί η ομοιότητα των χρηστών, όπως είναι ο συντελεστής συσχέτισης συνημίτονου (cosine correlation coefficient), η τροποποιημένη ομοιότητα συνημίτονου (modify cosine similarity), και ο συντελεστής συσχέτισης Pearson. Περισσότερο συχνή είναι η χρήστη του μέτρου Pearson (εξίσωση 9) για τον υπολογισμό της ομοιότητας μεταξύ των χρηστών.

$$sim_{u,v} = \frac{\sum_{i \in I}\left(r_{u,i}-\overline{r_u}\right)\left(r_{v,i}-\overline{r_v}\right)}{\sqrt{\sum_{i \in I}\left(r_{u,i}-\overline{r_u}\right)^2}\sqrt{\sum_{i \in I}\left(r_{v,i}-\overline{r_v}\right)^2}}$$

Εξίσωση 9

Όπου $r_{u,i}$ δηλώνει τη βαθμολογία του χρήστη u για το αντικείμενο i και $\overline{r_u}$ είναι ο μέσος όρος των βαθμολογιών που έχει δώσει ο χρήστης u για όλα τα αντικείμενα που έχει βαθμολογήσει. Αντίστοιχα $r_{v,i}$ είναι η βαθμολογία του χρήστη v για το αντικείμενο i και $\overline{r_v}$ είναι ο μέσος όρος των βαθμολογιών που έχει δώσει ο χρήστης u για όλα τα αντικείμενα που έχει βαθμολογήσει.

Για να υπολογίσουμε την πρόβλεψη βαθμολόγησης για ένα αντικείμενο i προς τον χρήστη u χρησιμοποιούμε την παρακάτω έκφραση:

$$p_{u,i} = \overline{r_u} + \frac{\sum_{v \in V} sim_{u,v}\left(r_{u,i}-\overline{r_u}\right)}{\sum_{v \in V}|sim_{u,v}|}$$

Εξίσωση 10

Όπου V είναι το σύνολο των k όμοιων χρηστών. Στους υπολογισμούς των προβλέψεων για κάθε αντικείμενο συμπεριλαμβάνονται μόνο όσοι έχουν βαθμολογήσει το αντικείμενο αυτό.

- **Αντιμετώπιση αδυναμιών συνεργατικής μεθόδου με την εμπιστοσύνη**

Η έννοια της εμπιστοσύνης έχει κερδίσει ερευνητικά την προσοχή όχι μόνο γιατί συμβάλλει στην παραγωγή καλύτερων αποτελεσμάτων από αυτά των κλασικών συνεργατικών μεθόδων που λαμβάνουν υπόψιν μόνο την ομοιότητα των χρηστών (*O'Donovan & Smyth 200)* αλλά και γιατί συνεισφέρει στη αντιμετώπιση των καταγεγραμμένων προβλημάτων και αδυναμιών των συνεργατικών μεθόδων αλλά και γενικότερα των συστημάτων υπόδειξης (*Massa 2003, Massa & Avesani 2007).*

Ως δήλωση εμπιστοσύνης ορίζεται ως μια ρητή βεβαίωση του γεγονότος ότι ένας χρήστης εμπιστεύεται έναν άλλο. Ο χρήστης Α μπορεί να προσδιορίσει τους χρήστες που εμπιστεύεται. Σύμφωνα με τον *Massa (2003)* μια δήλωση εμπιστοσύνης από το χρήστη Α προς Β χρήστη σημαίνει ότι ο χρήστης Α αξιολογεί σταθερά αξιόπιστα τα σχόλια και τις βαθμολογίες του χρήστη Β.

Ένα δίκτυο εμπιστοσύνης (trust network) ή ένα κοινωνικό δίκτυο (social network) δομείται με βάση τα συγκεντρωτικά στοιχεία των δηλώσεων εμπιστοσύνης που γίνεται από την πλευρά των χρηστών. Ένα δίκτυο εμπιστοσύνης αποτελεί ένα γράφημα στο οποίο κόμβοι είναι οι χρήστες και οι κατευθυνόμενες ακμές αποτελούν την εμπιστοσύνη που δηλώνουν οι χρήστες προς τους άλλους κόμβους του συστήματος. Διαισθητικά συμπεραίνεται ότι αν ο χρήστης Α εμπιστεύεται το χρήστη Β και ο Β τον Γ είναι δυνατόν να συμπεράνουμε κάτι για το πόσο μπορεί να εμπιστευτεί ο Α τον Γ. Αυτή η μεταφορά της εμπιστοσύνης μπορεί να χρησιμοποιηθεί για την εξαγωγή χρήσιμων πληροφοριών για τη διατύπωση συστάσεων και ξεπεραστούν οι αδυναμίες των συστημάτων υπόδειξης για μια πραγματική κοινότητα.

Όπως έχει αναφερθεί στις αδυναμίες των συνεργατικών μεθόδων συγκαταλέγεται η σποραδικότητα των δεδομένων στον πίνακα όπου αποθηκεύονται οι βαθμολογίες των χρηστών για τα αντικείμενα του συστήματος. Δηλαδή είναι πολύ μικρό το ποσοστό των βαθμολογημένων αντικειμένων από χρήστες σε πραγματικά συστήματα. Σε δημοφιλή συστήματα όπως το Epinions η σποραδικότητα ξεπερνά το 99.99% στο *Eachmovie* το 97.6% και στο *Movielens* το 95.8%.

Το πρόβλημα έγκειται στο ότι ο συντελεστής συσχέτισης Pearson μπορεί να υπολογιστεί μόνο σε επικαλυπτόμενα αντικείμενα (δηλαδή στα κοινά βαθμολογημένα αντικείμενα μεταξύ χρηστών). Έτσι, για οποιοδήποτε δεδομένο χρήστη, ο αριθμός των άλλων χρηστών με τους οποίους είναι δυνατόν να υπολογιστούν ομοιότητας είναι χαμηλή. Αντ 'αυτού, η μετάδοση εμπιστοσύνης πάνω από το κοινωνικό δίκτυο επιτρέπει να «καλυφθεί» μια μεγαλύτερη μερίδα της βάσης χρηστών. Με αυτό τον τρόπο είναι δυνατόν να υπολογισθεί ένα εναλλακτικό βάρος σχετικά με το πόσο πρέπει να λάβουμε άλλους χρήστες υπόψη κατά τον υπολογισμό μιας υπόδειξης (*Massa 2003*).

Η κρύα έναρξη (cold start) είναι το φαινόμενο όπου νέοι χρήστες εισέρχονται στο σύστημα και δεν έχουν αξιολογήσει αρκετά αντικείμενα ώστε να είναι δυνατή η παραγωγή υποδείξεων αξιόλογης ποιότητας. Οι χρήστες που βρίσκονται στη φάση της κρύας εκκίνησης είναι κρίσιμης σημασίας για τα συστήματα και μπορούν να ωφεληθούν ιδιαίτερα από την διάδοση εμπιστοσύνη εφ 'όσον έχουν τουλάχιστον ένα έμπιστο φίλο στην κοινότητα. Αυτό μπορεί να είναι ένας αποτελεσματικός μηχανισμός για τη γρήγορη και αποτελεσματική ενσωμάτωση νέων χρηστών, ειδικά σε σύγκριση τις παραδοσιακές τεχνικές όπου οι χρήστες χρειάζεται να αξιολογήσουν τουλάχιστον 10 αντικείμενα πριν να λάβουν υποδείξεις από το σύστημα

Τα συστήματα που λειτουργούν με συνεργατικές μεθόδους αντιμετωπίζουν προβλήματα κλιμάκωσης καθώς χρειάζεται υπολογισμός των γειτόνων επί του συνόλου των χρηστών για τον εκάστοτε τρέχοντα χρήστη. Ωστόσο, μπορεί να μειωθεί ο αριθμός των χρηστών που πρέπει να εξετάσει το σύστημα φιλτράροντάς τους ώστε οι εναπομείναντες να έχουν αξιολογηθεί ως αξιόπιστοι. Για παράδειγμα, θα μπορούσαν να ληφθούν υπόψιν μόνο οι χρήστες που βρίσκονται σε μικρή απόσταση στο κοινωνικό δίκτυο από τρέχοντα χρήστη ή στην εξέταση να συμπεριλαμβάνονται μόνο χρήστες με προβλεπόμενη εμπιστοσύνη μεγαλύτερη από ένα ορισμένο κατώφλι *(Massa & Avesani 2007)*.

Επίσης πολύ σημαντική είναι η συνεισφορά που μπορεί να έχει η εμπιστοσύνη στην αντιμετώπιση των επιθέσεων των κακόβουλων χρηστών (*Ziegler & Lausen, 2004)..* Σύμφωνα με τους *Eschenauer et al.,(2002)* μπορεί να βοηθήσει στην αύξηση της αντίστασης του συστήματος στις επιθέσεις. Δηλαδή μπορεί να χρησιμοποιηθούν για τον εντοπισμό κακόβουλων χρηστών λαμβάνοντας υπόψη μόνο αξιόπιστες βαθμολογίες βάσει του επιπέδου εμπιστοσύνης που λαμβάνουν οι χρήστες που τις έχουν δηλώσει. Θα πρέπει να επισημανθεί ωστόσο ότι δεν υπάρχει μια σφαιρική εικόνα για το ποιοι χρήστες είναι

αξιόπιστοι ή εμπιστοσύνης καθώς οι απόψεις μπορεί να ποικίλλουν ανάμεσα στους χρήστες. Στην αντιμετώπιση των κακόβουλων χρηστών σημαντική ίσως να έχει η δυσπιστία (distrust) που όμως δεν έχει ερευνηθεί καθότι νέα έννοια και δεν μπορούν να εξαχθούν συμπεράσματα για την αποτελεσματικότητά της.

Τέλος, βάσει του *Massa (2003)* ακόμη και αν η μετρική της εμπιστοσύνη μπορεί να εισαχθεί σε ένα ενιαίο κεντρικό διακομιστή, αυτές οι προσεγγίσεις συλλογής δεδομένων έχουν τα ακόλουθα τεράστια μειονεκτήματα. Εκφράζοντας πληροφορίες (τι αρέσει, πού αρέσει) σε έναν κεντρικό διακομιστή σημαίνει ότι μόνο ο διακομιστής θα είναι σε θέση να χρησιμοποιήσει αυτές τις πληροφορίες. Αυτό έχει ως αποτέλεσμα το προφίλ των χρηστών να διασπάται σε τμήματα και να διαμερίζεται σε πολλούς διαφορετικούς servers οι οποίοι δεν συνεργάζονται μεταξύ τους και κάθε μεμονωμένος διακομιστής υποφέρει ακόμη περισσότερο από σποραδικότητα στα δεδομένα. Επιπλέον, αυτό σημαίνει ότι ο χρήστης δεν μπορεί να μετακινηθεί από το ένα στο άλλο σύστημα υπόδειξης χωρίς να χάσει το προφίλ του (και με αυτό, τη δυνατότητα να λαμβάνουν καλές συστάσεις).

- **Τιμές εμπιστοσύνης**

Η εμπιστοσύνη αποτελεί πληροφορία που σχετίζεται με τις κοινωνικές σχέσεις και ως εκ τούτου, σε ένα web-based κοινωνικό δίκτυο αναπαριστάται ως μια ετικέτα για αυτή τη σχέση. Υπάρχει πολλή ελευθερία ως προς το ποια μορφή παίρνει αυτή η ετικέτα στη συνέχεια αναφέρονται οι πιθανές επιλογές. Παραδείγματος χάριν το e-cademy χρησιμοποιεί την απλούστερη δυνατή εκπροσώπηση της εμπιστοσύνης. Οι χρήστες δηλαδή έχουν δύο επιλογές: να μην κάνουν καμία δήλωση σχετικά με την εμπιστοσύνη, ή δηλώσουν ότι ένας φίλος είναι «αξιόπιστος». Αυτό δεν επιτρέπει οποιαδήποτε κατάταξη αξιοπιστίας ή αναξιοπιστίας. Επιτρέπει στους χρήστες απλά να υποδείξουν ποια άτομα εμπιστεύονται. Υπάρχουν ορισμένα είδη σχέσεων που εύκολα μπορούν να αναπαρασταθούν με τον συγκεκριμένο τρόπο. Για παράδειγμα, έστω ότι σχετίζονται με ένα πρόσωπο, αν έχουμε συναντήσει ένα άτομο, ή αν είμαστε συνάδελφοι, είναι μια σχέση που υπάρχει ή δεν υπάρχει. Η δήλωση της εμπιστοσύνη, ωστόσο, δεν είναι τόσο απλή.

Σύμφωνα με την Golbeck η αναπαράσταση της τιμής της εμπιστοσύνης μπορεί να συμπεριλαμβάνει:

- Ρητή δήλωση εμπιστοσύνης (όλοι λαμβάνουν τη δήλωση έχουν τιμή εμπιστοσύνης 1 ενώ για όσους δεν υπάρχει δήλωση η τιμή της εμπιστοσύνης προς αυτούς είναι 0)

- Κλίμακα με συνεχές εύρος πχ οι *Richardson et al. (*2003), που χρησιμοποίησαν ένα συνεχές εύρος [0-1]
- Κλίμακα με ετικέτες αντί για αριθμούς (π.χ. «πολύ χαμηλή εμπιστοσύνη», «χαμηλή εμπιστοσύνη», «μέτρια εμπιστοσύνη», υψηλή εμπιστοσύνη» και «πολύ υψηλή εμπιστοσύνη»).
- Τέλος αναφέρει ότι υπάρχει δυνατότητας να υπάρξουν και συστήματα κατάταξης εμπιστοσύνης που χρησιμοποιούν ρητές αξιολογήσεις των χρηστών (πχ σχόλια) και με τη χρήση μηχανισμών εκμαίευση προτιμήσεων να οικοδομήσουν ένα προφίλ της εμπιστοσύνης των χρηστών.

3.6 Ερευνητικές προσεγγίσεις

Από τους *(O' Donovan & Smyth ,2005)* προτείνεται η έννοια της εμπιστοσύνης σε δύο επίπεδα. Εμπιστοσύνη σε επίπεδο αντικειμένου (item level trust) και σε επίπεδο προφίλ χρήστη (profile level trust). Η εμπιστοσύνη σε επίπεδο προφίλ υπολογίζεται βάσει του ποσοστού των ορθών προβλέψεων που έχει κάνει ο χρήστης που διαχειρίζεται το συγκεκριμένο προφίλ. Αντίστοιχα η εμπιστοσύνη σε επίπεδο αντικειμένου υπολογίζεται βάσει της εμπιστοσύνης που έχουν δηλώσει οι χρήστες του συστήματος. Η σκέψη πάνω στην οποία βασίζεται η πρόταση είναι ότι χρήστες που παρέχουν ορθότερες προτάσεις είναι περισσότερο αξιόπιστοι σε σχέση με χρήστες που παρέχουν προτάσεις χαμηλής ποιότητας. Η εμπιστοσύνη σε επίπεδο προφίλ αποδείχθηκε πιο ανθεκτική σε κακόβουλες επιθέσεις και κυρίως σε τυχαίες επιθέσεις (τυχαία επιλογή αντικειμένων που δομούν τα κακόβουλα προφίλ και τυχαία επιλογή της βαθμολογίας τους).

Από τον *Ziegler (2005)* στη διατριβή του προτείνεται η χρήση εμπιστοσύνης παρόμοια με τη συγκεκριμένη (*Massa et al., 2007)* όπου η τοπική εμπιστοσύνη χρησιμοποιείται για την δόμηση της γειτονιάς των χρηστών. Από τους *(O' Donovan & Smyth ,2005)* προτείνεται επίσης ότι οι υβριδικές μέθοδοι (συνδυαστική χρήση τεχνικών που στηρίζονται στο περιεχόμενο και συνεργατικών μεθόδων) έχουν μεγαλύτερη αποτελεσματικότητα με σχέση με τη χρήση μεμονωμένης κατηγορίας τεχνικής.

Η *Goldbeck (2005)* στη διατριβή της μελετά την εμπιστοσύνη στα διαδικτυακά κοινωνικά δίκτυα (web-based social networks) και στο πώς αυτή μπορεί να υπολογισθεί. Ο αλγόριθμος που ανέπτυξε και συμπεριλαμβάνει την μετρική της εμπιστοσύνης είναι ο TidalTrust . Οι χρήστες δηλώνουν την εμπιστοσύνη τους για άλλους χρήστες

βαθμολογώντας μέσω κλίμακας. Η εμπιστοσύνη που προκύπτει είναι άμεση δηλαδή από χρήστη σε χρήστη χωρίς να προβλέπεται διάδοσή της.

Από τους *Montaner et al., (2002)* προτάσσεται ένα μοντέλο εμπιστοσύνης βάσει του οποίο οι χρήστες λαμβάνουν υποδείξεις από μία ομάδα χρηστών που στηρίζεται και δομείται στις δηλώσεις εμπιστοσύνης. Η εμπιστοσύνη εξαρτάται από την ικανοποίηση του χρήστη από το προτεινόμενο αντικείμενο και η τιμή εμπιστοσύνη μεταβάλλεται ανάλογα. Πρόβλημα σε αυτή τη μέθοδο αποτελεί η έλλειψη δηλώσεων εμπιστοσύνης μεταξύ των χρηστών στην αρχή της διαδικασίας υπόδειξης *(cold start)*.

Από τους *Massa & Avesani (2004)* προτάσσεται μέθοδος κατά την οποία όταν ένα αντικείμενο προτάσσεται σε κάποιον χρήστη στη συνέχεια ζητείται από αυτόν να αξιολογήσει την υπόδειξη. Η βαθμολογία που δίνει ο τρέχων χρήστης στο αντικείμενο που του προτάθηκε αποτελεί την εμπιστοσύνη που δείχνει ο χρήστη που βαθμολογεί προς το χρήστη που προτείνει. Η εμπιστοσύνη διαδίδεται στο δίκτυο χρησιμοποιώντας τις υπάρχουσες σχέσεις. Δηλαδή υπολογίζεται η εμπιστοσύνη ενός χρήστη προς ένα άλλο που όμως δεν ενώνονται άμεσα στο γράφο του συστήματος. Ως απόσταση δύο χρηστών ορίζεται ο αριθμός των ακμών που αποτελούν το μονοπάτι που τους ενώνει. Δηλαδή αν ο χρήστης Α εμπιστεύεται το χρήστη Β και ο Β τον C τότε μέσω της διάδοσης μπορεί να προβλεφθεί η εμπιστοσύνη του Α προς τον C. Αυτή η ιδέα αποτελεί μία αποτελεσματική προσπάθεια για τη δημιουργία ενός δικτύου εμπιστοσύνης . Όμως λόγω έλλειψης ορίων στη διάδοση μπορεί να προκληθεί υπερχείλιση και βγει εκτός ελέγχου η διάδοση εμπιστοσύνης. Επίσης όσο μεγαλύτερη είναι η διάδοση τόσο αυξάνεται η κάλυψη των χρηστών αλλά μειώνεται η ορθότητα των προβλέψεων. Για αυτό το λόγο είναι σημαντική η ισορροπία ανάμεσα στις δύο μετρικές.

Οι *Massa & Avesani, 2009* για να αντιμετωπίσουν το πρόβλημα της σποραδικότητας των δεδομένων και της κρύας έναρξης *(cold start)* πρότειναν τη χρήστη της εμπιστοσύνης ως βάρος στη διαδικασία πρόβλεψης αντί για το παραδοσιακό βάρος της ομοιότητας καθώς έτσι αυξάνεται η κάλυψη του δικτύου μέσω της διάδοσης και άθροισης της εμπιστοσύνης, Σκοπός είναι να χρησιμοποιηθούν στη διαδικασία οι πιο αξιόπιστοι χρήστες σύμφωνα με τις δηλώσεις του τρέχοντα χρήστη.

3.7 Υπολογισμός Εμπιστοσύνης

Δεδομένου ενός κοινωνικό δίκτυο η παροχή πληροφοριών σχετικά με την εμπιστοσύνη που απολαμβάνουν οι χρήστες σε τοπικό επίπεδο (από χρήστη σε χρήστη) ή καθολικά (πόσο αξιόπιστος θεωρείται από την κοινότητα συνολικά) μπορεί να παραχθεί με πολλούς τρόπους. Ο στόχος όμως εν γένει πάντα ο ίδιος: να υπάρξει τιμή εμπιστοσύνης για όλους τους χρήστες του δικτύου.

Στα περισσότερα κοινωνικά δίκτυα ή δίκτυα εμπιστοσύνης λίγες είναι οι άμεσες δηλώσεις εμπιστοσύνης από χρήστη σε χρήση λόγω του μεγάλου όγκου των ατόμων. Έτσι είναι περιορισμένο το πεδίο των αντικειμένων τα οποία μπορούν να προταθούν των οποίων η επιλογή στηρίζεται στις δηλώσεις αξιοπιστίας. Οι έμπιστοι χρήστες είναι καλοί στις υποδείξεις των *Massa & Avesani (2007),* οπότε είναι σημαντικό να αυξηθούν οι σχέσεις εμπιστοσύνης στο δίκτυο. Επειδή όπως αναφέρθηκε και προηγουμένως οι περισσότεροι χρήστες είναι άγνωστοι μεταξύ τους πρέπει να υπάρξει πρόβλεψη στο κατά πόσο ο χρήστης Α που δεν γνωρίζει άμεσα το χρήστη Β μπορεί να τον εμπιστευτεί. Για να μπορέσει να υπάρξει πρόβλεψη δημιουργείται μία αλυσίδα ή ένα μονοπάτι στο γράφο ανάμεσα στο κόμβο του χρήστη Α και σε αυτό του Β μέσω ενδιάμεσων κόμβων-χρηστών ακολουθώντας τις δηλώσεις εμπιστοσύνης και μεταφέροντας έτσι την τιμή της.

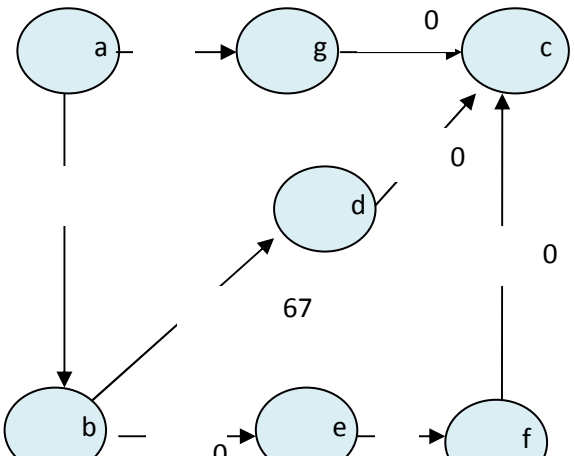
Η εμπιστοσύνη όταν δεν είναι άμεση (direct) διαδίδεται (trust propagation) στο δίκτυο βασιζόμενη στη ιδιότητα της μεταβατικότητας (transitivity) και στη συνέχεια αθροίζεται με διάφορες τεχνικές ώστε να προβλεφθεί η τιμή εμπιστοσύνης από τον Α στο Β (trust aggregation) μέσω της συνθεσιμότητας (composability). Η ιδιότητα της μεταβατικότητας δεν ισχύει σε όλες τις περιπτώσεις. Για παράδειγμα αν ο Α εμπιστεύεται τον Β για το γούστο του στη μουσική και ο Β εμπιστεύεται τον Γ για τις γνώσεις του σε οικονομικά ζητήματα δεν συνεπάγεται ότι ο Α θα εμπιστεύεται τον Γ σε οικονομικά ζητήματα. Η εμπιστοσύνη για να μεταβιβασθεί μέσω του δικτύου θα πρέπει να αφορά παρεμφερή πεδίο αν όχι το ίδιο.

Η άμεση εμπιστοσύνη μπορεί να δηλωθεί από ένα χρήστη για έναν άλλο μέσω κλίμακας αξιολόγησης *(Massa & Avesani 2007)*. Σε αυτό το μοντέλο δεν λαμβάνεται υπόψιν ο χρόνος της δήλωσης έτσι σε περίπτωση που ένας χρήστης αλλάξει την τιμή εμπιστοσύνης του προς έναν άλλο η τιμή θα ανανεωθεί , αντικαθιστώντας την προηγούμενη. Κατά τους *Massa & Avesani (2007)* η άμεση εμπιστοσύνη είτε υπάρχει είτε όχι. Δηλαδή δεν δηλώνεται πόσο εμπιστεύεται ένας χρήστης τον άλλο αλλά αν τον εμπιστεύεται ή όχι .Αν τον εμπιστεύεται

εισέρχεται στο «Web of Trust» του χρήστη ενώ αν δεν τον εμπιστεύεται στη «Blacklist» του. Η μη δήλωση εμπιστοσύνης δεν αποτελεί έλλειψη εμπιστοσύνης αλλά αδιαφορία. Η Golbeck στο TidalTrust χρησιμοποιεί δεκάβαθμη κλίμακα εμπιστοσύνης όπου το 1 δηλώνει χαμηλή εμπιστοσύνη και το 10 υψηλή. Τέλος υπάρχουν και μοντέλα [T6] που δεν στηρίζουν την τιμή της άμεσης εμπιστοσύνης σε ρητή απόκριση του χρήστη στο κατά πόσο εμπιστεύεται έναν άλλο αλλά την εκμαιεύουν συγκρίνοντας τις κοινές βαθμολογίες τους .

Αντίστοιχα υπάρχουν διαφορετικές οπτικές για τον υπολογισμό της διαδιδόμενης εμπιστοσύνης. Σημείο σύγκλισης όμως είναι ότι τα μονοπάτια με μεγαλύτερη εμπιστοσύνη προσφέρουν υπόδειξης καλύτερης ποιότητας όπως τα πιο κοντινά μονοπάτια (Golbeck et al. 2004, *Massa & Avesani 2007)* Ενώ ο πιο συνήθης τρόπος υπολογισμού της είναι μέσω πιθανοτήτων, υπολογίζοντας το γινόμενο των επιμέρους τιμών εμπιστοσύνης που προκύπτουν από το μονοπάτι που ενώνει το χρήστη Α και το χρήστη Β. Δηλαδή αν ο Α εμπιστεύεται τον Β με 0.8 και ο Β τον Γ με 0.6 , ο Α εμπιστεύεται τον Γ με πιθανότητα 0.8* 0.6 = 0.48 . Άλλες προσεγγίσεις χρησιμοποιούν κλασικούς τελεστές άθροισης, όπως ο μέγιστος (*maximum*), ο ελάχιστος (*minimum*), το σταθμισμένο άθροισμα (*weighted sum*), ο μέσος όρος (*average*), ή ο σταθμισμένος μέσος όρος (*weighted average*) των τιμών της εμπιστοσύνης κατά μήκος το μονοπατιού από το χρήστη Α στο Γ (*Golbeck 2003*) . Βασικό πλεονέκτημα των σταθμισμένων τελεστών είναι ότι θεωρούν μερικούς κόμβους (χρήστες) πιο σημαντικούς από άλλους δίνοντάς τους μεγαλύτερο βάρος και κατά συνέπεια δημιουργώντας μια διαδικασία υπολογισμού της εμπιστοσύνης περισσότερο διάφανη. Έρευνες (*Massa & Avesani 2007,* Golbeck et al. 2004, *Ziegler & Lausen, 2004*) έχουν δείξει ότι η απόσταση της διάδοσης επηρεάζει σημαντικά την ορθότητα (accuracy).

Η διάδοση και η συνάθροιση της εμπιστοσύνης συχνά συνδυάζονται και η τελική εκτίμησή της εξαρτάται από τον τρόπο υλοποίησης *(Kantor, Ricci, Rokach, & B. Shapira, 2010)*. Στα πραγματικά δίκτυα συχνά υπάρχουν εναλλακτικές διαδρομές που ενώνουν δύο κόμβους. Αυτά τα μονοπάτια δεν είναι αναγκαίο να έχουν το ίδιο μήκος ούτε να μεταφέρουν την ίδια τιμή εμπιστοσύνης (Εικόνα 3).



Εικόνα 3 Δίκτυο εμπιστοσύνης όπως οι κόμβοι συνδέονται με πολλαπλές διαδρομές

Όπως φαίνεται στην εικόνα 3 υπάρχουν τρεις τρόποι να διαδοθεί η εμπιστοσύνη από το χρήστη a στον χρήστη *c* μέσω του *b*. Να χρησιμοποιηθεί ένας τελεστής άθροισης από τον *b* στον *d* και από τον *d* στον *c* , για το δεύτερο μονοπάτι ένας από τον *b* στον *e* από τον *e* στον *f* και από τον *f* στον *c* ενώ στο τρίτο μονοπάτι χρησιμοποιείται ένα τελεστής άθροισης από τον *g* στον *c* Στο τέλος αθροίζονται τα δύο μεταδιδόμενα αποτελέσματα εμπιστοσύνης. Σε αυτό το παράδειγμα η εμπιστοσύνη αρχικά διαδίδεται και στη συνέχεια αθροίζεται (*first propagate then aggregate*) (*Kantor, Ricci, Rokach, & B. Shapira, 2010)*

Εναλλακτική επιλογή η αντίστροφη διαδικασία, δηλ. αρχικά να αθροιστεί και στη συνέχεια να διαδοθεί η εμπιστοσύνη (*First Aggregate Then Propagate*), σε αυτή την περίπτωση ο ενδιάμεσος χρήστης *b* θα πρέπει αρχικά να συναθροίσει τις τιμές που έλαβε από τους *d* και *e* και στη ακολούθως να διαδώσει τη νέα τιμή στον *a*. Σε αυτή την περίπτωση οι κόμβοι του γράφου (χρήστες) έχουν μεγαλύτερη υπευθυνότητα σε σχέση με το πρώτο σενάριο και ότι ο υπολογισμός της εμπιστοσύνης μπορεί να πραγματοποιηθεί σε ένα κατανεμημένο περιβάλλον, χωρίς οι χρήστες να χρειάζεται να εκφράσουν την προσωπική τους τιμή προς τους άλλους. Υπολογίζοντας το παραπάνω παράδειγμα και χρησιμοποιώντας το *γινόμενο* ως τελεστή διάδοσης προκύπτει ότι για το πρώτο σενάριο (*First propagate then aggregate*) η εμπιστοσύνη του *a* στον *c* βάσει της διαδρομής *a→b→d→c*, είναι ίση με: 1*0.5*0.4=0.2, η εμπιστοσύνη του *a* στον *c* βάσει της διαδρομής *a→g→c*, είναι ίση με: 1*0.9=0.9 ενώ μέσω της διαδρομής *a→b→e→f→c*, είναι ίση με: 1*0.8*0.6*0.7=0.34. Με χρήση του *μέσου όρου* ως τελεστή άθροισης, η τελική εκτίμηση εμπιστοσύνης θα είναι ίση με: (0.9+0.2+0.34)/3≈ 0.48. Τέλος βάσει του δεύτερου σεναρίου (*First Aggregate Then Propagate*) ο χρήστης *b* θα περάσει τη τιμή (0.2+0.34)/2≈ 0.27 στον χρήστη *a* και ο *g* 0.9/1=0.9 με η τελική τιμή εμπιστοσύνης (0.9+0.27)/2≈ 0.59.

Όσο αυξάνεται το μήκος του μονοπατιού μειώνεται η αξιοπιστία του ενώ όσο μεγαλύτερη είναι η διάδοση της εμπιστοσύνης τόσο αυξάνεται η κάλυψη των χρηστών του δικτύου. Ουσιαστικά υπάρχει ένα συμβιβασμός μεταξύ κάλυψης (coverage) και ορθότητας (accuracy) .Μοντέλα που λαμβάνουν υπόψιν το μήκος του μονοπατιού (distance) για τον υπολογισμό της εμπιστοσύνης είναι αναφέρεται στο *Massa & Avesani (2007)* και η Golbeck που λαμβάνει επιλέγει μόνο τα κοντινότερα μονοπάτια αγνοώντας τα υπόλοιπα.

3.8 Εξάγοντας εμπιστοσύνη από ένα δίκτυο

Τα περισσότερα συστήματα που επιχειρούν να ενσωματώσουν την έννοια της εμπιστοσύνης στη διαδικασία παραγωγής υποδείξεων ζητούν από αυτούς ρητές δηλώσεις για τους άλλους τρόπους που έχουν ήδη αναφερθεί όπως δήλωση φιλίας, κλίμακα αξιολόγηση κτλ. Τέτοια παραδείγματα είναι τα το το Moleskiing *(Massa et al., 2005)* ένα site για ορειβατικό σκι όπου οι χρήστες καταγράφουν την εμπιστοσύνη τους σε μια κλίμακα 1 έως 9 και το epinions.com όπου γίνονται δηλώσεις εμπιστοσύνης αλλά και έλλειψης εμπιστοσύνης. Ένα άλλο πολύ γνωστό παράδειγμα είναι FilmTrust (*Golbeck, 2006),* ένα online κοινωνικό δίκτυο στο οποίο οι χρήστες καλούνται να αξιολογήσουν τους γνωστούς τους σε μια κλίμακα από το 1 έως το 10 εκτός της δυνατότητας αξιολόγησης των αντικειμένων του συστήματος. Όλα αυτά τα συστήματα μετατρέπουν την εμπιστοσύνη άμεση ή παραγόμενη σε βάρος που χρησιμοποιείται για να δηλώσει πόσο πρέπει ο κάθε χρήστης να θεωρηθεί σημαντικός στη διαδικασία παραγωγής υποδείξεων του κάθε τρέχοντα χρήστη. Στη συνέχεια θα γίνει αναφορά στις πιο συχνά χρησιμοποιούμενες στρατηγικές, δηλαδή την κλασική εκδοχή του σταθμισμένου μέσου όρου και εκδοχές συνεργατικών μεθόδων που ενσωματώνουν την εμπιστοσύνη.

3.8.1 Trust-based weighted mean

Σε ένα σύστημα υποδείξεων που δεν αποτελεί δίκτυο εμπιστοσύνης, ένας απλός αλγόριθμός που παράγει προτάσεις στους χρήστες του προσπαθεί να εκτιμήσει κατά πόσο ο τρέχων χρήστης βρίσκει ενδιαφέρον το αντικείμενο *i*. Για το σκοπό αυτό υπολογίζεται η μέση βαθμολογία που έχει λάβει το συγκεκριμένο αντικείμενο λαμβάνοντας υπόψη κάθε βαθμολογία $r_{u,i}$ που έχει πάρει από τους χρήστες του συστήματος u . Η απλή αυτή στρατηγική μπορεί να βελτιωθεί ενσωματώνοντας αυτή βασική σύσταση μπορεί να βελτιωθεί με τη χρήση ενός σταθμισμένου μέσου που στηρίζεται στην εμπιστοσύνη. Ειδικότερα, με τη συμπερίληψη του σταθμισμένου μέσου $t_{a,u}$, που αντικατοπτρίζει το

βαθμό που κάθε χρήστης που προτείνει είναι αξιόπιστος , ο αλγόριθμος επιτρέπει τη διάκριση μεταξύ των πηγών που συμμετέχουν στη παραγωγή συστάσεων. Είναι λογικό ότι δίνεται μεγαλύτερη βαρύτητα στις αξιολογήσεις που προέρχονται από χρήστες υψηλής εμπιστοσύνης. Ο τύπος δίδεται από την Εξίσωση 3, στην οποία $p_{a,i}$ συμβολίζει την προβλεπόμενη βαθμολογία του αντικειμένου στόχου i για το τρέχοντα χρήστη, και το R^T παριστά το σύνολο των χρηστών που έχουν αξιολογήσει το *i* και για τους οποίους η τιμή εμπιστοσύνης που τους δίνεται από τον τρέχοντα χρήστη υπερβαίνει μία ορισμένη τιμή κατωφλίου.

$$p_{a,i} = \frac{\sum_{u \in R^T} t_{a,u} r_{u,i}}{\sum} t_{a,u}$$

Εξίσωση 11

3.8.2 TidalTrust

Αυτός ο αλγόριθμος αποτελεί το κεντρικό σημείο του συστήματος υπόδειξής της Golbeck, Η καινοτομία αυτού του αλγορίθμου αυτού που ονομάζεται TidalTrust έγκειται στον τρόπο που η εμπιστοσύνη *ta,u* μεταφέρεται στο γράφο. Σύμφωνα με το (*Golbeck et al., 2006)* οι συγγραφείς πριν την υλοποίηση του αλγορίθμου έτρεξαν αρκετά πειράματα κρύβοντας κάθε φορά μία σχέση εμπιστοσύνης του δικτύου και παρατήρησαν ότι οι διαδρομές με τους λιγότερους ενδιάμεσους κόμβους (shortest path) δίνουν ακριβέστερες εκτιμήσεις για την εμπιστοσύνη και ότι τα μονοπάτια που έχουν υψηλότερες τιμές εμπιστοσύνης αποδίδουν κι αυτές πολύ καλύτερα αποτελέσματα

Επομένως, λαμβάνοντας υπόψη μόνο την πρώτη παρατήρηση επιλέγοντας μόνο μικρότερες διαδρομές θα πρέπει να λαμβάνονται καλύτερα αποτελέσματα. Ωστόσο, σε ορισμένες περιπτώσεις μόνο λίγοι χρήστες θα είναι προσπελάσιμοι αν τεθεί όριο στο μήκος του μονοπατιού κυρίως σε αραιά συστήματα όπου υπάρχουν λίγες δηλώσεις εμπιστοσύνης μεταξύ των χρηστών. Αυτή η αντιστάθμιση αντιμετωπίζεται με τη χρήστη μεταβλητής στο μέγιστο επιτρεπόμενο μέγεθος μονοπατιού που απαιτείται για τη σύνδεση του τρέχοντα χρήστη με το χρήστη u. Η μεταβλητή αποτελεί το βάθος του μονοπατιού του κάθε αλγορίθμου. Ανάλογα το δίκτυο και τους υπολογιστικούς σκοπούς η μέγιστη τιμή των βημάτων του μονοπατιού ποικίλλει.

Ένας τρόπος για την αντιμετώπιση της δεύτερη παρατήρησης (υψηλότερες τιμές εμπιστοσύνη στην πορεία αποδίδουν καλύτερες εκτιμήσεις εμπιστοσύνης) είναι να περιορίσει τις προσλαμβάνουσες πληροφορίες, ώστε αυτές να προέρχονται από τους πιο

έμπιστους χρήστες. Ωστόσο, κάθε χρήστης έχει τη δική της συμπεριφορά για την παραγωγή της εμπιστοσύνη τους προς τους άλλους (ένας χρήστης μπορεί να δίνει τη μέγιστη τιμή αρκετά συχνά, ενώ κάποιος άλλος ποτέ), και, επιπλέον αρκετές φορές συμβαίνει να υπάρχουν μόνο λίγα μονοπάτια με την ίδια υψηλή τιμή εμπιστοσύνης . Αυτός είναι ο λόγος επέλεξαν να ενσωματώσουν μια τιμή που αντιπροσωπεύει τη δύναμη του μονοπατιού (δηλαδή την ελάχιστη βαθμολογία εμπιστοσύνης της διαδρομής), και να υπολογίζουν τη μέγιστη δύναμη του μονοπατιού πάνω από όλα τα μονοπάτια που οδηγούν στους χρήστες που έχουν βαθμολογήσει τα αντικείμενα στόχους. Αυτό το μέγιστο τότε επιλέγεται ως το ελάχιστο όριο εμπιστοσύνης για τη συμμετοχή στη διαδικασία.

Η βασική εξίσωση του TidalTrust δίνεται από την Εξίσωση 12, στην οποία ο όρος WOT^+ αντιπροσωπεύει το σύνολο των χρηστών για τους οποίους έχει γίνει δήλωση εμπιστοσύνης η τιμή της οποίας υπερβαίνει μία ορισμένη τιμή κατωφλίου *max*. Αυτό σημαίνει ότι κάθε χρήστης κατά τη διαδικασία υπολογίζει την εμπιστοσύνη του προς κάθε άλλο χρήστη ως σταθμισμένος μέσος όρος, και λαμβάνει υπόψη τις πληροφορίες που προέρχονται από χρήστες για τους οποίους ο ίδιος έχει δώσει βαθμολογίες τουλάχιστον όσο το κατώφλι *max*.

$$t_{a,u} = \frac{\sum_{v \in WOT^+(a)} t_{a,v}\, t_{v,u}}{\sum_{v \in WOT^+(a)} t_{a,v}}$$

Εξίσωση 12

Ο TidalTrust είναι ένας αναδρομικός αλγόριθμος . Η τιμή της εμπιστοσύνη $t_{a,u}$ υπολογίζεται αναδρομικά ως ο σταθμισμένος μέσος όρος των τιμών εμπιστοσύνης $t_{v,u}$, για όλους τους χρήστες v που αποτελούν πρώτο σύνδεσμο στη συντομότερη διαδρομή από τον *a* έως το *u*. Οι χρήστες εξασφαλίζουν ότι το μέγιστο βάθος διαδρομής δεν υπερβαίνεται παρακολουθώντας το μήκος της τρέχουσας διαδρομής. Αυτός ο αλγόριθμος αυτός ανήκει στην κατηγορία βαθμωτών (gradual) προσεγγίσεων εμπιστοσύνης και είναι παράδειγμα τοπικής μετρικής εμπιστοσύνης (local trust metric).

3.8.3 Trust-based collaborative filtering

Ενώ το φιλτράρισμα της Golbeck είναι ένα παράδειγμα εφαρμογής σταθμισμένου μέσου όρου, μια άλλη κατηγορία συστημάτων που βασίζονται στην εμπιστοσύνη συνδέεται στενότερα με τεχνικές αλγορίθμων συνεργατικής μεθόδου. Στη συνεργατική μέθοδο η βαθμολογία του αντικειμένου στόχου *i* για το χρήστη *a* μπορεί να προβλεφθεί χρησιμοποιώντας τις αξιολογήσεις των γειτόνων του (βάσει της ομοιότητας των χρηστών) που είναι ήδη εξοικειωμένοι με το στοιχείο *i*. Η κλασικός τύπος δίνεται από την Εξίσωση 13.

$$p_{a,i} = \overline{r_a} + \frac{\sum_{u \in R^+} w_{a,u}\left(r_{u,i} - \overline{r_u}\right)}{\sum_{u \in R^+} w_{a,u}}$$

Εξίσωση 13

$$p_{a,i} = \overline{r_a} + \frac{\sum_{u \in R^T} t_{a,u}\left(r_{u,i} - \overline{r_u}\right)}{\sum_{u \in R^T} t_{a,u}}$$

Εξίσωση 14

Η άγνωστη βαθμολογία $p_{a,i}$ για το αντικείμενο *i* και τον χρήστη στόχος *α* υπολογίζεται με βάση τη μέση βαθμολογία $\overline{r_a}$ του χρήστη *α* των αντικειμένων που έχει ήδη βαθμολογήσει καθώς και για την βαθμολογία $r_{u,i}$ από τους άλλους χρήστες για *u* για το αντικείμενο *i*. Η εξίσωση λαμβάνει επίσης υπόψη την ομοιότητα $w_{a,u}$ μεταξύ των χρηστών *α* και *u*, η οποία συνήθως υπολογίζεται βάσει του συντελεστή συσχέτισης Pearson (PCC). Στην πράξη, τις περισσότερες φορές χρησιμοποιούνται μόνο οι χρήστες με θετική συσχέτιση $w_{a,u}$ που έχουν αξιολογήσει το *i*. Αυτό το σύνολο χρηστών ορίζεται ως R^+. Ωστόσο, αντί για υπολογισμό των βαρών με τη χρήστη του PCC, είναι εφικτή η χρήση βαρών που προέρχονται από τις αναπτυσσόμενες σχέσεις εμπιστοσύνης του τρέχοντα χρήστη στο δίκτυο. Η εξίσωση 14 προσαρμόζει την εξίσωση 13 αντικαθιστώντας τα βάρη $w_{a,u}$ που υπολογίζονται με το PCC από τις τιμές εμπιστοσύνης $t_{a,u}$. Η στρατηγική αυτή υποστηρίζεται επίσης από το γεγονός ότι η εμπιστοσύνη και ομοιότητας συσχετίζονται *(Massa & Avesani 2007, Golbeck 2004)*.

3.8.4 MoleTrust

Η εξίσωση 14 αποτελεί τη βάση του αλγορίθμου σύστασης *(Massa & Avesani , 2007)* η οποία ενσωματώνει μία νέα μετρική εμπιστοσύνης που ονομάζεται MoleTrust. Αυτή η μετρική υλοποιείται σε δύο στάδια. Στο πρώτο στάδιο, οι κύκλοι στο δίκτυο εμπιστοσύνης αφαιρούνται, ενώ το δεύτερο στάδιο περιλαμβάνει τον υπολογισμό της εμπιστοσύνης. Δεδομένου ότι είναι συχνή η περίπτωση όπου ένας μεγάλος αριθμός διαδοχικών διαδόσεων πρέπει να εκτελεσθούν κατά τη διάρκεια των πειραμάτων (π.χ. ότι τα μεγάλα σύνολα δοκιμών από Epinions.com), είναι πολύ πιο αποτελεσματικό να αφαιρεθούν οι κύκλοι εμπιστοσύνη εκ των προτέρων, έτσι ώστε κάθε χρήστης χρειάζεται να επισκεφθεί μία φορά μόνο για να αποκτήσει πρόβλεψη εμπιστοσύνης.

Η αφαίρεση των κύκλων μετατρέπει το αρχικό δίκτυο εμπιστοσύνης σε ένα κατευθυνόμενο και μη κυκλικό γράφημα, και ως εκ τούτου η πρόβλεψη για την εμπιστοσύνη $t_{a,u}$ μπορεί να εξασφαλισθεί κάνοντας μια απλή διαδρομή στο γράφημα: πρώτον, η εμπιστοσύνη των χρηστών με απόσταση 1 υπολογίζεται (δηλαδή, άμεση εμπιστοσύνη), στη συνέχεια υπολογίζεται η εμπιστοσύνη των χρηστών σε απόσταση 2, κ.λπ. Σημειώστε ότι λόγω της μη κυκλικής φύσης του γραφήματος, η τιμή της εμπιστοσύνη ενός χρήστη σε απόσταση *x* εξαρτάται μόνο από τις τιμές εμπιστοσύνης των χρηστών που υπολογίστηκαν ήδη στην απόσταση x -1.

Η εμπιστοσύνη των χρηστών με απόσταση 2 ή περισσότερο υπολογίζεται με παρόμοιο τρόπο με αυτόν του αλγορίθμου της Golbeck. Ωστόσο υπάρχουν διαφορές. Στον TidalTrust, ο χρήστης *u* προστίθεται στο WOT^{+} *(α)* μόνο εάν ανήκει στη συντομότερη διαδρομή από το χρήστη στόχο *α* για το αντικείμενο *i*. Από την άλλη, στον αλγόριθμο MoleTrust, το σύνολο WOT^{+} *(α)* περιλαμβάνει όλους τους χρήστες οι οποίοι έχουν αξιολογήσει το αντικείμενο στόχος *i* και για τους οποίους είναι εφικτός ο υπολογισμός εμπιστοσύνης μέσω άμεσης ή έμμεσης σχέσης που στηρίζεται στη διάδοση της αξιοπιστίας μέσω μονοπατιού στο δίκτυο. Αλλά η εμπιστοσύνη δεν υπολογίζεται για άπειρες αποστάσει: πριν από την έναρξη της διαδικασίας υπολογισμού, πρέπει να εκχωρείται μια τιμή d στην παράμετρο της «διάδοσης στον ορίζοντα» (propagation horizon). Μέσω αυτής της εκχώρηση ς, μόνο οι χρήστες που είναι προσβάσιμοι σε απόσταση d λαμβάνονται υπόψη. Μία άλλη σημαντική παράμετρος εισόδου του MoleTrust είναι το όριο εμπιστοσύνης για συμμετοχή στη διαδικασία (σε αντίθεση με τη δυναμική τιμή κατωφλίου max στον TidalTrust), η οποία για παράδειγμα ορίζεται σε 0,6 (η τιμή αυτή ανήκει στο [0,1]).

Αντίστοιχα με τον TidalTrust ο αλγόριθμος MoleTrust ανήκει στην κατηγορία των βαθμωτών τοπικών μετρικών εμπιστοσύνης. Στα πειράματά τους, οι Massa και Avesani κατέδειξαν ότι ο MoleTrust παρέχει καλύτερες εκτιμήσεις εμπιστοσύνης από καθλικές μετρικές εμπιστοσύνης όπως αυτές που χρησιμοποιούνται από το eBay, ειδικά όταν πρόκειται για εκτίμηση της εμπιστοσύνης προς αμφιλεγόμενη χρήστες (οι οποίοι αξιολογούνται ως αξιόπιστοι από μία ομάδα και αναξιόπιστοι από μία άλλη). Απέδειξαν επίσης ότι ο MoleTrust παράγει ακριβέστερες προβλέψεις για τους νέους χρήστες του συστήματος (cold start users), σε σύγκριση με μία παραδοσιακή τεχνική συνεργατικής μεθόδου.

Οι προσεγγίσεις των *Golbeck (2004)* και *Massa & Avesani (2007),*είναι δύο χαρακτηριστικά παραδείγματα τεχνικών υπόδειξης που έχουν βελτιώσει την αποτελεσματικότητα παραδοσιακών τεχνικών με τη χρήση της εμπιστοσύνης.

3.9 Automatic Trust Generation

Οι αλγόριθμοι που συζητήθηκαν στην προηγούμενη ενότητα απαιτείται ρητή δήλωση εμπιστοσύνης από τους χρήστες. Κατά συνέπεια, οι εφαρμογές που χρησιμοποιούν ένα τέτοιο αλγόριθμο πρέπει να παρέχουν ένα μέσο για να λαμβάνουν τις απαραίτητες πληροφορίες. Όπως π.χ. το FilmTrust ή Moleskiing. Ωστόσο, αυτό ενδέχεται να μην είναι πάντα δυνατό ή εφικτό. Σε τέτοιες περιπτώσεις, μέθοδοι που συνάγουν αυτόματα εκτιμήσεις εμπιστοσύνης, χωρίς να χρειάζεται η ρητή εισαγωγή της πληροφορίας από τους χρήστες θα μπορούσε να είναι μια καλύτερη λύση (*O'Donovan & Smyth ,2005*) Συνήθως οι προσεγγίσεις αυτές βασίζουν τον υπολογισμό της εμπιστοσύνη στην προηγούμενη βαθμολογική συμπεριφορά των χρηστών του συστήματος. Πιο συγκεκριμένα, η απόφαση σε ποιο βαθμό ένας συγκεκριμένος χρήστης θα πρέπει να συμμετάσχει στη διαδικασία υπόδειξης επηρεάζεται στο κατά πόσο έχει πραγματοποιήσει ακριβείς υποδείξεις κατά το παρελθόν. Γνωστό παράδειγμα αυτής της προσέγγισης αποτελεί η έρευνα των *O'Donovan & Smyth (2005)*

3.9.1 Profile- & item-level trust

Ένας χρήστης με ιστορικό καλώ υποδείξεων μπορεί να θεωρηθεί ως πιο αξιόπιστος από ότι άλλοι χρήστες με λιγότερο καλές αποδόσεις. Για να είναι σε θέση να επιλέξει τους πιο αξιόπιστους χρήστες του συστήματος, ο O'Donovan εισήγαγε δύο μετρικές εμπιστοσύνης, την εμπιστοσύνη σε επίπεδο προφίλ (*profile-level trust*) και την εμπιστοσύνη σε επίπεδο αντικειμένου *item-level trust)*. αντανακλώντας τη γενικότερη αξιοπιστία του συγκεκριμένου χρήστη *u*, και την αξιοπιστία του *u* χρήστη σε σχέση με ένα συγκεκριμένο αντικείμενο *i* αντίστοιχα. Και για τις δύο μετρικές εμπιστοσύνη πρέπει να υπολογισθούν η ορθότητα των συστάσεων του χρήστη *u* για τον τρέχοντα χρήστη *α*. Ειδικότερα, μια πρόβλεψη pa, i που παράγεται μόνο από τις πληροφορίες που προέρχονται από τον χρήστη *u* (επομένως ο *u* είναι ο μοναδικός που συνεισφέρει στην υπόδειξη) θεωρείται σωστή εάν η τιμή pa, i για το χρήστη *α* είναι εντός του διαστήματος λάθους *ε* της πραγματική βαθμολογία ra του χρήστη α.

Η εμπιστοσύνη σε επίπεδο προφίλ t_u^p για το χρήστη *u* τότε ορίζεται ως το ποσοστό των σωστών συστάσεων στις οποίες συνέβαλε ο *u*. Παρατήρηση αποτελεί ότι αυτό είναι ένα πολύ γενικό μέτρο εμπιστοσύνης. Στην πράξη συμβαίνει συχνά ότι ο χρήστης *u* αποδίδει καλύτερα συνιστώντας ένα σύνολο συγκεκριμένων στοιχείων. Για το σκοπό αυτό, προτείνεται επίσης από τον O'Donovan μια περισσότερο εκλεπτυσμένη μετρική την εμπιστοσύνη σε επίπεδο αντικειμένου t_u^i, η οποία μετρά το ποσοστό των ορθών συστάσεων για το αντικείμενο *i*. Ως εκ τούτου, σε τέτοιες αυτοματοποιημένες προσεγγίσεις, οι τιμές δεν εμπιστοσύνης δεν υπολογίζονται μέσω διάδοσης και άθροισης, αλλά με βάση τις αξιολογήσεις που δόθηκαν στο παρελθόν. Οι μετρικές που προτείνονται από τον O'Donovan είναι καθολική εμπιστοσύνης. Βάσει του τρόπου με τον οποίο λαμβάνονται οι τιμές εμπιστοσύνης μπορούν να θεωρηθούν ως πιθανοτικές.

3.9.2 Trust-based filtering

Παρόμοια με τις άλλες τεχνικές που ενσωματώνουν την εμπιστοσύνη, οι τιμές που λαμβάνονται μέσω της εμπιστοσύνης χρησιμοποιείται ως βάρη στη διαδικασία παραγωγής υποδείξεων. Ακριβώς όπως ο *Massa (2004)* και ο *O'Donovan et al (2005*). Επικεντρώθηκε στις προσαρμογές που μπορούν να γίνουν μέσω της εμπιστοσύνης στη συνεργατική μέθοδο . Μια εναλλακτική λύση της χρησιμοποίησης ως βάρος της εμπιστοσύνης αντί του βαθμού ομοιότητας είναι η χρήση της εμπιστοσύνης ως φίλτρο επιλογή των γειτόνων, έτσι ώστε μόνο οι πιο αξιόπιστοι γείτονες συμμετέχουν στη διαδικασία σύστασης. Αυτή η στρατηγική ονομάζεται φιλτράρισμα με βάση την εμπιστοσύνη, εξίσωση 15 στην οποία ο όρος $w_{a,u}$, δηλώνει την ομοιότητα όπως προκύπτει από PCC και $\mathbf{R^{T+}} = \mathbf{R^T} \cap \mathbf{R^+}$.

$$p_{a,i} = \overline{r_a} + \frac{\sum_{u \in R^{T+}} w_{a,u}\left(r_{u,i} - \overline{r_u}\right)}{\sum_{u \in R^T} w_{a,u}}$$

Εξίσωση 15

Με άλλα λόγια, οι χρήστες των οποίων η εμπιστοσύνη σε επίπεδο αντικειμένου/ προφίλ υπερβαίνει ένα ορισμένο όριο, και που έχουν μια θετική συσχέτιση με τον τρέχοντα χρήστη, λαμβάνονται υπόψη.

Οι *O'Donovan & Smyth (2005*) έδειξαν ότι η εμπιστοσύνη που βασίζεται φιλτράρισμα επιτυγχάνει μεγαλύτερη ακρίβεια από ό, τι το συνεργατικό φιλτράρισμα από την άποψη του μέσου όρου λαθών. Επιπλέον, ο αλγόριθμος χρησιμοποιεί την εμπιστοσύνη σε επίπεδο

προφίλ αποδίδει λιγότερα λάθη σε σχέση με την παραδοσιακή τεχνική σχεδόν στο 70% του συνόλου των προβλέψεων.

Η πρόταση των *O'Donovan & Smyth (2005*)αποτελεί ένα αντιπροσωπευτικό παράδειγμα στην ομάδα των στρατηγικών που χρησιμοποιούν αυτόματη δημιουργία εμπιστοσύνης.

Τέλος υπάρχει και η έρευνα Τ6 όπου η εμπιστοσύνη του χρήστη σπάει σε άμεση εμπιστοσύνη και εμπιστοσύνη υπόδειξης, Η άμεση εμπιστοσύνη επηρεασμένοι από τους *O'Donovan & Smyth (2005*) δεν προέρχεται από ρητή δήλωση του χρήστη για την αξιοπιστία κάποιου άλλου αλλά από τις βαθμολογίες του παρελθόντος. Η εμπιστοσύνη υπόδειξης μεσοσταθμικά από την άμεση εμπιστοσύνη που δίνουν στο συγκεκριμένο προφίλ/ αντικείμενο οι γείτονες του τρέχοντα χρήστη. Στο τέλος η άμεση εμπιστοσύνη και η υπόδειξης προσθέτονται έχοντας η κάθε μία ένα βάρος (α και β όπου α+β=1) ώστε ανάλογα με την περίσταση να επιλέγεται ποιος τύπος εμπιστοσύνης πρέπει να ληφθεί πιο σοβαρά υπόψιν στον υπολογισμό της τιμής αξιοπιστίας.

4 Κακόβουλες πρακτικές

4.1 Εισαγωγή

Καθώς τα συστήματα υποδείξεων πρέπει να είναι ανοικτά στην άποψη των χρηστών είναι πολύ δύσκολη η δημιουργία συστημάτων που δεν θα είναι ευάλωτα σε κακόβουλες επιθέσεις. Κατά συνέπεια είναι αναγκαία η ανίχνευση και η περιθωριοποίηση των κακόβουλων προφίλ και των συνεπειών της ύπαρξής τους ώστε να μην λαμβάνονται υπόψιν κατά τη διαδικασία παραγωγής υποδείξεων προς τους χρήστες.

Οι επιθυμίες και ανάγκες των χρηστών μπορούν να καλυφθούν από μία γκάμα προσφερόμενων επιλογών. Συχνά όμως δεν υπάρχει χρόνος ούτε η διάθεση για την αξιολόγηση όλων των εναλλακτικών ώστε να προκύψει η βέλτιστη πρόταση. Τα συστήματα υποδείξεων καλούνται να πράξουν την εξαντλητική αξιολόγηση των εναλλακτικών και να λύσουν αποδοτικά το πρόβλημα της υπερπληροφόρησης. Τα συστήματα υπόδειξης εκτός της προφανούς χρησιμότητας που προσφέρουν στους χρήστες αποτελούν ένα σημαντικό σύστημα για τους ιδιοκτήτες του. Εκτός του ότι βελτιώνουν τις προσφερόμενες υπηρεσίες προς το χρήστη, οδηγούν σε πιστότητα (customer loyalty) που σχετίζεται έντονα με την αύξηση των πωλήσεων.

Οι υψηλές βαθμολογίες σχετίζονται με καλό επίπεδο πωλήσεων και για αυτό το λόγο τέτοια συστήματα αποτελούν συχνά στόχο προμηθευτών ή εν γένει κακόβουλων χρηστών ώστε να επηρεάσουν τις υποδείξεις που παράγονται. Τα συστήματα συνεργατικής υπόδειξης αποτελούν τον πιο συνήθη τρόπο προσωποποίησης στο ίντερνετ και ταυτόχρονα είναι γνωστά λόγω του ότι είναι πολύ ευάλωτα σε επιθέσεις εισαγωγής ψεύτικων προφίλ.
Τα συστήματα που χρησιμοποιούνται στο ηλεκτρονικό εμπόριο είναι αναγκαίο να προστατεύονται γιατί διατρέχουν μεγάλη η πιθανότητα επίθεσης που θα βάλει σε κίνδυνο την ποιότητα των προβλέψεων και την εμπιστοσύνη των χρηστών προς αυτό.

Την παρούσα χρονική στιγμή υπάρχουν αρκετά εμπορικά συστήματα υπόδειξης που χρησιμοποιούνται στο ηλεκτρονικό εμπόριο για πρόταση ταινιών (πχ movielens.org, tivo.org), μουσικής (πχ udioscrobbler.com) αλλά και για ερευνητικούς σκοπούς (πχ Pocketlens) και κυβερνητικούς (πχ το recommender system της ΝΑΣΑ για την υπόδειξη συσχετιζόμενων τεχνικών αναφορών). Σε τέτοιου είδους συστήματα οι χρήστες χτίζουν προφίλ βαθμολογώντας αντικείμενα του συστήματος.

Οι αλγόριθμοι συνεργατικής μεθόδου λειτουργούν βασιζόμενοι στη σκέψη ότι παρόμοιοι άνθρωποι μοιάζουν και στις επιλογές τους. Δηλαδή άτομα με παρόμοια συμπεριφορά είναι πιθανόν να αρέσκονται στις ίδιες διασκεδάσεις και να κάνουν παρόμοιες επιλογές. Κατά κανόνα αλγόριθμοι αυτής της κατηγορίας λειτουργούν σε δύο στάδια:

- Το πρώτο στάδιο είναι αυτό της πρόβλεψης όπου ο αλγόριθμος απλά υπολογίζει κατά πόσο ένα αντικείμενο ή μία ομάδα αντικειμένων θα αρέσουν στον τρέχοντα χρήστη (δηλαδή το χρήστη για τον οποίο προορίζεται η πρόβλεψη). Τα αντικείμενα μπορεί να έχουν επιλεχθεί από το χρήση μέσω αναζήτησης ή περιήγησης στο σύστημα.
- Το δεύτερο στάδιο είναι αυτό της υπόδειξης όπου ο αλγόριθμος δημιουργεί μία λίστα στην οποία τα αντικείμενα που την αποτελούν κατατάσσονται σειριακά βάσει της βαθμολογίας που έλαβαν για το πόσο προβλέπεται ότι θα είναι αρεστά στο χρήστη. Οι βαθμολογία αυτή προκύπτει από βαθμολογίες άλλων χρηστών που πιθανώς να είναι άγνωστοι στον τρέχοντα χρήστη και που επιλέγονται μέσω της συσχέτισης ανάμεσα στις δικές του βαθμολογίες και των υπολοίπων χρηστών.

Όσοι παράγουν και εμπορεύονται αντικείμενα επιθυμούν να έχουν την καλύτερη δυνατή πορεία στις αγορές και να πουλήσουν. Στα συστήματα υπόδειξης υπάρχει το κίνητρο στους παραγωγούς να επιθυμούν τα προϊόντα τους να προτείνονται συχνότερα από αυτά των άλλων. Αυτό μπορεί να επιτευχθεί παράγοντας προϊόντα υψηλής ποιότητας που ικανοποιούν αποδοτικά τις ανάγκες των χρηστών κι αυτοί με τη σειρά τους να αξιολογούν με υψηλή βαθμολογία το προϊόν. Όμως κακόβουλοι παραγωγοί μπορεί να επιλέξουν δρόμο απάτης για να πετύχουν το σκοπό τους. Δηλαδή να προσπαθήσουν να επηρεάσουν το σύστημα υπόδειξης με τέτοιο τρόπο ώστε το προϊόν τους να προτάσσεται πολύ συχνά στους χρήστες ανεξαρτήτως αν έχει λάβει με αδιάφθορο τρόπο τις απαιτούμενες βαθμολογίες και ανεξαρτήτως ποιότητας του αντικειμένου. Ένα τέτοιο παράδειγμα κακόβουλου παραγωγού τεράστιας φήμης είναι η περίπτωση της Sony Pictures το 2001 όπου παραδέχθηκε ότι προσπάθησε να διαβάλει σύστημα υπόδειξης ταινιών εισάγοντας ψεύτικα προφίλ κριτικών στο σύστημα που ήταν ανύπαρκτα άτομα με σκοπό την προώθηση νέων ταινιών που είχαν μόλις κυκλοφορήσει. Επίσης η εταιρεία Amazon έχει ανακοινώσει δύο τρία περιστατικά όπου έγιναν προσπάθεια διαβολής του συστήματος υπόδειξής της . Ανάλογα περιστατικά έχει αναδείξει και το e-Bay όπου συχνά βρίσκεται σε θέση να πρέπει να αντιμετωπίσει χρήστες που προσπαθούν με διάφορους τρόπους να εκμεταλλευτούν το σύστημα υπέρ των επιδιώξεών τους.

Οι επιθέσεις στα συστήματα υπόδειξης μπορεί να επηρεάσουν την ποιότητα των προβλέψεων για πολλούς χρήστες , μειώνοντας κατά συνέπεια την ικανοποίησή τους έναντι του συστήματος. Όταν ο χρήστης αισθανθεί ότι χάνει χρόνο ή χρήμα χρησιμοποιώντας σύστημα αυτό, θα στραφεί σε εναλλακτικές λύσεις δηλαδή σε άλλα συστήματα που προσφέρουν παρόμοιες υπηρεσίες) δημιουργώντας έτσι πρόβλημα στον ιδιοκτήτη του συστήματος ο οποίος πρέπει να αντιμετωπίσει την επίθεση γρήγορα και αποτελεσματικά για να μην χάσει τους χρήστες του.

Η απόδοση οποιουδήποτε συστήματος μπορεί να μετρηθεί βάσει κάποιας συνάρτησης χρησιμότητας που μοντελοποιεί τη χρησιμότητα μίας υπόδειξης. Προφανώς η w()συνάρτηση δομείται βάσει των οπτικών των εμπλεκόμενων μερών. Στις συνεργατικές μεθόδους μπορούν να ανιχνευτούν τρεις κατηγορίες:

- Αρχικά έχουμε τους τελικούς χρήστες (**end-users**) για τους οποίους παράγονται οι υποδείξεις. Για αυτή την κατηγορία χρηστών σημασία έχει οι υποδείξεις που τους προτείνονται να είναι συμβατές με το γούστο και τις επιλογές τους.
- Δεύτερη κατηγορία είναι ο ιδιοκτήτης του συστήματος (**owner**) που τον ενδιαφέρει ο αριθμός των συναλλαγών που γίνονται μέσω του συστήματος. Για εκείνον δεν έχει τόση σημασία η ορθότητα των προτάσεων όσο το σύστημα να του φέρνει νέους πελάτες.
- Τρίτη κατηγορία είναι οι εξωτερικές ενδιαφερόμενες οντότητες που έχουν έμμεσο ενδιαφέρον για τις συναλλαγές που πραγματοποιούνται μέσω του συγκεκριμένου συστήματος. Για ένα σύστημα υπόδειξης βιβλίων τέτοια οντότητα μπορεί να είναι ένας συγγραφέας . Αυτές οι οντότητες ενδιαφέρονται για τις προτάσεις που γίνονται για τα προϊόντα που τους ενδιαφέρουν αλλά δεν έχουν ευθεία πρόσβαση στη βάση δεδομένων.

4.2 Βιβλιογραφική σύνοψη

Οι *Chirita et al., (2005)* αναφέρουν τεχνικές που σκοπό έχουν να ανιχνεύσουν τα κακόβουλα προφίλ βάσει του μοτίβου των βαθμολογιών των χρηστών που σχετίζονται με κάθε προφίλ. Εισάγουν τη μετρική RDMA και ένα νέο αλγόριθμο για την ανίχνευση των κακόβουλων προφίλ. Η προσέγγιση έχει μεγαλύτερη επιτυχία σε προφίλ που έχουν βαθμολογήσει πολλά αντικείμενα ενώ έχει χαμηλότερο βαθμό επιτυχίας σε μικρά προφίλ. Οι *Burke et al., (2006)* προτείνει έναν αλγόριθμο που λαμβάνει υπόψιν πολλές ιδιότητες που διαχωρίζουν τα κακόβουλα από τα αυθεντικά προφίλ και οδηγεί σε καλύτερα αποτελέσματα από αυτά των *Chirita et al., (2005)*. Ο αλγόριθμος ανιχνεύει προφίλ που έχουν δημιουργηθεί βάσει μοντέλων για τα οποία ο αλγόριθμος έχει εκπαιδευτεί. Μειονέκτημα του συγκεκριμένου αλγορίθμου είναι ότι δυσκολεύεται να ανιχνεύσει τα κακόβουλα προφίλ όταν έχουν ενσωματώσει θόρυβο στις βαθμολογίες τους.

Οι *O΄Mahony et al., (2005)* ασχολήθηκαν με την ανίχνευση θορύβου στα δεδομένα χρησιμοποιώντας τη signal detection theory η οποία βάσει των συμπερασμάτων τους αποδείχθηκε ιδιαίτερα αποτελεσματική στην αύξηση της ευρωστίας ενός συστήματος. Ο θόρυβος δεν ταυτίζεται με τους κακόβουλους χρήστες αλλά και πάλι μειώνει την αξιοπιστία του συστήματος.

Επίσης λαμβάνουν ως αφετηρία τη δημιουργία γειτονιάς για κάθε χρήστη μέσω του αλγορίθμου k-NN αλλά αντί να χρησιμοποιήσουν ως βάρος (w) την ομοιότητα μεταξύ των χρηστών χρησιμοποιούν την χρησιμότητα (utility). Ως χρησιμότητα ορίζουν μία καθολική μεταβλητή που μετρά την χρησιμότητα κάθε χρήστη ως γείτονα στο σύστημα.

Οι *Massa & Avesani, (2004)* εισήγαγαν την έννοια της εμπιστοσύνης ως τρόπο σύνδεσης των χρηστών πέραν της ομοιότητας. Η χρήση της εμπιστοσύνης ως βάρος στη διαδικασία πρόβλεψης αντί για το παραδοσιακό βάρος της ομοιότητας αυξάνει την ορθότητα και τη ευρωστία του συστήματος χρησιμοποιώντας στη διαδικασία τους πιο αξιόπιστους χρήστες.

Οι *O' Donovan & Smyth (2006)* χρησιμοποίησαν την έννοια της εμπιστοσύνης , στην καθολική της μορφή (αριθμός επιτυχημένων προβλέψεων που συμμετείχε ο ενεργός χρήστης προς τον τρέχοντα χρήστη προς όλες τις προβλέψεις που συμμετείχε ο ενεργός χρήστης προς τον τρέχοντα χρήστη) ως βάρος για τον υπολογισμό της πρόβλεψης και διαπίστωσαν ότι έτσι αυξάνεται η ανθεκτικότητα του συστήματος μειώνοντας τον αντίκτυπο της επίθεσης στις παραγόμενες προβλέψεις σε σχέση με τη χρήση της παραδοσιακής συνεργατικής μεθόδου που χρησιμοποιεί ως βάρος την ομοιότητα των χρηστών

4.3 Επιθέσεις

Υπάρχουν διάφορες μορφές επίθεσης στα συστήματα υπόδειξης με πιο επιρρεπή τα συστήματα συνεργατικής υπόδειξης. Όμως αυτή η ευπάθεια εξισορροπείται από ένα μεγάλο πλεονέκτημα, τη δυνατότητα που δίδεται σε χρήστες με περίεργα και ιδιαίτερα ενδιαφέροντα να αναγνωρίσουν ανάμεσα στο σύνολο των χρηστών παρόμοια προφίλ και να λάβουν προτάσεις . Λόγω της ποικιλομορφίας των απόψεων ανάμεσα στους χρήστες είναι δύσκολο να κατηγοριοποιηθεί με σαφήνεια αν ένα προφίλ ανήκει σε εκκεντρικό ή κακόβουλο χρήστη, έτσι είναι ανέφικτο να υπάρξει για όλα τα διφορούμενα προφίλ ορθή διάκριση.

Πολλά συστήματα επιτρέπουν την δωρεάν πρόσβαση στο σύστημα μέσω απλής εγγραφής. Αυτή η διαδικασία μπορεί να αποτελέσει σημείο εκμετάλλευσης από τους επιτιθέμενους δημιουργώντας πολλαπλά προφίλ στο ίδιο σύστημα και εισάγοντας βαθμολογίες που είναι επιλεγμένες ώστε το να επηρεάσουν. Οι επιθέσεις εισαγωγής προφίλ (Profile injection attacks) προσθέτουν κάποιο αριθμό προφίλ (1%-5% των συνολικών προφίλ του

συστήματος) τα οποία θα πρέπει να ανιχνευτούν και να υπάρξει προστασία απέναντι στην παρουσία τους. Αυτού του είδους οι επιθέσεις αναφέρονται και ως shilling ή spam attacks. Οι επιθέσεις με εισαγωγή προφίλ μπορεί να κατηγοριοποιηθεί σε δύο ομάδες: την εισαγωγή κακόβουλων προφίλ που βαθμολογούν ένα αντικείμενο πολύ υψηλά με σκοπό την αύξηση της πιθανότητας το αντικείμενο αυτό να προταθεί για υπόδειξη σε μεγάλο τμήμα των χρηστών του συστήματος (push επιθέσεις) και η εισαγωγή κακόβουλων προφίλ που σκοπό έχουν τη μείωση τη δημοφιλία ενός αντικειμένου ώστε να μην προτάσσεται στους χρήστες (nuke επίθεση).

Ως επίθεση (attack) ορίζεται ως ένας μετασχηματισμός που αντιστοιχίζει τη βάση δεδομένων του συστήματος σε μία νέα βάση δεδομένων. Ο πιο δημοφιλής και κοινός τρόπος επίθεσης αποτελείται από τη δημιουργία μίας ομάδας μεροληπτικών προφίλ που εισάγονται στο σύστημα. Ο επιτιθέμενος εγγράφεται πολλαπλά στο σύστημα και λαμβάνει τη δυνατότητα χειρισμού πολλών λογαριασμών. Έτσι μπορεί να εισάγει μεροληπτικά δεδομένα στη βάση που αποτελούν θόρυβο ανάλογα με το σκοπό της επίθεσής του. Κάθε ένας από τους λογαριασμούς του κακόβουλου χρήστη που λαμβάνει μέρος στην επίθεση ονομάζεται προφίλ επίθεσης (attack profile). Το αντικείμενο ή τα αντικείμενα που τίθενται ως στόχοι είναι συνήθως μη δημοφιλή αντικείμενα , με χαμηλό μέσο όρο βαθμολογίας και με λίγες βαθμολογίες. Σε όλες τις περιπτώσεις θεωρείται ότι ο επιτιθέμενος δεν έχει πρόσβαση στη βάση δεδομένων πέραν τις αλληλεπίδρασής του με τη διεπαφή του συστήματος.

Το κόστος μίας επίθεσης εξαρτάται από το σύστημα και υπάρχουν διάφορα κριτήρια που μπορούν να χρησιμοποιηθούν για την κατασκευή της συνάρτησης κόστους όπως ο αριθμός των προφίλ που εισήχθησαν στο σύστημα. Συνήθως δεν είναι δυνατή η εισαγωγή περισσότερου του 1% των προφίλ σε πραγματικές συνθήκες. Αν το κόστος σχετίζεται με το μέγεθος της προσπάθειας που καταβάλλεται για το μετασχηματισμό της βάσης δεδομένων τότε μπορεί να μοντελοποιηθεί ως συνάρτηση αύξουσας μονοτονίας των ζευγαριών χρήστη-αντικειμένου που έχουν τροποποιηθεί από την επίθεση. Επίσης είναι δυνατόν να συμπεριληφθεί το πραγματικό κόστος αν οι βαθμολογίες μπορούν να εισαχθούν στο σύστημα μόνο με την απόκτηση του προϊόντος που έχει ως στόχο η επίθεση.

Από την οπτική του τελικού χρήστη μία αλλαγή στην προβλεπόμενη βαθμολογία ενός αντικειμένου από τη μεσολάβηση μίας επίθεσης δεν οδηγεί αναγκαία σε δυσαρέσκεια του

για την υπόδειξη ενώ συχνά δεν γίνεται καν αντιληπτή η διαφοροποίηση. Άλλωστε η ορθότητα των προβλέψεων είναι υποκειμενικό μέτρο και κατά συνέπεια μία επιτυχημένη επίθεση δεν συνεπάγεται διαφωνία στη γνώμη για το τι είναι καλό ή κακό μεταξύ των χρηστών και των επιτιθέμενων.

Για να αναλυθεί η ευπάθεια των αλγορίθμων του collaborative filtering πρέπει πρώτα να γίνει αντιληπτή η φύση των επιθέσεων που δέχονται. Αρχικός στόχος είναι επιθέσεις που έχουν ως κίνητρο οικονομικά οφέλη για αυτόν που οργανώνει την επίθεση. Τα οικονομικά οφέλη μπορούν να προκύψουν είτε αυξάνοντας τιε πωλήσεις δικών του προϊόντων (push στρατηγική) είναι μειώνοντας τις πωλήσεις των ανταγωνιστών (nuke στρατηγική). Κάθε προφίλ της επίθεσης θα επιλέξει υψηλή ή χαμηλή βαθμολογία ανάλογα τη στρατηγική για το αντικείμενο στόχο και θα βαθμολογήσει επίσης ένα σύνολο άλλων αντικειμένων ανάλογα με το μοντέλο επίθεσης που έχει επιλεχθεί.

Η κατανομή των βαθμολογιών των προφίλ επίθεσης επηρεάζουν την αποτελεσματικότητα της. Ως συνιστώσες της κατανομής ορίζονται ο βαθμός δημοφιλίας του αντικειμένου, ο βαθμός αρεσκείας του αντικειμένου στους χρήστες και η εντροπία των βαθμολογιών.

- Ο βαθμός αρεσκείας προκύπτει από το μέσο όρο των βαθμολογιών που έχει λάβει το αντικείμενο από τους χρήστες. Όσο υψηλότερη μέση βαθμολογία έχει ένα αντικείμενο τόσο συχνότερα προτείνεται στους χρήστες.
- Ο βαθμός δημοφιλίας προκύπτει από τον αριθμό των βαθμολογιών που έχει λάβει. Όταν ένα αντικείμενο δεν είναι δημοφιλές (δηλαδή δεν έχει λάβει πολλές βαθμολογίες) είναι πιο επιρρεπές σε επιθέσεις με σκοπό τον επηρεασμό των προβλέψεων και προτάσεων για το συγκεκριμένο .
- Ο βαθμός εντροπίας σχετίζεται με το πόσο διαφέρουν οι βαθμολογίες που έχει λάβει ένα αντικείμενο. Όσο υψηλότερη εντροπία έχει τόσο πιο ευάλωτο είναι σε επιθέσεις.

Για να γίνει αντιληπτό γιατί οι αλγόριθμοι που παράγουν τα ψεύτικα προφίλ είναι επιτυχημένοι πρέπει να μελετηθούν και να γίνουν κατανοητοί οι στόχοι τους. Πρωταρχικός στόχος είναι η μεγιστοποίηση της τιμής πρόβλεψης του αντικειμένου στόχου στους περισσότερους χρήστες του συστήματος. Αυτό μπορεί να επιτευχθεί με δύο τρόπους:

- Κατασκευή προφίλ με μέτριο βαθμό συσχέτισης με πολλούς χρήστες του συστήματος
- Κατασκευή προφίλ με πολύ υψηλό βαθμό συσχέτισης (ομοιότητα) με ένα συγκεκριμένο και μικρό κομμάτι του συνόλου των χρηστών.

Για να επιτευχθούν οι παραπάνω στόχοι τα προφίλ επίθεσης θα πρέπει να δημιουργηθούν με συγκεκριμένες προδιαγραφές. Οι περισσότερες τεχνικές επιθέσεων περιλαμβάνουν την βαθμολόγηση των αντικειμένων που τα αποτελούν γύρω από το μέσο όρο του συστήματος (κατανομή Gauss) που έχει σαν αποτέλεσμα τη μείωση της απόκλισης από τις βαθμολογίες των άλλων χρηστών , εκτός από το αντικείμενο που έχει επιλεχθεί ως στόχος της επίθεσης. Αρκετά μοντέλα επίθεσης έχουν μελετηθεί (average attack, Random attack, Bandagon attack, segment attack) και τα αποτελέσματα τους έδειξαν ότι μπορούν να δημιουργήσουν σημαντικά προβλήματα από τη μη αντιμετώπισή τους. Ακόμα και η εισαγωγή 1% κακόβουλων προφίλ μπορεί να επηρεάσει το σύστημα και να ανεβάσει πολύ το αντικείμενο που αποτελεί το στόχο της επίθεσης υψηλά στις προτεινόμενες υποδείξεις. Τέτοιες επιθέσεις είναι πολύ αποτελεσματικές ιδιαίτερα σε μη δημοφιλή αντικείμενα (αντικείμενα με μικρό αριθμό βαθμολογιών) όπου η μαζική βαθμολόγησή τους από τα προφίλ της επίθεσης θα επηρεάσει την επίδοσή τους.

Αν το σύστημα παράγει υποδείξεις μέσω του αλγορίθμου Resnick το κακόβουλο προφίλ που εισάγεται πρέπει να μοιάζει στους χρήστες στους οποίους απευθύνεται η επίθεση με σκοπό να επηρεάσει τις επιλογές τους. Για παράδειγμα έστω ότι ένα προμηθευτής θέλει να προωθήσει ένα προϊόν του. Το ιδανικό σενάριο θα ήταν τα προφίλ της επίθεσης να είχαν πλήρη ομοιότητα με τους χρήστες που αποτελούν το στόχο της επίθεσης κι έτσι να μεγιστοποιηθεί η βαθμολογία πρόβλεψης του αντικειμένου που θέλει να προωθήσει ο επιτιθέμενος. Οι παράμετροι της επίθεσης όπως αριθμός των κακόβουλων προφίλ που θα εισαχθούν στο σύστημα, ο αριθμός των αντικειμένων που θα βαθμολογηθούν και οι βαθμολογίες τροποποιούνται ανάλογα την επίθεση.

Τα προφίλ που συμμετέχουν στην επίθεση έχουν υψηλό RDMA καθώς ο αριθμητής (η διαφορά της βαθμολογίας από το μέσο όρο) θα είναι υψηλή ενώ ο παρονομαστής (το άθροισμα των βαθμολογιών που έχουν δοθεί στο συγκεκριμένο αντικείμενο) θα είναι χαμηλό. Έτσι συνολικά όλο το κλάσμα θα έχει υψηλή τιμή και οι επιτιθέμενοι ξεχωρίζουν

βάσει αυτής της τιμής και το σύστημα δεν περιλαμβάνει τα συγκεκριμένα προφίλ στη διαδικασία παραγωγής προτάσεων για τους υπόλοιπους χρήστες.

Ένα πολύ συνηθισμένο μέτρο για την ανίχνευση επίθεσης είναι το prediction shift που μετρά τη μέση διαφοροποίηση στην τιμή της πρόβλεψης εντός αντικειμένου πριν και μετά την επίθεση. Με τη χρήση του συγκεκριμένου μέτρου βρέθηκε ότι οι Average επιθέσεις είναι πολύ πιο αποτελεσματικές από τις Random. Και οι αλγόριθμοι k-NN που χρησιμοποιούν ως μέτρο ομοιότητας το Pearson είναι ιδιαίτερα ευάλωτοι σε επιθέσεις ψεύτικων προφίλ.

4.4 Θόρυβος vs κακόβουλη επίθεση

Η αποδοτική αναζήτηση πληροφορίας έχει ερευνηθεί και ερευνάται σε υψηλό βαθμό. Δεν ισχύει το ίδιο όμως για την ασφάλεια του συστήματος που προσφέρει την πληροφορία και αποτελεί εξίσου σημαντικό ζήτημα.

Όπως αναφέρθηκε η ασφάλεια είναι θεμελιώδους σημασίας για όλα τα συστήματα και τις εφαρμογές του Web. Ενώ κανείς θα πίστευε ότι λειτουργούν με ασφάλεια χτίζοντας απροσπέλαστους αλγορίθμους στη πραγματικότητα δεν είναι έτσι κυρίως λόγω της ανοικτής φύσης του τρόπους λειτουργίας τους. Καθώς δεν γίνεται να υποθέσεις τα κίνητρα χρήσης όλων των χρηστών , ούτε να εξασφαλιστεί η καλή πρόθεσή τους δεν μπορεί να εξασφαλιστεί ότι η το σύνολο των δεδομένων που εισάγονται σε ένα τέτοιο σύστημα αντανακλούν τις αληθινές επιλογές και σκέψεις των χρηστών. Έτσι στα δεδομένα θεωρείται ότι ένας βαθμός θορύβους υπάρχει φυσιολογικά και θα πρέπει να αναμένεται. Γενικά ο θόρυβος στα δεδομένα χωρίζεται σε δύο κατηγορίες:

- Φυσικός θόρυβος: Ο θόρυβος που συγκαταλέγεται σε αυτή την κατηγορία σχετίζεται με τους τρόπους λήψης των προτιμήσεων των χρηστών από τα συστήματα υπόδειξης. Σε πολλά συστήματα δηλαδή, ζητείται από τους χρήστες να βαθμολογήσουν αντικείμενα που αγόρασαν ή αξιολόγησαν. Καθώς η ανθρώπινη φύση είναι επιρρεπής στα λάθη και οι περισσότεροι χρήστες θεωρούν την βαθμολόγηση βαρετή κάποια λάθη στη βάση είναι φυσιολογικό να αναμένονται. Επίσης σε συστήματα που χρησιμοποιούν έμμεσα βαθμολογικά σχήματα υπάρχει η ίδια αν όχι και μεγαλύτερη πιθανότητα θορύβου στα δεδομένα. *Δηλαδή δημιουργείται από την ατελή συμπεριφορά*

του χρήστη (απρόσεκτη ή erroneous) και από τις διάφορες τεχνικές συλλογής δεδομένων.

- Κακόβουλος θόρυβος: Μεγαλύτερης σημασία και σοβαρότητα δίνεται στην πιθανότητα εισαγωγής μεροληπτικών βαθμολογικών στη βάση δεδομένων του συστήματος. Είναι ευνόητο ότι κακόβουλες οντότητες έχουν κίνητρα να δημιουργήσουν επιθέσεις εναντίον των συστημάτων υπόδειξης δεδομένων των πλεονεκτημάτων που μπορεί να επιτευχθούν με μία επιτυχημένη επίθεση. Πολλά συστήματα υπόδειξης δραστηριοποιούνται σε εμπορικό περιβάλλον ισχυρά κίνητρα δίνονται σε κακόβουλες οντότητες ώστε να προσπαθήσουν να ελέγξουν τον τρόπο λειτουργίας τους. Παραδείγματος χάριν ένας εκδοτικός οίκος επιθυμεί να προωθήσει ένα συγκεκριμένο βιβλίο. Θα προσπαθήσει να αυξήσει τις θετικές αξιολογήσει ς και βαθμολογίες για το εν λόγω βιβλίο ενώ παράλληλα θα καταβάλλει προσπάθεια να μειώσει τις θετικές αξιολογήσεις άλλων βιβλίων της ιδίας κατηγορίας ώστε να προτάσσεται περισσότερο το δικό του στους χρήστες του συστήματος . *Δηλαδή δημιουργείται από τη συνειδητή επιθυμία του χρήστη να εισάγει μεροληπτικά δεδομένα στο σύστημα με σκοπό να το διαβάλλει.*

4.4.1 Φυσικός θόρυβος

Η βάση δεδομένων στα συστήματα υπόδειξης μοντελοποιείται ως ένας πίνακας χρήστη-αντικειμένου που δείχνει τις προτιμήσεις ενός συνόλου χρηστώ A για ένα σύνολο αντικειμένων I . Έστω $r_{ij} \epsilon R$ η βαθμολογία που έδωσε ο χρήστης i στο αντικείμενο j όπου R είναι το σύνολο των βαθμολογιών (διακριτές οι τιμές του συνόλου). Θεωρούμε $U_g \subset U$ το σύνολο των αυθεντικών χρηστών του συστήματος και $U_a \subset U$ το σύνολο των προφίλ επίθεσης που υπάρχουν στο σύστημα.

Το θέμα που προκύπτει είναι ο διαχωρισμός αν ένα προφίλ περιλαμβάνει φυσικό ή κακόβουλο θόρυβο. Στη βιβλιογραφία έχουν αναφερθεί τεχνικές όπως το κατά πόσο κοντά βρίσκεται μία βαθμολογία r_{ij} στο μέσο όρο των βαθμολογιών του χρήστη i ή στο μέσο όρο των συνολικών βαθμολογιών που έχει λάβει το αντικείμενο j . Όμως σε συστήματα που στηρίζονται στις βαθμολογίες για να υπολογίσουν τις προτάσεις αυτές οι τεχνικές δεν

θα ήταν αποτελεσματικές καθώς η κατανομή των βαθμολογιών των χρηστών δεν είναι κατά κανόνα ομοιόμορφα κατανεμημένη γύρο από το μέσο όρο της κλίμακας βαθμολογίας του συστήματος.

Μία πιο αξιόπιστη προσέγγιση αποτελεί το Mean Absolute Error (MAE) *(Breese et al., 1998)* .Το συγκεκριμένο μέτρο χρησιμοποιείται για να μετρήσει τη συνέπεια c (consistency) της βαθμολογίας $r_{u,v}$ που έδωσε ο χρήστης u για το αντικείμενο v και της προβλεπόμενης βαθμολογίας $p_{u,v}$ βάσει της παρακάτω εξίσωσης:

$$[c(G,T)]_{\perp}(u,v) = [(\ [|r]_{\perp}(u,v) - p_{\perp}(u,v)\ |)/(r\ [max -]_{\perp} r\ [min]_{\perp})]$$

Εξίσωση 16

Όπου $p_{u,v}$ είναι η προβλεπόμενη τιμή για το αντικείμενο v από το χρήστη u και $r\frac{max}{}\ r\,min$ η μεγαλύτερη και η μικρότερη επιτρεπόμενη τιμή της κλίμακας βαθμολογίας του συστήματος. Η βαθμολογία $r_{u,v}$ θεωρείται θόρυβος αν ισχύει η παρακάτω σχέση :

$$c(G,T)_{u,v} > th$$

Εξίσωση 17

όπου th είναι η τιμή κατωφλίου.

4.4.2 Κακόβουλος θόρυβος

Ο κακόβουλος θόρυβος προκαλείται από την εισαγωγή προφίλ επίθεσης στο σύστημα με στόχο να επηρεάσουν τις προβλέψεις με παράγονται με τον τρόπο που επιθυμεί ο επιτιθέμενος. Στόχος και σε αυτή την περίπτωση είναι ο διαχωρισμός γνησίων και κακόβουλων προφίλ. Και σε αυτή την περίπτωση η θεωρία (**signal detection theory)** μπορεί να χρησιμοποιηθεί. Αν δεν υπάρχουν επικαλύψεις μεταξύ των κατανομών των κακόβουλων και των γνήσιων χρηστών η επιλογή ενός κατωφλιού όπου για ανώτερες τιμές ο χρήστης χαρακτηρίζεται ως κακόβουλος είναι εύκολη διαδικασία. Όμως είναι φυσιολογικό να υπάρχουν επικαλύψεις των κατανομών και τότε η επιλογή του κατωφλιού του χωρίζει γνήσιους και κακόβουλους χρήστες είναι προβληματική . Σε αυτή την περίπτωση μαζί με τις επιτυχίες «hits» δηλαδή τις σωστές αποφάνσεις για το ότι ένας

χρήστης είναι κακόβουλος συχνά έχουμε και λανθασμένους συναγερμούς (false alarm) που σημαίνει ότι αυθεντικοί χρήστες έχουν λανθασμένα ταυτοποιηθεί ως κακόβουλοι. Για αυτό χρειάζεται σωστή επιλογή κατωφλιού που να οδηγεί σε υψηλό ποσοστό επιτυχιών «hits» και ταυτόχρονα χαμηλό ποσοστό λανθασμένων συναγερμών «false alarms». Η επιλογή πολύ υψηλής κατωφλιού οδηγεί σε μείωση της ορθότητας και της κάλυψης οπότε θα πρέπει να κρατηθεί μία ισορροπία να επιτύχουμε τη μεγιστοποίηση όλων των διαστάσεων που μας ενδιαφέρουν.

Αν αφαιρεθεί ο φυσικός θόρυβος από τα δεδομένα τότε το σύστημα θα γίνει πιο ακριβές (**accurate**) στις προβλέψεις του ενώ αν αντιμετωπιστούν αποτελεσματικά οι επιθέσεις των κακόβουλων θα αυξηθεί η ευρωστία (**robustness**) του συστήματος.

4.5 Διαστάσεις επίθεσης

Κάθε επίθεση έχει έναν αριθμό ιδιοτήτων που μπορεί να φανούν χρήσιμες για την περιγραφή και τη σύγκριση διαφορετικών επιθέσεων.

Σκοπός επίθεσης

Διαφορετικές επιθέσεις πιθανόν να έχουν διαφορετικό σκοπό καθώς αυτός απορρέει από τις προθέσεις του οργανωτή της επίθεσης. Αν και ο γενικός σκοπός των επιθέσεων είναι να επηρεαστούν με κάποιο τρόπο οι προβλέψεις αντικειμένων που γίνονται για τους χρήστες. Δύο βασικοί σκοποί είναι η προώθηση (**push**) δηλαδή η πρόταση των επιλεγμένων προϊόντων σε όσο το δυνατόν περισσότερους χρήστες και αποδυνάμωση (**nuke)** των στοχευόμενων προϊόντων με σκοπό να προτάσσονται όσο το δυνατόν αραιότερα. Επίσης μπορεί να μην είναι σκοπός ένα αντικείμενο αλλά όλο το σύστημα. Δηλαδή επιθυμία του επιτιθέμενου να είναι να επηρεάσει τη συνολική λειτουργία του συστήματος ώστε να υπολειτουργεί ή να παράγει προτάσεις χαμηλής ποιότητας.

Στόχος επίθεσης

Οι επιθέσεις μπορεί να απευθύνονται σε ένα υποσύνολο χρηστών ή σε ένα υποσύνολο αντικειμένων του συστήματος υπόδειξης. Καλό για τον επιτιθέμενο είναι να περιορίσει το αποτελέσματα της επίθεσης σε ένα μικρό και συγκεκριμένο υποσύνολο αντικειμένων ώστε να είναι πιο στοχευόμενη η επίθεση και ταυτόχρονα να μην κινήσει υποψίες στους διαχειριστές του συστήματος. Επιπρόσθετα θα ήταν επικερδές να περιορίσει την επίδραση και σε μικρό υποσύνολο χρηστών που όμως έχουν τα επιθυμητά χαρακτηριστικά ώστε να μεγιστοποιηθεί ο αντίκτυπος της επίθεσης. Παραδείγματος χάριν αφενός θα φαινόταν

ύποπτο να υπάρξει πρόταση του συστήματος για ένα cd με heavy metal μουσική σε ένα χρήστη που έχει αξιολογήσει θετικά μόνο κομμάτια κλασικής μουσικής. Αφετέρου δεν είναι το ίδιο αποτελεσματική η συγκεκριμένη επίθεση σε σχέση με τα αποτελέσματα που θα είχε αν το σύνολο των χρηστών που έλαβαν τη συγκεκριμένη πρόταση ήταν όλοι οπαδοί της rock.

Απαιτούμενη γνώση

Οι επιθέσεις ενδέχεται να απαιτούν κάποιο επίπεδο γνώσης για τα αντικείμενα , τους χρήστες, τις βαθμολογίες και τους αλγορίθμους του συστήματος υπόδειξης για να είναι επιτυχημένες. Μία επίθεση που έχει σχεδιαστεί έχοντας γνώση για τα δεδομένα είναι πιο αποτελεσματική από μία επίθεση που γίνεται χωρίς να έχει εσωτερικές πληροφορίες. Επιπλέον πληροφορίες όπως ο βαθμός αραιότητας των βαθμολογιών, οι κατανομές των βαθμολογιών, οι παράμετροι των χρησιμοποιούμενων αλγορίθμων μπορούν να βοηθήσουν σημαντικά στην επιλογή των παραμέτρων της ίδιας της επίθεσης ώστε αυτή να γίνει πιο αποτελεσματική και να ελαχιστοποιήσει τη δυνατότητα ανίχνευσής της.

Κόστος

Μία επίθεση έχει ένα σχετιζόμενο κόστος που εξαρτάται από την προσπάθεια που χρειάστηκε και το επίπεδο πληροφοριών που χρειάστηκε για την εκτέλεσή της. Έχοντας το κόστος της επίθεσης και ένα κατάλληλο μέτρο αξιολόγησης της αποτελεσματικότητας της επίθεσης είναι δυνατή η αξιολόγηση της με βάσει το δείκτη κόστος/κέρδος ώστε να φανεί αν η επίθεση άξιζε να πραγματοποιηθεί. Διαστάσεις που προσμετρούνται στη δημιουργία της συνάρτησης κόστους είναι οι παρακάτω:

- Μέγεθος επίθεσης (ο αριθμός των κακόβουλων προφίλ που εισήχθησαν στο σύστημα)
- Δυσκολία αλληλεπίδρασης με το σύστημα. Παραδείγματος χάριν συστήματα που περιλαμβάνουν ημι-αυτόματες τεχνικές όπως η CAPTCHA πιθανόν να έχουν μεγαλύτερο κόστος από επίθεση σε κάποιο σύστημα που δεν έχει.
- Η απόκτηση πληροφοριών για το σύστημα, τους αλγορίθμους, τις βαθμολογίες και τους χρήστες.
- Οποιοσδήποτε άλλος πόρος χρειάζεται για το σχεδιασμό και την εκτέλεση της επίθεσης

Αλγοριθμική εξάρτηση

Κάποιες επιθέσεις μπορεί να έχουν σχεδιαστεί με σκοπό την εκμετάλλευση συγκεκριμένων αδυναμιών των αλγορίθμων του υπό επίθεση συστήματος ενώ κάποιες άλλες είναι πιο γενικής μορφής και μπορούν να εκτελεστούν σε διαφορετικά συστήματα έχοντας αποτελέσματα έναντι διαφόρων αλγορίθμων. Οι επιθέσεις που έχουν σχεδιαστεί έναντι συγκεκριμένων αλγορίθμων μπορεί να χρειάζονται λιγότερους πόρους αλλά από την άλλη απαιτούν περισσότερες και λεπτομερείς πληροφορίες για τον τρόπο λειτουργίας των αλγορίθμων του συστήματος και των ρυθμίσεών του.

Δυνατότητα ανίχνευσης.

Ιδιότητες που προκύπτουν από την επίθεση μπορεί να ευθύνονται για την ανίχνευσή τους σε χρήστες και διαχειριστές του συστήματος. Γενικά ο επιτιθέμενος επιθυμεί να οργανώσει και να εκτελέσει την επίθεσή του με τέτοιο τρόπο ώστε να είναι όσο το δυνατό μικρότερη η δυνατότητα ανίχνευσης της από τους διαχειριστές. Έτσι θα μπορεί για μεγαλύτερο χρονικό διάστημα να επηρεάζει κατά το δοκούν τις προβλέψεις και τις υποδείξεις στους χρήστες προτού γίνει αντιληπτή και απενεργοποιηθεί η δραστηριότητά της. Η σημασία της δυνατότητας ανίχνευσης εξαρτάται από το σκοπό της επίθεσης.

Μοντέλο επίθεσης

Το μοντέλο επίθεσης εξειδικεύει τα χαρακτηριστικά των βαθμολογιών του προφίλ επίθεσης.

Μέγεθος προφίλ

Ο αριθμός των βαθμολογημένων αντικειμένων ενός προφίλ αποτελεί το μέγεθος προφίλ. Η προσθήκη βαθμολογιών σε ένα υπάρχον προφίλ είναι χαμηλότερου κόστους σε σχέση με την εισαγωγή ενός νέου προφίλ στο σύστημα. Όμως τα προφίλ που έχουν μεγάλο μέγεθος έχουν ρίσκο να θεωρηθούν κακόβουλα καθώς τις περισσότερες φορές οι γνήσιοι χρήστες δεν βαθμολογούν παρά ένα μικρό τμήμα του συνόλου των αντικειμένων της βάσης. Κανείς δεν μπορεί να έχει διαβάσει όλα τα βιβλία που έχουν εκδοθεί ούτε μπορεί να έχει δει όλες τις ταινίες που έχουν κυκλοφορήσει. Έτσι μεγάλα προφίλ επίθεσης με πολλά βαθμολογημένε αντικείμενα είναι εύκολο να ξεχωρίσουν στη βάση και να αποδοθούν ως κακόβουλα.

Μέγεθος επίθεσης

Ως μέγεθος επίθεσης ορίζεται ο αριθμός των προφίλ που έχουν εισαχθεί στο σύστημα στα πλαίσια μιας κακόβουλης ενέργειας. Θεωρείται ότι ένας εκλεπτυσμένος επιτιθέμενος θα πραγματοποιεί την εισαγωγή των προφίλ αυτόματα. Για αυτό ο αριθμός των προφίλ είναι κρίσιμης σημασίας γιατί είναι εφικτή η αποτροπή της επίθεσης εισάγοντας στην διαδικασία εγγραφής βήματα που απαιτούν ανθρώπινη παρέμβαση αυξάνοντας έτσι ο διαχειριστής το κόστος εγγραφής στο σύστημα

4.6 Στρατηγικές επίθεσης

Βάσει της βιβλιογραφίας δύο είναι οι βασικές στρατηγικές κακόβουλης επίθεσης. Οι product push και product nuke επιθέσεις. Ο αντικειμενικός στόχος αυτών των επιθέσεων είναι να προωθήσουν ή να υποβιβάσουν τις προβλέψεις που γίνονται για συγκεκριμένα αντικείμενα.

Οι στόχοι των επιθέσεων ώθησης (push) και (nuke) είναι η προώθηση και η denote αντίστοιχα των αντικειμένων που αποτελούν το στόχο της επίθεσης. Οι επιτιθέμενοι πρέπει να λαμβάνουν υπόψιν τους δύο βασικές παραμέτρους πρώτον η επιλογή των αντικειμένων που θα βαθμολογηθούν από τα προφίλ επίθεσης και δεύτερον ποια θα είναι η βαθμολογία που θα αποδοθούν στα επιλεχθέντα αντικείμενα. Προδιαγραφές ενός προφίλ επίθεσης: πρώτον θα πρέπει να υπάρχει υψηλός βαθμός ομοιότητας με τους αυθεντικούς χρήστες ώστε να μπορεί να επηρεάζει τις παραγόμενες προβλέψεις από το σύστημα προς τους χρήστες που αποτελούν το στόχο της επίθεσης. Δεύτερον επειδή το αποτέλεσμα της συσχέτισης Pearson ανήκει στο κλειστό και συνεχές διάστημα [-1,1] είναι σημαντικό κάθε προφίλ επίθεσης να σχετίζεται αρνητικά ή θετικά με τα προφίλ των γνήσιων χρηστών. Σε διαφορετική περίπτωση δεν θα υπάρξει αντίκτυπος της επίθεσης ή ακόμα χειρότερα για τον επιτιθέμενο αντί για προώθηση του προϊόντος θα έχει nuke και το αντίθετο.

Ας δώσουμε ένα παράδειγμα δημοφιλούς επίθεσης . Επιλέγοντας δημοφιλή αντικείμενα που είτε αρέσουν είτε δεν αρέσουν στον πληθυσμό των χρηστών και στη συνέχεια βαθμολογώντας τα ανάλογα μπορούν να επιτευχθούν οι προδιαγραφές που περιγράφηκαν προηγουμένως. Έστω μία επίθεση προώθησης , σκοπός είναι η ανάθεση της χαμηλότερης βαθμολογίας στα αντικείμενα που δεν αρέσουν r_{min} , στα αντικείμενα που αρέσουν r_{min} +1 ενώ στο αντικείμενο που αποτελεί το στόχο της επίθεσης βαθμολογία r_{max} . Έτσι δημιουργείται θετική συσχέτιση μεταξύ του κακόβουλου προφίλ και των αυθεντικών.

Αντίστοιχες θα είναι και οι ενέργειες σε nuke επίθεση με αντιστροφή των βαθμολογιών μεταξύ των αντικειμένων που αρέσουν και δεν αρέσουν.

Η δυσκολία της παραπάνω στρατηγικής έγκειται στην αναγνώριση των αντικειμένων που είναι δημοφιλή με θετικό ή αρνητικό τρόπο. (αρέσουν /δεν αρέσουν). Σε κάποιες περιπτώσεις κάτι τέτοιο είναι εύκολο όπως για παράδειγμα στο σύστημα MovieLens που προβάλει το μέσο όρο των συνολικών βαθμολογιών που έχουν δοθεί σε κάθε αντικείμενο του συστήματος από τους χρήστες. Επίσης είναι δυνατή η εξόρυξη χρήσιμων πληροφοριών και από άλλες πηγές όπως ο αριθμός χρηστών που έχουν αξιολογήσει θετικά ή αρνητικά το αντικείμενο κά.

Random product push/nuke attack

Οι επιθέσεις push και nuke επιχειρούν να ωθήσουν την πρόβλεψη της βαθμολογίας του αντικειμένου επίθεσης ή της ομάδας αντικειμένων επίθεσης σε μία συγκεκριμένη τιμή στόχος. Ένα παράδειγμα τέτοιας επίθεσης θα μπορούσε να είναι ένας συγγραφέας που θέλει να προωθήσει τα βιβλία του εισάγοντας υψηλές βαθμολογίες στο σύστημα υπόδειξης. Σε αυτή την περίπτωση ο επιτιθέμενος επιχειρεί την επίθεση χτίζοντας ψεύτικα προφίλ με το αντικείμενο στόχο να λαμβάνει την υψηλότερη τιμή βάσει του συστήματος βαθμολόγησης. Ο αριθμός των αντικειμένων, τα αντικείμενα και οι βαθμολογίες τους επιλέγονται τυχαία. Και αυτή η επίθεση είναι χαμηλής αποτελεσματικότητας.

Focussed product push/nuke attack

Σε αυτού του τύπου την επίθεση στόχος είναι η εστίαση σε συγκεκριμένα αντικείμενα αλλά σε αυτή την περίπτωση ο σκοπός επιτυγχάνεται μέσω της εκμετάλλευσης της σχέσης συσχέτισης Pearson. Η εκμετάλλευση στηρίζεται στο ότι η συσχέτιση των χρηστών προκύπτει μόνο μέσω της σύγκρισης των βαθμολογιών αντικειμένων που και οι δύο έχουν αξιολογήσει. Συνεπώς μπορεί να προκύψει υψηλή συσχέτιση χρηστών με λίγα κοινά αντικείμενα. Αυτή η αδυναμία σχετίζεται τόσο με τη μείωση της ακρίβειας των προβλέψεων αλλά και της ευρωστίας του συστήματος.

Ο βαθμός συσχέτισης που προκύπτει από τη χρήστη του Pearson δύο χρηστών με δύο μόνο κοινά αντικείμενα είναι πάντα -1 ή 1, δηλαδή πλήρως θετική ή αρνητική. Έτσι να ένας επιτιθέμενος επιλέξει δύο αντικείμενα για τα οποία μπορεί να γνωρίζει τις βαθμολογίες των

χρηστών μαζί με το αντικείμενο που θέλει να προωθήσει τότε είναι υψηλή η πιθανότητα να επιτύχει η επίθεση.

Ζητούμενο είναι τα προφίλ επίθεσης συσχετίζονται με τον ίδιο τρόπο θετικά ή αρνητικά με τους γνήσιους χρήστες που επιτελούν το στόχο της επίθεσης. Αυτό προϋποθέτει ότι ο επιτιθέμενος έχει γνώσει των βαθμολογιών που έχει βάλει ο χρήστης σε συγκεκριμένα αντικείμενα. Υπάρχουν κάποιες μέθοδοι πχ οι ευρετικές που βοηθούν σε αυτό το πρόβλημα. Είναι λογική η υπόθεση ότι κάποια αντικείμενα που έχουν βαθμολογηθεί πολλές φορές (πχ τα δημοφιλή αντικείμενα) γενικά αξιολογούνται υψηλότερα του μέσου όρου. Τέτοια αντικείμενα για αυτό το λόγο είναι κατάλληλες επιλογές για την κατασκευή επιθέσεων. Και αυτή η επίθεση είναι χαμηλής αποτελεσματικότητας αλλά με καλύτερα αποτελέσματα από τα προηγούμενα.

Ευρωστία συστήματος

Η ευρωστία (robustness) *(Groot et all 2000)* αποτελεί το βαθμό κατά τον οποίο ένα σύστημα ή ένα υποσύστημα λειτουργεί ορθά παρουσία θορύβου στα δεδομένα ή κάτω από πιεστικές περιβαλλοντολογικές συνθήκες.

Οι συνεργατικές μέθοδοι στηρίζονται στην ανανέωση των δεδομένων όποτε οι χρήστες εισάγουν μία νέα βαθμολογία χωρίς να υπάρχει ασφάλεια ότι τα δεδομένα που εισάγονται αντιπροσωπεύουν τη πραγματική γνώμη του χρήστη. Επειδή η διαδικασία της βαθμολόγησης μπορεί να είναι επαχθής , οι χρήστες μπορεί να είναι απρόσεκτοι στις τιμές που εισάγουν και για αυτό ανακρίβειες κατά συνέπεια θόρυβος στα δεδομένα πρέπει να αναμένεται. Επίσης μπορεί να υπάρξει περίπτωση που κακόβουλοι χρήστες επιτίθενται επίτηδες σε συστήματα υπόδειξης και τους προκαλούν προβλήματα λειτουργίας. Ανάλογα με την εφαρμογή, το κόστος από μία επίθεση ποικίλει για τους διαχειριστές και του πελάτες του συστήματος.

Οι περισσότερες εφαρμογές υπόδειξης λειτουργούν στο διαδίκτυο όπου είναι απίθανο να ελεγχθεί αν τα δεδομένα που εισέρχονται σε αυτές είναι ειλικρινή. Έτσι εξαρτάται από τον αλγόριθμο για το κατά πόσο αντιμετωπίζει το πρόβλημα του θορύβου των δεδομένων.
Αναφέρονται δυο οπτικές της ευρωστίας . Η πρώτη αφορά την ορθότητα (accuracy) της υπόδειξης, δηλαδή οι προτάσεις που παράχθηκαν μετά την επίθεση αρέσουν στους χρήστες; Ενώ η δεύτερη οπτική της ευρωστίας είναι η σταθερότητα (stability) η οποία

εξετάζει αν υπήρξαν διαφοροποιήσεις στις προβλέψεις που παράχθηκαν πριν και μετά την επίθεση ανεξαρτήτως αν οι προτάσεις άρεσαν ή όχι στους χρήστες.

Ενώ σταθερότητα και ορθότητα είναι διακριτές έννοιες δεν είναι πλήρως ανεξάρτητες. Για παράδειγμα αν ένας σύστημα παραγωγής υποδείξεων έχει άριστα επίπεδα ορθότητας ανεξαρτήτως παρουσίας θορύβου στα δεδομένα τότε θα πρέπει να έχει και άριστα επίπεδα σταθερότητας. Από την άλλη έστω ένα αντικείμενο που δεν αρέσει σε κανέναν. Σύμφωνα με τους κανόνες των συστημάτων υπόδειξης η διαδικασία του να προτείνεις σε όλους ή να μην προτείνεις σε κανέναν είναι αμφότερα σταθερές αλλά ενώ η πρώτη είναι ορθή ενώ η δεύτερη είναι λανθασμένη.

Η μέτρηση της ευρωστίας ενός συστήματος μπορεί να πραγματοποιηθεί με δύο τρόπους. Είτε με τη μέτρηση της ορθότητας των υποδείξεων που προτάσσονται στους χρήστες του συστήματος -δηλαδή απαντώντας την ερώτηση αν στο χρήστη Χ άρεσε το αντικείμενο Ψ που του προτάθηκε-, είτε μετρώντας την σταθερότητα του συστήματος. Ακόμα κι αν δεν είναι δυνατή η αξιολόγηση του αν το αντικείμενο Ψ άρεσε στο χρήστη Χ το γεγονός ότι το σύστημα άλλαξε τις προτάσεις του προς το χρήστη Χ μετά την επίθεση ή λόγω θορύβου αρκεί να προκαλέσει ανησυχία στους διαχειριστές ακόμα κι αν η υπόδειξη πριν την επίθεση ήταν λανθασμένη.

Ένα εύρωστο σύστημα υπόδειξης θα πρέπει να έχει τα παρακάτω χαρακτηριστικά:

1. Υψηλή ευστάθεια σε μικρές επιθέσεις (επιθέσεις που αποτελούνται από μικρό αριθμό προφίλ που εισάγονται στο σύστημα)
2. Καλή έως μέτρια ευστάθεια σε επιθέσεις μεσαίου μεγέθους δηλαδή αριθμός προφίλ επίθεσης μικρότερος του 5% των χρηστών του συστήματος.
3. Χαμηλή μέση επιρροή στην προβλεπόμενη ορθότητα (Mean Average Error) σε αντικείμενα που δεν είναι αντικείμενα στόχος της επίθεσης.
4. Πολύ υψηλή σταθερότητα στην εισαγωγή τυχαίου θορύβου στα δεδομένα
5. Όχι μείωση ορθότητας σε περίπτωση που δεν υπάρχει επίθεση στο σύστημα
6. Δυνατότητα δημιουργίας εμπιστοσύνης έστω και μερικώς μεταξύ των χρηστών
7. Επεκτασιμότητα ώστε να είναι δυνατή η υποστήριξη εκατοντάδων χιλιάδων γνήσιων χρηστών και μερικών χιλιάδων κακόβουλων

8. Δυνατότητα αντιμετώπισης παράλληλων επιθέσεων που έχουν στόχο διαφορετικά αντικείμενα
9. Δυνατότητα γενίκευσης της γνώσης που λήφθηκε από την εκπαίδευση του μοντέλου για την αντιμετώπιση διαφορετικών τύπων επίθεσης
10. Προσπάθεια για όσο το δυνατόν μεγαλύτερη ανεξαρτησία από παραμέτρους, ο αλγόριθμος θα πρέπει να επιλέγει το κατώφλι σε κάθε επίθεσης και να μην απαιτείται ανθρώπινη χειροκίνητη παρέμβαση.

4.7 Προφίλ επίθεσης

Γενική μορφή προφίλ επίθεσης

Σκοπός του επιτιθέμενου είναι να δημιουργήσει μία επίθεση που να χρειάζεται ελάχιστες πληροφορίες και γνώση για το σύστημα ενώ ταυτόχρονα η αποτελεσματικότητά της θα είναι μέγιστη.

Η γενική μορφή μίας επίθεσης προώθησης (**push)** είναι η εξής:

Αποτελείται από ένα σύνολο μ βαθμολογιών που αντιστοιχούν σε μ αντικίμενο όπου μ είναι το σύνολο των αντικειμένων του συστήματος. Το προφίλ επίθεσης των μ βαθμολογιών μπορεί να σπάσει σε τέσσερες ομάδες αντικειμένων: α) το αντικείμενο στόχο i_t (target-item),μία ομάδα επιλεγμένων αντικειμένων I_S (selected items), μία ομάδα αντικειμένων πλήρωσης που συνήθως επιλέγονται τυχαία I_F (filler items) και μία ομάδα αντικειμένων που δεν έχουν ακόμα αξιολογηθεί από το χρήστη I_E (unrated items) . Κατά την επίθεση το σύνολο των επιλεγμένων αντικειμένων (selected items) παραμένουν ίδια για όλα τα προφίλ επίθεσης. Για παράδειγμα έστω επίθεση που εισάγει στο σύστημα δέκα προφίλ. Αυτά τα προφίλ θα έχουν ίδιο αντικείμενο επίθεσης και επιλεγμένα αντικείμενα ενώ θα διαφέρουν στα τυχαία αντικείμενα καθώς η επιλογή τους είναι τυχαία. Τα μοντέλα επίθεσης ορίζονται βάσει των κανόνων που ακολουθούνται για την επιλογή των αντικειμένων των τεσσάρων ομάδων που προαναφέρθηκαν και των βαθμολογιών τους. Σε κάποιες επιθέσεις το σύνολο των επιλεγμένων αντικειμένων () μπορεί να είναι κενό.

Selected Items I_S	Filler Items I_F	Unrated Items I_E	Target Item i_t

Πίνακας 3 Γενική μορφή προφίλ επίθεσης

Οι περισσότεροι ερευνητές του παρελθόντος εστίαζαν στο σύνολο των επιλεγμένων αντικειμένων για τη δημιουργία ενός νέου μοντέλου επίθεσης θα πρέπει να επισημανθεί και η σημασία των αντικειμένων πλήρωσης (Filler Items)

4.7.1 Παράδειγμα επίθεσης

Όπως αναφέρθηκε ξανά, τα συστήματα συνεργατικής υπόδειξης που στηρίζονται στην ομοιότητα των χρηστών, η γειτονιά κάθε χρήστη προκύπτει από την ομοιότητα του τρέχοντος με τους υπολοίπους. Όσοι χρήστες έχουν υψηλή συσχέτιση με τον τρέχοντα δημιουργούν τη γειτονιά του και αποτελούν πηγή των υποδείξεων του. Έστω ένα σύστημα υπόδειξης τραγουδιών και μία χρήστης η Alice. Έστω ότι το σύστημα θέλει να προβλέψει τη βαθμολογία της για το αντικείμενο 5 (το αντικείμενο αυτό προφανώς δεν έχει ακόμα βαθμολογηθεί από εκείνη). Στον πίνακα 2 παρατίθεται το προφίλ της Alice και πέντε άλλων αυθεντικών χρηστών. Βάσει των ομοιοτήτων προκύπτει ότι ο χρήστης 4 είναι πιο κοντά στην Alice αλλά δεν έχει δει το αντικείμενο 5 και για αυτό δεν μπορεί να χρησιμοποιηθεί για πρόβλεψη. Οπότε ως δεύτερο στη σειρά ομοιότητας είναι ο χρήστης 4 και η πρόβλεψη για την Alice είναι 2.

	Αντικείμενο 1	Αντικείμενο 2	Αντικείμενο 3	Αντικείμενο 4	Αντικείμενο 5	Βαθμός συσχέτισης με Alice
Alice	4	4	1	3	;	
Χρήστης 1	2	3		2	1	0.5
Χρήστης 2	4	1	2		2	0.1889
Χρήστης 3		3	4	4	4	-0.7559
Χρήστης 4	4		1	3		1
Χρήστης 5	3	4	3	4	2	0.4080

Πίνακας 4 Προφίλ χρηστών συστήματος πριν την επίθεση

Έστω τώρα ότι στο σύστημα η κακόβουλη χρήστης Eve εισάγει δύο ψεύτικα προφίλ με σκοπό να προωθήσει το αντικείμενο 5. Στον πίνακα 3 φαίνονται οι νέες συσχετίσεις της Alice με τους χρήστες του συστήματος (γνήσιους και κακόβουλους). Πλέον ο χρήστης με την υψηλότερη ομοιότητα με την Alice που να έχει αξιολογήσει το αντικείμενο 5 είναι η Eve1 δημιουργώντας πρόβλεψη για 5 ενώ πριν την επίθεση το η πρόβλεψη ήταν 2, ακριβώς το αντίθετο. Το παράδειγμα με απλό τρόπο έδειξε πόσο μπορεί να αλλοιώσει μία επίθεση την ικανότητα του συστήματος να παράγει ορθές προβλέψεις το οποίο προκαλεί σε δυσαρέσκεια από την πλευρά του χρήστη και που προοδευτικά οδηγεί στη μείωση της εμπιστοσύνης του. Αντίστοιχα αποτελέσματα μπορούν να προκύψουν και σε συστήματα υπόδειξης που κατασκευάζουν προβλέψεις βασιζόμενα στην ομοιότητα αντικειμένων.

	Αντικείμενο 1	Αντικείμενο 2	Αντικείμενο 3	Αντικείμενο 4	Αντικείμενο 5	Βαθμός συσχέτισης με Alice
Alice	4	4	1	3	;	
Χρήστης 1	2	3		2	1	0.5
Χρήστης 2	4	1	2		2	0.1889
Χρήστης 3		3	4	4	4	-0.7559
Χρήστης 4	4		1	3		1
Χρήστης 5	3	4	3	4	2	0.4080
Eve1	4	4	3	4	5	0.9428
Eve2	4	3	2	4	5	0.7385

Πίνακας 5 Προφίλ χρηστών συστήματος μετά την επίθεση

Επιλογή αντικειμένου στόχου

Τα περισσότερα μοντέλα επίθεσης δεν θεωρούν σημαντική την επιλογή του σωστού αντικειμένου για να το θέσουν ως αντικείμενο επίθεσης. Αυτό οφείλεται στο γεγονός ότι τα μοντέλα λαμβάνουν υπόψιν τους μόνο τις βαθμολογίες των χρηστών. Σύμφωνα με τους *Mobasher et al., (2005)* δεν είναι κάθε αντικείμενο της βάσης καλός υποψήφιος για να

αποτελέσει το στόχο της επίθεσης. Καλός υποψήφιος ορίζεται το αντικείμενο που έχει υψηλές πιθανότητες θα επιλεχθεί από το χρήστη όταν αυτό του προταθεί. Οι παράγοντες που συντελούν στη δημιουργία ενός καλού υποψηφίου μπορούν να εξαχθούν από τη μελέτη του τρόπου επιλογής αντικειμένων από το χρήστη (παραδείγματος χάριν μίας ταινίας, ενός εστιατορίου κτλ). Τέτοιοι παράγοντες που επηρεάζουν τη συμπεριφορά του χρήστη είναι οι βαθμολογίες και οι αξιολογήσεις για ένα αντικείμενο, η άποψη των experts και δημοφιλών χρηστών, Οι αξιολογήσεις και οι βαθμολογίες δεν είναι απαραίτητο ότι πρέπει να βρίσκονται στο συγκεκριμένο σύστημα αλλά και σε άλλα. Πχ όταν ένας χρήστης θελήσει να αγοράσει μία ταινία από το Amazon, πέραν των αξιολογήσεων σε αυτό θα κοιτάξει και σε άλλα sites όπως το IMDB.

Τέτοιοι παράγοντες πρέπει να λαμβάνονται υπόψιν γιατί ένα αντικείμενο με χαμηλή μέση βαθμολογία και αξιολογήσεις αρνητικές από χρήστες και ειδικούς δεν αποτελεί καλό υποψήφιο στόχο για επίθεση. Μπορεί μετά την επίθεση το συγκεκριμένο αντικείμενο να μπει στη λίστα προτάσεων των χρηστών αλλά παρόλα αυτά οι πιθανότητες να το επιλέξει ο χρήστης είναι χαμηλές.

Αντικείμενα που είναι καλύτεροι υποψήφιοι πρέπει να αρέσουν ή έστω να είναι υπό αμφισβήτηση το αν είναι θετικές οι αρνητικές οι κριτικές. Αν ένα αντικείμενο αρέσει είναι περισσότερες οι πιθανότητες να το επιλέξει ο χρήστης ενώ αν το αντικείμενο δεν είναι ούτε αρεστό ούτε μη αρεστό πάλι υπάρχουν πιθανότητες μετά την επίθεση να μετατραπεί η άποψη για αυτό σε θετική. Η υπόθεση που βασίζεται ο παραπάνω συλλογισμός είναι ότι ένας χρήστης είναι πιθανότερο να αγοράσει ένα δημοφιλές αντικείμενο που από τη θέση 8 της λίστας προτάσεων ανέβηκε μετά την επίθεση στη θέση 2.

Επιλογή αντικειμένων πλήρωσης

Πολλά μοντέλα επίθεσης στοχεύουν στη δημιουργία προφίλ που να έχουν τη μέγιστη δυνατή ομοιότητα με τους γνήσιους χρήστες ενός συστήματος (πχ average, Random, Bandagon κά). Όταν ένα προφίλ έχει μεγάλη ομοιότητα με τον τρέχοντα χρήστη επιλέγεται για τη γειτονιά του και έτσι μπορεί να επηρεάσει την προβλεπόμενη βαθμολογία το αντικείμενο που το ενδιαφέρει (αντικείμενο στόχος). Σε όσο περισσότερες γειτονιές χρηστών συμπεριληφθούν τα προφίλ επίθεσης τόσο πιο αποτελεσματική θεωρείται η επίθεση. Οι επιθέσεις Bandagon και segment προσπαθούν να επιτύχουν το στόχο της υψηλής συσχέτισης με τη βαθμολόγηση δημοφιλών αντικειμένων . Βρέθηκε ότι οι Random και average επιθέσεις ήταν πιο αποτελεσματικές *(Dellarocas, 2000).* Σύμφωνα με τους

Mobasher et al., (2005) αυτό μπορεί να οφείλεται σε άλλους παράγοντες και όχι στη στρατηγική που χρησιμοποιήθηκε για την αύξηση της ομοιότητας χωρίς όμως να αναφέρει ποιοι είναι οι παράγοντες αυτοί. Επίσης καλή στρατηγική είναι επίτευξη υψηλής συσχέτισης όχι με περισσότερους χρήστες του συστήματος γενικά αλλά με τους χρήστες που έχουν ήδη βαθμολογήσει το αντικείμενο προς επίθεση.

Συνεργατική μέθοδος που στηρίζεται στην ομοιότητα αντικειμένων

Σύμφωνα με τον *Dellarocas (2000).* τα συστήματα υπόδειξης συνεργασίας που στηρίζονται σε ομοιότητα αντικειμένων είναι πιο ενθετικά σε επιθέσεις από αυτά που στηρίζονται σε ομοιότητα χρηστών. Αυτό σύμφωνα με τους *Mobasher et al., (2005)* μπορεί να οφείλεται στο ότι οι επιθέσεις προσπαθούν να βελτιώσουν την ομοιότητα μεταξύ χρηστών και όχι αντικειμένων.

Βάσει της παραπάνω σκέψης το σύστημα θα πρέπει να βελτιώνει την ομοιότητα αντικειμένων δίνοντας σημασία στην κατανομή των βαθμολογιών του αντικειμένου στόχου. Η βελτίωση της συσχέτισης προκύπτει από την κατάλληλη επιλογή αντικειμένων πλήρωσης.

4.8 Τεχνικές-μοντέλα επίθεσης

Random Attack

Στην τυχαία επίθεση στόχος είναι η μείωση της συνολικής απόδοσης του συστήματος. Αυτής της κατηγορίας οι επιθέσεις δεν εστιάζουν σε συγκεκριμένους χρήστες ή συγκεκριμένα αντικείμενα αλλά στοχεύουν σε όλους τους χρήστες και τα αντικείμενα ισότιμα σε μία προσπάθεια να περιορίσει τη γενική ορθότητα του συστήματος. Ως ένα πραγματικό σενάριο θεωρούμε το διαχειριστή ενός συστήματος που είναι ανταγωνιστικός και θέλει να «χτυπήσει» ένα αντίπαλο σύστημα για να προσελκύσει περισσότερους πελάτες. Η στρατηγική επίθεσης είναι καθαρή, ο αριθμός των αντικειμένων που επιλέγονται για τα προφίλ επίθεσης καθώς και τα αντικείμενα και οι βαθμολογίες τους επιλέγονται τυχαία. Τα προφίλ επίθεσης αποτελούνται από l-1 αντικείμενα που επιλέγονται τυχαία (filler items) από το χώρο που ενδιαφέρει την κακόβουλη οντότητα να παρέμβει. Οι βαθμολογίες που δίδονται στα επιλεγμένη αντικείμενα είναι κι αυτές τυχαίες ακολουθώντας τη ομοιόμορφη κατανομή με μέσο όρο το μέσο του συστήματος, δηλαδή το μέσο όρο όλων των βαθμολογιών των χρηστών του. Αυτού του τύπου η επίθεση δεν χρειάζεται ιδιαίτερες γνώσεις για τις ιδιότητες του συστήματος που είναι προς επίθεση, Αυτές οι επιθέσεις λόγω της απλοϊκότητάς τους δεν έχουν κριθεί ιδιαίτερα

αποτελεσματικές. Πιθανή εξήγηση είναι ότι λόγω της τυχαιότητας οι συσχετίσεις μπορεί να είναι θετικές ή αρνητικές με τους χρήστες,

Average attacks

Σε αυτού του είδους τις επιθέσεις το σύνολο των επιλεγμένων αντικειμένων είναι κενό. Τα αντικείμενα πλήρωσης επιλέγονται τυχαία και σε κάθε ένα από αυτά τίθεται τιμή ίση με το μέσο του αντικειμένου όπως αυτό βαθμολογήθηκε από τους χρήστες του συστήματος. Η αποτελεσματικότητα των συγκεκριμένων επιθέσεων είναι από τις υψηλότερες. Για να οργανωθεί μία average επίθεση είναι απαραίτητη υψηλή γνώση των ιδιοτήτων του συστήματος καθώς χρειάζεται ο μέσος όρος όλων των αντικειμένων πλήρωσης (filler items). Αν και έρευνες έδειξαν ότι αυτές οι επιθέσεις είναι αποτελεσματικές ακόμα και με μικρό αριθμό τέτοιων αντικειμένων πλήρωσης (Williams et al., 2007).

Bandagon attacks

Σε αυτό το μοντέλο επίθεσης εντός των επιλεγμένων αντικειμένων (selected items) συμπεριλαμβάνονται και αντικείμενα που απολαμβάνουν υψηλή δημοφιλία. Κατά συνέπεια αυτά τα προφίλ έχουν υψηλότερη πιθανότητα να είναι μεγαλύτερη ομοιότητα με περισσότερους χρήστες. Στα επιλεγμένα αντικείμενα και στο αντικείμενο στόχος δίνονται οι υψηλότερες βαθμολογίες. Όπως και στις τυχαίες επιθέσεις έτσι κι εδώ τα αντικείμενα πλήρωσης επιλέγονται τυχαία και δίνονται σε αυτά βαθμολογίες ίσες με το μέσο όρο βαθμολογιών που δίνουν οι χρήστες του συστήματος. Θα μπορούσε να θεωρηθεί ότι οι Bandagon επιθέσεις είναι εξέλιξη των τυχαίων επιθέσεων. Ούτε αυτές υψηλό επίπεδο γνώσης του συστήματος και λαμβάνουν ως εισερχόμενες πληροφορίες δεδομένα που είναι διαθέσιμα από δημόσιες πηγές.

Segmented attacks

Αυτή η κατηγορία επίθεσης χρησιμοποιείται για την προώθηση του αντικειμένου στόχου σε όσους χρήστες είναι πιο πιθανόν να επηρεαστούν από τη συγκεκριμένη υπόδειξη. Ως τμήμα (segment) ορίζεται μια ομάδα χρηστών με τάση να τους αρέσουν αντικείμενα με παρόμοια χαρακτηριστικά. Ομάδα χρηστών που βαθμολογούν θετικά γνωστά εστιατόρια με ελληνική κουζίνα αποτελούν ένα τμήμα χρηστών που ενδιαφέρονται για την ελληνική κουζίνα. Όταν ένας επιτιθέμενος θέλει να προωθήσει ένα εστιατόριο με ελληνικά πιάτα θα προσπαθήσει να προσεγγίσει τους συγκεκριμένους χρήστες καθώς υπάρχουν περισσότερες πιθανότητες να τους επηρεάσει θετικά. Ένα κομμάτι των επιλεγμένων

αντικειμένων (selected items) είναι δημοφιλή σε αυτή την ομάδα χρηστών ενώ όλα τα επιλεγμένα καθώς και το αντικείμενο στόχος βαθμολογούνται με την υψηλότερη βαθμολογία. Τα αντικείμενα πλήρωσης (filler items) επιλέγονται τυχαία και λαμβάνουν τη χαμηλότερη βαθμολογία. Έχει αποδειχθεί ότι είναι από τις πιο αποτελεσματικές στρατηγικές και δεν απαιτεί σημαντικές γνώσεις του συστήματος καθώς οι πληροφορίες που χρειάζεται για να δημιουργηθούν τα προφίλ επίθεσης μπορούν να ληφθούν από δημόσιες πηγές.

Η επίθεση segment δημιουργείται ειδικά για τη μεγιστοποίηση της ομοιότητας μεταξύ του αντικειμένου στόχου και άλλων αντικειμένων που ο επιτιθέμενος θεωρεί ότι αρέσουν στους χρήστες. Για αυτό είναι καίρια η επιλογή των χρηστών που θα δεχτούν την επίθεση ώστε να είναι αυξημένες οι πιθανότητες να είναι ευάλωτοι στην ακολουθούμενη στρατηγική επιλογή αντικειμένων. Αν και η επίθεση segment έχει καλό βαθμό διείσδυσης και επηρεασμού δεν είναι τόσο αποτελεσματική η average.

Δημοφιλής επίθεση (Popular attack)

Το βασικό κίνητρο που κρύβεται πίσω από επιθέσεις της συγκεκριμένης κατηγορίας είναι ότι τα δημοφιλέστερα αντικείμενα του χώρου που ενδιαφέρει τον κακόβουλο χρήστη να επέμβει αποτελούν τους πιο κατάλληλους υποψήφιους για χρήστη τους από τα προφίλ επίθεσης. Για παράδειγμα αν έχουμε ένα σύστημα υπόδειξης μουσικής τα μουσικά κομμάτια θα είναι χωρισμένα σε κατηγορίες όπως κλασική, ροκ, pop κτλ. Αν κάποιος θέλει να επιτεθεί στο συγκεκριμένο σύστημα με σκοπό την προώθηση ενός ροκ μουσικού κομματιού είναι λογικό η επίθεση να εστιάσει την προσπάθεια σε άτομα που στο παρελθόν έχουν αγοράσει ή αξιολογήσει θετικά ροκ κομμάτια. Έτσι περιορίζεται ο χώρος των αντικειμένων που θα εξεταστούν για επιλογή στα προφίλ επίθεσης. Αυτή η κατηγορία επιθέσεων είναι σε γενικές γραμμές εύκολη διαδικασία να αναγνωρισθούν αλλά παρουσιάζουν σημαντικά πλεονεκτήματα για τον επιτιθέμενο:

Στα προφίλ των γνήσιων χρηστών είναι αναμενόμενο να υπάρχει σημαντικός αριθμός αντικειμένων που έχουν βαθμολογηθεί και είναι δημοφιλή. Βαθμολογώντας ο επιτιθέμενος δημοφιλή αντικείμενα αυξάνει σημαντικά την ομοιότητά του με περισσότερους χρήστες του συστήματος.

Προφίλ που ακολουθούν τη συγκεκριμένη στρατηγική έχουν υψηλή πιθανότητα να συμπεριληφθούν σε πολλές γειτονιές χρηστών που ενδιαφέρον τον επιτιθέμενο μειώνοντας

έτσι το κόστος επίθεσης όσον αφορά τον αριθμό των προφίλ επίθεσης αλλά και του μεγέθους επίθεσης κάθε προφίλ (δηλαδή του αριθμού των αντικειμένων που έχουν βαθμολογηθεί από κάθε προφίλ)

Τα δημοφιλή αντικείμενα τείνουν να έχουν συνεπείς και υψηλές βαθμολογίες οπότε οι βαθμολογίες που δίνονται είναι πιο κοντά σε αυτές των γνήσιων χρηστών.

Probe Attack

Οι επιθέσεις που στηρίζονται στη δημιουργία προφίλ που βαθμολογούν δημοφιλή αντικείμενα είναι εύκολο να αναγνωρισθούν ιδίως αν έχει δημιουργηθεί μεγάλος αριθμός τέτοιων προφίλ. Ακόμα και να ο επιτιθέμενος προσπαθήσει να διαφοροποιήσει τα αντικείμενα που βαθμολογούνται σε κάθε προφίλ υπάρχουν σημεία που προδίδουν ότι αφορούν επίθεση. Για αυτό χρειάζεται μία στρατηγική που δημιουργεί λιγότερες υποψίες. Η συγκεκριμένη επίθεση αναφέρθηκε πρώτη φορά στο [13] και χρησιμοποιεί τα εξερχόμενα στοιχεία του συστήματος υπόδειξης ως μέσο για την επιλογή επόμενων αντικειμένων προς βαθμολόγηση για το χτίσιμο των προφίλ επίθεσης και τη βαθμολόγησή τους . Βαθμολογώντας αρχικά ένα μικρό αριθμό αντικειμένων ο κακόβουλος χρήστης αρχίζει να ανακρίνει το σύστημα και σταδιακά να χτίζει προφίλ που η κατανομή των βαθμολογιών τους μοιάζει πολύ με αυτή των γνήσιων χρηστών στο σύστημα. Για αυτό το λόγο η πιθανότητα υψηλής ομοιότητας γνήσιων χρηστών και κακόβουλων είναι μεγάλη ενώ μειώνεται η πιθανότητα ανίχνευσης και ταυτοποίησής τους ως κακόβουλα .

(AverageBot Attack) (shilling attacks)

Τη βαθμολόγηση όλων των αντικειμένων του συστήματος περιλαμβάνει η AverageBot στρατηγική. Αν στόχος της επίθεσης είναι η ώθηση (push) ενός αντικειμένου τότε λαμβάνει την υψηλότερη δυνατή βαθμολογία r_{max}. Όλα τα υπόλοιπα αντικείμενα βαθμολογούνται τυχαία με τη χρήση κανονικής κατανομής και έχοντας ως μέσο το μέσο όρο των βαθμολογιών του συγκεκριμένου αντικειμένου στη βάση και τυπική απόκλιση ίση με την τυπική απόκλιση όλων των αντικειμένων. Υπάρχει και η RandomBot (πάλι από το shilling) που όμως δεν είναι αποδοτική επίθεση.

Reverse-Bandagon

Αυτού του τα τύπου οι επιθέσεις όπως και οι love/-hate που ακολουθούν σχεδιάζονται συγκεκριμένα για nuke στρατηγικές. Η reverse-Bandagon αποτελεί μία διαφοροποίηση της Bandagon . Σε αυτή την επίθεση τα επιλεγμένα αντικείμενα τείνουν να λαμβάνουν από τους

χρήστες του συστήματος ιδιαίτερα χαμηλές βαθμολογίες. Η επίθεση βαθμολογεί τελικά χαμηλά το αντικείμενο στόχο αλλά και όλα τα άλλα που συγκροτούν το προφίλ. Συσχετίζει δηλαδή το αντικείμενο στόχο με αντικείμενα που λαμβάνουν ευρέως χαμηλές βαθμολογίες.

love/-hate attacks

Μία love/hate επίθεση ούτε κι αυτή χρειάζεται γνώση της πάνω στο σύστημα και την κατανομή των βαθμολογιών. Αποτελείται από προφίλ επίθεσης όπου το αντικείμενο στόχος έχει λάβει τη χαμηλότερη δυνατή επιτρεπόμενη βαθμολογία ενώ τα υπόλοιπα αντικείμενα που τα συγκροτούν την υψηλότερη δυνατή. Αν και απλή τεχνική είναι ιδιαίτερα αποτελεσματική σε nuke στρατηγική.

4.9 Κατηγοριοποίηση των προφίλ επίθεσης

Από τη στιγμή που υπάρχει γνώση για τους τύπους των επιτυχημένων επιθέσεων μπορεί να μεταχειρισθεί η ανίχνευση της επίθεσης ως πρόβλημα κατηγοριοποίησης όπου το ζητούμενο είναι να κατηγοριοποιηθούν τα προφίλ στα γνωστά μοντέλα επίθεσης.

4.9.1 Ανίχνευση ιδιοτήτων

Για το διαχωρισμό των προφίλ σε κακόβουλα ή μη χρησιμοποιούνται ιδιότητες που προκύπτουν από τα προφίλ. Οι ιδιότητες χωρίζονται σε γενικές και σχετιζόμενες με τον τύπο επίθεσης. Οι γενικές βασίζονται σε περιγραφικές στατιστικές που σκοπό έχουν την ανίχνευση χαρακτηριστικών που διαχωρίζουν τα γνήσια προφίλ από τα προφίλ επίθεσης. Ενώ οι ιδιότητες που σχετίζονται με τον τύπο επίθεσης προσπαθούν να ανιχνεύσουν χαρακτηριστικά των προφίλ που το συσχετίζουν με κάποιο γνωστό μοντέλο επίθεσης. Τέλος ορίζονται οι ιδιότητες που σχετίζονται και εμφανίζονται σε πολλά προφίλ και στοχεύουν στην εύρεση συγκεντρώσεων “concentrations” στο σύνολο των προφίλ.

4.9.2 Γενικές ιδιότητες

Είναι αναμενόμενο ότι η στατιστική ταυτότητα των προφίλ επίθεσης θα διαφέρει από εκείνη των γνήσιων χρηστών. Αυτή η διαφοροποίηση πηγάζει από δύο σημεία α)τη βαθμολογία του αντικειμένου στόχου και β) την κατανομή των βαθμολογιών για τα αντικείμενα πλήρωσης (filler items). Είναι αποδεκτό πως είναι πολύ δύσκολο αν όχι ακατόρθωτο να έχει κάποιος πλήρεις γνώσεις για τη λειτουργία και τις ιδιότητες του συστήματος που επιθυμεί να επιτεθεί. Έτσι τα κακόβουλα προφίλ θα διαφέρουν από εκείνα

των γνήσιων χρηστών. Τέτοιες διαφοροποιήσεις μπορεί να αφορούν την ύποπτη απομάκρυνση των βαθμολογιών από το μέσο όρο των βαθμολογιών του συστήματος ή την ύπαρξη πολλών βαθμολογημένων αντικειμένων από κάποια προφίλ. Έτσι δημιουργώντας ιδιότητες που μπορούν να ποσοτικοποιήσουν τις παραπάνω ανωμαλίες μπορεί να ενισχυθεί η ανίχνευση κακόβουλων προφίλ.

4.9.3 Μετρικές γενικών ιδιοτήτων

- **Rating Deviation from Mean Agreement (RDMA)** χρησιμοποιείται για να μετρήσει την απόκλιση των βαθμολογιών των αντικειμένων του κάθε προφίλ από το μέσο deviation κάθε αντικειμένου σταθμισμένο δια τον αριθμό των βαθμολογιών:

$$RDMA_u = \frac{\sum_{i=0}^{N_u} \frac{|r_{u,i} - \overline{r_i}|}{R_{U,i}}}{N_u}$$

Εξίσωση 18

Όπου N_u είναι ο αριθμός των αντικειμένων που έχει βαθμολογήσει ο χρήστης u , $r_{u,i}$ είναι η βαθμολογία του u για το αντικείμενο i , η μέση βαθμολογία που έχει δοθεί από τους χρήστες στο αντικείμενο i και $R_{U,i}$ ο αριθμός των βαθμολογιών που έχει λάβει το αντικείμενο i από το σύνολο των χρηστών του συστήματος.

- **Weighted Degree of Agreement (WDA) χρησιμοποιείται** για να αθροίσει τις διαφορές των βαθμολογιών ενός προφίλ από τη μέση βαθμολογία κάθε αντικειμένου προς τη συχνότητα βαθμολόγησής του. Ουσιαστικά αποτελεί τον αριθμητή του RDMA.

Weighted Deviation from Mean Agreement (WDMA) χρησιμοποιείται για την ανίχνευση ανωμαλιών σταθμίζοντας υψηλά τη deviation των βαθμολογιών αντικειμένων που εμφανίζονται αραιά. Αυτή η ιδιότητα ίσως είναι η πιο χρήσιμη από όσες έχουν αναφερθεί. Η διαφορά με το RDMA είναι ότι ο αριθμός των βαθμολογιών που έχουν δοθεί για κάθε αντικείμενο υψώνεται στο τετράγωνο στην άθροιση μειώνοντας το βάρος όσων αντικειμένων έχουν βαθμολογηθεί πολλές φορές. Το WDMA υπολογίζεται ως εξής:

$$\mathbf{WDMA}_u = \sqrt{\frac{\sum_{i=0}^{N_u} \frac{|r_{u,i} - \overline{r_i}|}{R_{U,i}2}}{N_u}}$$

Εξίσωση 19

$r_{u,i}$

Όπου U είναι το σύνολο των χρηστών του συστήματος u ο χρήστης, είναι η βαθμολογία του u για το αντικείμενο i , η μέση βαθμολογία που έχει δοθεί από τους χρήστες στο αντικείμενο i και $R_{U,i}$ ο αριθμός των βαθμολογιών που έχει λάβει το αντικείμενο i από το σύνολο των χρηστών του συστήματος.

- **Length Variance (LenghtVAr)** σκοπό έχει να μετρήσει πόσο απέχει το μέγεθος ενός προφίλ (τον αριθμό των βαθμολογημένων αντικειμένων δηλαδή) από το μέσο όρο του μεγέθους των προφίλ του συστήματος. Αν το εξεταζόμενο προφίλ είναι πολύ μεγάλο είναι ύποπτο καθώς είναι απίθανο ένας χρήστης να έχει δει και αξιολογήσει τόσα πολλά αντικείμενα για τα οποία πρέπει να εισάγει δεδομένα χειρονακτικά . Το συγκεκριμένο χαρακτηριστικό είναι αποτελεσματικό για την ανίχνευση προφίλ με πολλά αντικείμενα πλήρωσης (filler items)

- **Degree of Similariy with Top Neighbors (DegSim)** βρίσκει τη μέση ομοιότητα των k πιο κοντινών γειτόνων. Έρευνες έχουν δείξει ότι τα προφίλ επίθεσης μοιάζουν περισσότερο με αυτά των κορυφαίων 25 στενότερων γειτόνων παρά με τα προφίλ των κοινών χρηστών. Επίσης υπάρχει και μία δεύτερη συνιστώσα του DegSim που μειώνει τη μέση ομοιότητα αν ο γείτονας έχει λιγότερα από d αντικείμενα κοινά. Αυτό το χαρακτηριστικό είναι περισσότερο χρήσιμο για την αξιολόγηση μικρών προφίλ.

Ιδιότητες που σχετίζονται με τον τύπο της επίθεσης

Ερευνητικές εργασίες έχουν δείξει ότι οι γενικές ιδιότητες δεν επαρκούν για το διαχωρισμό των γνήσιων προφίλ από τα εκκεντρικά προφίλ που όμως είναι γνήσια. [13] ιδίως για μικρά προφίλ που περιέχουν λίγα αντικείμενα πλήρωσης (filler items). Τέτοιες επιθέσεις μπορούν να είναι επιτυχημένες στον επηρεασμό των αποτελεσμάτων υπόδειξης για αυτό είναι αναγκαία η ενίσχυση των γενικών ιδιοτήτων με ιδιότητες που σχεδιάζονται ειδικά με

σκοπό να ταιριάξουν χαρακτηριστικά των προφίλ με χαρακτηριστικά που απορρέουν από τα διαφορετικές μοντέλα επιθέσεις .

Οι επιθέσεις μπορούν να χαρακτηριστούν βάσει του πώς δομούνται τα προφίλ που τις αποτελούν. Δηλαδή πώς κατασκευάζονται τα τμήματα του αντικείμενο στόχος (target item), αντικείμενα πλήρωσης (filler items), αντικείμενα επιλογής (selection items). Οι ιδιότητες που σχετίζονται με τον τύπο επίθεσης προσπαθούν να αναγνωρίσουν την «υπογραφή» κάθε τύπου επίθεσης.

Για την ανίχνευση της υπογραφής κάθε επίθεσης υπάρχου κάποιες μετρικές που μπορούν να φανούν χρήσιμες και να χρησιμοποιηθούν σε διάφορα μοντέλα ανίχνευσης επιθέσεων. Αυτές οι ιδιότητες έχουν σχεδιαστεί να αναγνωρίζουν χαρακτηριστικά των αντικειμένων πλήρωσης που ίσως να υποδεικνύουν ότι το προφίλ δεν δημιουργήθηκε από γνήσιο χρήστη. Τέτοιες ιδιότητες είναι:

4.9.4 Μετρικές ιδιοτήτων που σχετίζονται με τον τύπο της επίθεσης

- **Filler Mean Variance(FVM):** η διασπορά της βαθμολογίας κάθε αντικειμένου που υποτίθεται ότι ανήκει στα αντικείμενα πλήρωσης από το μέσο όρο βαθμολογίας του κάθε αντικειμένου. Η σκέψη πίσω από τη συγκεκριμένη μετρική αφορά την ανίχνευση ανωμαλιών από υψηλή ή χαμηλή διασπορά ανάμεσα στη μέση βαθμολογία κάθε αντικειμένου και της βαθμολογίας των αντικειμένων πλήρωσης του υπο αμφισβήτηση προφίλ . Το FVM για ένα χρήστη u και για ένα μοντέλο επίθεσης m υπολογίζεται ως εξής:

$$\mathbf{FVM}_{u,m} = \sum_{i \in P_{u,F_m}} \frac{\left(r_{u,i} - \overline{r_i}\right)^2}{\left|P_{u,F_m}\right|} = 0$$

Εξίσωση 20

P_{u,F_m}

Όπου είναι κομμάτι του προφίλ του χρήστη που έχει θεωρηθεί ότι είναι αντικείμενο πλήρωσης F του μοντέλο m , $\mathbf{r_{u,i}}$ η βαθμολογία του χρήστη u για το αντικείμενο $\mathbf{i}$ και $\overline{\mathbf{r_i}}$ ο μέσος όρος βαθμολογιών του αντικειμένου $\mathbf{i}$ από όλους τους

χρήστες και $|P_{u,r_m}|$ ο αριθμός των βαθμολογιών που βάση της υπόθεσης περιλαμβάνονται στο τμήμα πλήρωσης του προφίλ P_u του μοντέλου m .

- **Filler Mean Difference**, που είναι η μέση απόλυτη τιμή της διαφοράς μεταξύ της βαθμολογίας του χρήστη και της μέσης τιμής των βαθμολογιών του αντικειμένου που βάσει της υπόθεσης αποτελεί κομμάτι του τμήματος πλήρωσης.

- **Filler Average Correlation**, μετρά την συσχέτιση μεταξύ των βαθμολογιών των αντικειμένων πλήρωσης του προφίλ και της μέσης βαθμολογίας για κάθε αντικείμενο. Αυτές οι ιδιότητες χρησιμοποιούνται για την ανίχνευση του καλύτερου ταιριάσματος κάθε προφίλ με συγκεκριμένου τύπου επίθεση με βάσει την υπόθεση βέβαια ότι αποτελεί το συγκεκριμένο προφίλ ένα κομμάτι της επίθεσης.

Διαχείριση επίθεσης

Αφού έχουν ανιχνευτεί τα προφίλ επίθεσης το ερώτημα που γεννάται αφορά τη στάση του συστήματος ώστε να εξαλείψει ή να μειώσει την επιρροή της επίθεσης στο σύστημα. Κατά το ιδεατό σενάριο το σύστημα θα αναγνώριζε όλα τα κακόβουλα προφίλ και στη συνέχεια θα τα αγνοούσε ώστε να μην διοχετευτούν μεροληπτικές βαθμολογίες στη διαδικασία πρόβλεψης. Όμως σε ένα πραγματικό σενάριο η ανίχνευση των κακόβουλων προφίλ δεν μπορεί να είναι 100% ορθή. Μπορεί κάποια προφίλ εκκεντρικών χρηστών να θεωρηθούν λανθασμένα κακόβουλα ή και κάποια κακόβουλα να μην ανιχνευτούν από το σύστημα. Μία προσπάθεια ποσοτικοποίησης αυτής της αμφιβολίας την μετατρέπει σε βάρος. Δηλαδή ανάλογα με το βαθμό υποψίας για την μεροληψία του, κάθε προφίλ λαμβάνει ένα βάρος το οποίο συνυπολογίζεται στην διαδικασία παραγωγής προβλέψεων.

Αν και συνήθως η έρευνα εστιάζει στην άμεση επιρροή του συστήματος από τις επιθέσεις δεν αφορά μόνο αυτό η ευρωστία του συστήματος. Ένα σύστημα για να θεωρηθεί εύρωστο πρέπει όχι μόνο να αντέχει την άμεση κακόβουλη επίθεση σε συγκεκριμένα αντικείμενα αλλά πρέπει να προσφέρει ορθές προβλέψεις για όλα τα αντικείμενα που περιλαμβάνει το σύστημα.

4.9.5 Μετρικές αξιολόγησης κατηγοριοποίησης

Για τη σύγκριση διαφορετικών αλγορίθμων κατηγοριοποίησης σημαντική είναι η εύρεση μέτρων που αξιολογούν την απόδοση της κατηγοριοποίησης. Μία ορθή κατηγοριοποίηση θα εξαλείψει την επιρροή της επίθεσης. Διάσταση με βαρύτητα που αν δεν προσεχθεί μπορεί να μειώσει την ορθότητα των προβλέψεων είναι και η κατηγοριοποίηση γνησίων προφίλ ως κακόβουλα.

Για την μέτρηση της απόδοσης της κατηγοριοποίησης χρησιμοποιούνται η ειδικότητα (specificity) και η ευαισθησία. (sensitivity).

$$\mathbf{specificity} = \frac{\mathbf{\#\,true\,positives}}{\{\mathbf{\#true\,positives} + \mathbf{\#false\,negatives}\}}$$

Εξίσωση 21

$$\mathbf{sensitivity} = \frac{\mathbf{\#true\,negatives}}{\{\mathbf{\#\,true\,negatives} + \mathbf{\#false\,positives}\}}$$

Εξίσωση 22

Όπου $\mathbf{true\,positives}$ ο αριθμός των προφίλ που κατηγοριοποιήθηκαν ορθά, $\mathbf{false\,positives}$ ο αριθμός των προφίλ που κατηγοριοποιήθηκαν λανθασμένα ως κακόβουλα , $\mathbf{true\,negatives}$ τα προφίλ που σωστά κατηγοριοποιήθηκαν ως αυθεντικά και $\mathbf{false\,negatives}$ όσα προφίλ κατηγοριοποιηθήκαν ως αυθεντικά ενώ δεν είναι. Έτσι η υπολογίζει το ποσοστό των προφίλ επίθεσης που κατηγοριοποιήθηκαν ορθά ενώ το το ποσοστό των γνήσιων προφίλ που κατηγοριοποιήθηκαν ορθά. Πέραν των παραπάνω μετρικών που αφορούν τον υπολογισμό ορθών κατηγοριοποιήσεων σημαντικός είναι και ο υπολογισμός της επιρροής στην αντιλαμβανόμενη ορθότητα των υποδείξεων κατηγοριοποιήσεις που λανθασμένα έθεσαν γνήσια προφίλ ως κακόβουλα. Αξιολογούμε τη συγκεκριμένη επιρροή με τη κοινώς χρησιμοποιούμενη μετρική MAE η οποία αξιολογεί την αντιλαμβανόμενη ορθότητα των υποδείξεων.

Ένας αλγόριθμος κατηγοριοποίησης μπορεί να αυξήσει την ευρωστία του συστήματος. Η επιλογή του αλγορίθμου κατηγοριοποίησης παίζει επίσης σημαντικό ρόλο. Εξετάζοντας τρεις γνωστούς αλγορίθμους κατηγοριοποίησης k-NN, C4.5 και SVN προέκυψε ότι οι SVN και C4.5 σχεδόν τέλεια μπορούσαν να εντοπίσουν τα κακόβουλα προφίλ (sensitivity) ενώ ο k-NN είχε κάποια προβλήματα όταν ο αριθμός των αντικειμένων πλήρωσης του προφίλ ήταν μικρό. Αντίθετα κατά την αξιολόγηση του καλύτερα αποτελέσματα έδειξε ο k-NN όπου είχε λιγότερες λανθασμένες αξιολογήσει γνήσιων προφίλ ως κακόβουλα σε σχέση με τους SVN και C4.5. Συνδυαστικά για sensitivity και specificity καλύτερη επίδοση είχε ο SVN.

4.9.6 Μετρικές ευρωστίας

Οι μετρικές είναι πολύ σημαντικές καθώς γιατί είναι πιθανόν οι προβλέψεις να επηρεάσουν τις αποφάσεις του χρήστη με ένα συγκεκριμένο τρόπο, αγορά ή μη αγορά ενός προϊόντος (ή τουλάχιστον σκέψεις για αγορά ή όχι)

- **Prediction Shift**

Για τη μέτρηση της αποτελεσματικότητας μίας επίθεσης ένα μέτρο που χρησιμοποιείται ευρύτατα είναι η μεταβολή πρόβλεψης (Prediction Shift). Η μεταβολή πρόβλεψης για το αντικείμενο στόχος είναι η διαφορά της μέσης προβλεπόμενης βαθμολογίας του αντικειμένου πριν και μετά την επίθεση για όλους τους χρήστες. Ως μέση μεταβολή πρόβλεψης ονομάζεται η μέση αλλαγή πρόβλεψης πριν και μετά την επίθεση για όλα τα αντικείμενα στόχους. Βάσει του *(Dellarocas, 2000)* ορίζεται:

Έστω U το σύνολο των χρηστών , I το σύνολο των αντικειμένων του συστήματος και $\Delta_{u,i}$ η μεταβολή πρόβλεψης του χρήστη u για το αντικείμενο i .Δηλαδή όπου είναι η τιμή πρόβλεψης μετά την επίθεση και $p_{u,i}$ η πρόβλεψη πριν την επίθεση. Η μέση μεταβολή πρόβλεψης υπολογίζεται ως εξής:

$$\Delta_i = \frac{\sum_{u \in U} \Delta_{u,i}}{|U|}$$

Εξίσωση 23

Η πρόβλεψη μεταβολής ενός μοντέλου είναι η μέση πρόβλεψη μεταβολής για όλα τα αντικείμενα στόχους και υπολογίζεται ως εξής:

$$\Delta_{\square} = \square \frac{\sum_{i \in I} \Delta_i}{|\mathbf{I}|}$$

Εξίσωση 24

- **Hit ratio**

Μετρά την επίδραση των προφίλ επίθεσης στις κορυφαίες k προτάσεις. Το αποτέλεσμα των αλγορίθμων υπόδειξης είναι μία λίστα με τις πλέον κατάλληλες υποδείξεις για κάθε χρήστη και η μετρική αυτή βρίσκει το κομμάτι των χρηστών που επηρεάστηκαν από την επίθεση. Έστω $H_{u\,i}$ = 1 όπου i αντικείμενο που εμπεριέχεται στις κορυφαίες k προτάσεις του χρήστη ενώ σε αντίθετη περίπτωση $H_{u\,i}$ = 0. Το hit ratio παίρνει συνεπώς διακριτές τιμές 0-1 ενώ όταν εκφράζεται ως ποσοστό υπολογίζεται ως εξής:

$$H = \frac{100}{N} \times \sum_{u} \Delta H_{u,i,}$$

Εξίσωση 25

όπου N ο αριθμός των χρηστών.

- **Normalised Absolute Error NAE)**

Έστω ένα σύνολο A από ζευγάρια χρηστών-αντικειμένων που δεν έχουν βαθμολογηθεί. Θεωρούμε ότι οι βαθμολογίες που μπορούν να αποδοθούν κυμαίνονται μεταξύ $R_{max\square}$ και $R_{min\square}$. Ευρωστία του συνόλου A ορίζουμε το κανονικοποιημένο απόλυτο λάθος (normalised absolute Error NAE) της πρόβλεψης πριν και μετά την επίθεση T με την εξής έκφραση:

$$\max\left\{\left|p_{a,j} - R_{max\square}\right|, \max\left\{\left|p_{a,j} - R_{min\square}\right|\right\}\right)$$

Εξίσωση 26

Επιπρόσθετα η ευρωστία ενός συστήματος συνεργατικής υπόδειξης για το σύνολο A που υπόκειται σε επίθεση κόστους c δίνεται από την παρακάτω έκφραση:

$$Robust(A, c) = min_{(i,j) \notin A a,j} Robust(a, j, c)$$

Εξίσωση 27

Αξιοσημείωτο είναι ότι η έννοια της ευρωστίας δεν συμπεριλαμβάνει την έννοια της ορθής βαθμολόγησης για τα ζευγάρια χρήστη-αντικείμενο. Δηλαδή εδώ φαίνεται και η διαφοροποίηση των εννοιών της ορθότητας (accuracy) και της ευρωστίας (robustness)του συστήματος.

- **MAE**

Ο μέσος όρος της απόλυτης τιμής της διαφοράς μεταξύ των βαθμολογιών πρόβλεψης και των πραγματικών βαθμολογιών που έχουν δώσει οι χρήστες. Η απόδοση του αλγορίθμου αξιολογείται πριν και μετά την επίθεση ώστε να προσδιοριστεί ο αντίκτυπος της επίθεσης στις προβλέψεις. Αν και το MAE θεωρείται κλασική επιλογή για τη ν αξιολόγηση ενός συστήματος υπόδειξης πολλοί ερευνητές θεωρούν ότι υπάρχει κενό ανάμεσα σε αυτό που μετράει και σε αυτό που θα έπρεπε να μετράει. Τα συστήματα υπόδειξης είναι εργαλείο λήψης αποφάσεων και γενικά είναι πιο σοβαρό το ότι ένα αντικείμενο προτάθηκε από το σύστημα παρά η πρόβλεψη της βαθμολογίας του.

Σύμφωνα με τον Shafer το πιο συχνό αποτέλεσμα της δράσης των συστημάτων υπόδειξης είναι η απόκτηση ενός προϊόντος. Για αυτό θα πρέπει να γίνει εστίαση στην ποιότητα των υποδείξεων και όχι των προβλέψεων.

Επιπλέον επειδή ο χρήστης ασχολείται μόνο με τις k πιο υψηλές προτάσεις της λίστας με τις προτάσεις και δεν ψάχνει παραπέρα, μόνο τα αντικείμενα που είναι υψηλά στη λίστα θα πρέπει να χρησιμοποιηθούν για την αξιολόγηση της ποιότητας του αλγορίθμου.

- **Stability of Prediction**

Η μετρική ορίστηκε στο [14] το οποίο μετρά τον αριθμό των προβλέψεων για τα αντικείμενα που έχουν τεθεί ως στόχοι και που η τιμή τους δεν μεταβλήθηκε πάνω από ένα κατώφλι μετά την επίθεση. Το POA θεωρεί όλες τις μεταβολές στις προβλέψεις ισότιμες , σαν να τέχουν την ίδια σημασία. Διαισθητικά όμως φαίνεται ότι δεν είναι αυτή η πραγματικότητα. Δηλαδή έστω μια κλίμακα 1 έως 5, αν ένα αντικείμενα πριν την επίθεση είχε πρόβλεψη 2 και μετά την επίθεση η τιμή γίνει 3 έχει χαμηλότερη αξία από την μεταβολή της πρόβλεψης από το 4 στο 5.

- **Power of Attack**

Ένα άλλο μέτρο πάλι από το [14] είναι το POA το που ορίζεται ως η μέση μεταβολή της πρόβλεψης ως προς την τιμή στόχο που έχει θέσει ο επιτιθέμενος. Συνήθως η τιμή στόχος είναι η υψηλότερη δυνατή r_{max} ή χαμηλότερη . Το POA έχει το ίδιο μειονέκτημα με το Stability of Prediction, δηλαδή δίνει αξία μόνο στο μέτρο της απόλυτης μεταβλητής και όχι στη σημασία της μεταβολής.

$$POA(A_j, j, T) = 1 - \frac{1}{(|A_j|) \sum_{a \in A_j} k_{a,j}}$$

Εξίσωση 28

Όπου αν $p'_{aj} = R_{target}$ ενώ αλλιώς Ορίζουμε το μέσο POA για όλες τις επιθέσεις με κόστος c για το υποσύνολο A_j ως:

$$POA(A_j, j, c) = 1 - \frac{1}{(|T_c|) \sum_{T \in T_c} POA(A_j, j, T)}$$

Εξίσωση 29

Και έστω I το σύνολο των αντικειμένων του A τότε ως μέσο POA μίας επίθεσης με κόστος c στο σύνολο A ορίζεται ως:

$$POA(A, c) =$$

Εξίσωση 30

- **Expected Top-N Oppucancy**

Η μετρική ορίζεται ως ο αριθμός των φορών που ένα αντικείμενο στόχος εισέρχεται στη λίστα με τα κορυφαία αντικείμενα βάσει πρόβλεψης προς υπόδειξη. Προσμετρείται για όλους τους χρήστες . Για παράδειγμα έστω ότι υπάρχει λίστα με τα 5 κορυφαία αντικείμενα προς υπόδειξη (Πίνακας 6).

Rank	Item	Pred
1	B	5.0

2	E	4.9
3	D	4.7
4	A	4.5
4	C	4.5
4	F	4.5

Πίνακας 6 Λίστα με κορυφαία αντικείμενα

Έστω ότι τα αντικείμενα που έχουν τεθεί ως στόχοι είναι τα E και F, ενώ τα A,C,F ανταγωνίζονται για την τέταρτη και πέμπτη θέση της λίστας λόγω ισοβαθμίας. Η επιλογή των δύο αντικειμένων που θα καλύψουν τις θέσεις γίνεται τυχαία οπότε θεωρούμε ότι κάθε αντικείμενο έχει 2/3 πιθανότητα να συμπεριληφθεί. Οπότε επειδή το E έχει σίγουρη παρουσία στο λίστα παίρνει βαθμό 1 και το F $\overline{0.666}$ οπότε το Expected Top-N Oppucancy για το συγκεκριμένο χρήστη είναι $\overline{1.666}$.

- **Good prediction (GP)**

Ως καλή πρόβλεψη ενός αντικειμένου $\mathbf{i}$ ορίζεται ο αριθμός των φορών που αληθεύει η παρακάτω έκφραση:

$$\epsilon \mathbf{U_j}$$

Όπου το δ_g αποτελεί την τιμή του κατωφλίου , $\mathbf{U_j}$ είναι το σύνολο των γνήσιων χρηστών που έχουν βαθμολογήσει το αντικείμενο $\mathbf{i}$. Συχνή επιλογή κατωφλιού είναι το $\mathbf{r_{max}}$ -1

- **Bad prediction (BP)**

Στην περίπτωση της nuke επίθεσης χρησιμοποιείται αντίστοιχα το μέτρο «κακής πρόβλεψης» Bad prediction BP. Ως κακή πρόβλεψη ενός αντικειμένου $\mathbf{i}$ ορίζεται ο αριθμός των φορών που αληθεύει η παρακάτω έκφραση μετά την εκδήλωση της επίθεσης :

$$\epsilon \mathbf{U_j}$$

Εξίσωση 31

Όπου το δ_g αποτελεί την τιμή του κατωφλίου , U_i είναι το σύνολο των γνήσιων χρηστών που έχουν βαθμολογήσει το αντικείμενο i. Συχνή επιλογή κατωφλιού είναι το $[(r]_{max} + r_{min}])/2$

4.9.7 Μετρικές ανίχνευσης επίθεσης

Στη συνέχεια αναφέρονται μετρικές που βοηθούν στην αναγνώριση επιθέσεων στο σύστημα:

- **Number Of Prediction Differences (NPD)**

Η μετρική NPD ορίζεται για κάθε χρήστη ως ο αριθμών των αλλαγών στις προβλέψεις του συστήματος που προκύπτουν από την αφαίρεση ενός χρήστη από το σύστημα.

Για τους περισσότερους χρήστες το NPD είναι ιδιαίτερα χαμηλό ενώ λίγα άτομα μόνο έχουν υψηλή τιμή. Μετά την επίθεση τα άτομα που πριν είχαν τις υψηλότερες τιμές NPD τώρα θα πέσει λίγο η τιμή τους ενώ τις υψηλότερες πλέον θα έχουν οι κακόβουλοι λόγω της προσπάθειάς τους να μοιάζουν με όσο το δυνατόν περισσότερους χρήστες κι έτσι να μπορούν να συμπεριληφθούν σε όσο το δυνατόν περισσότερες γειτονιές τους. Παρόλα αυτά όμως υπάρχει μία επικάλυψη γνήσιων και κακόβουλων χρηστών που σε έρευνες όπως η παρούσα ανέρχεται στο 0.3%.

- **St&ard Deviation in User's Ratings**

Η μετρική υπολογίζει το βαθμό που μία βαθμολογία ενός χρήστη απέχει από το μέσο όρο όλων των βαθμολογιών του. Οι επιθέσεις RandomBot είναι πολύ εύκολο να αναγνωρισθούν λόγω της κατανομής των βαθμολογιών του χρήστη με τυπική απόκλιση κοντά στο μηδέν. Οι περισσότεροι χρήστες έχουν μεγαλύτερη τιμή εντροπίας στις βαθμολογίες τους και κάποιες φορές δίνουν υπερβολικές τιμές όπως τη μικρότερη δυνατή βαθμολογία. Παρόλα αυτά κάποιοι χρήστες έχουν χαμηλή τυπική απόκλιση και για αυτό τα προφίλ επίθεσης προσπαθούν να αυξήσουν την εντροπία τους ώστε να μιμηθούν τη βαθμολογική συμπεριφορά αυτών των χρηστών και να αποφύγουν την αναγνώριση. Έτσι αυξάνοντας την εντροπία μειώνουν την δύναμη της επίθεσης καθώς μειώνεται ο βαθμός της ομοιότητας των κακόβουλων προφίλ με τα αυθεντικά. Έτσι η ανάλυση του μοτίβου των βαθμολογιών έχει θετικό αποτέλεσμα στην αντίσταση κατά των επιθέσεων αφού αναγκάζει τα κακόβουλα προφίλ να «καμουφλάρουν» τη συμπεριφορά τους και τελικά να μειώνουν την επίδρασή τους στο σύστημα.

- **Degree of Agreement with Others**

Ο βαθμός συμφωνίας είναι η μέση απόκλιση των βαθμολογιών ενός χρήστη από το μέσο όρο της βαθμολογίας κάθε αντικειμένου.

Εξίσωση 32

Όπου R_{ia} είναι η βαθμολογία που έδωσε το χρήστης i στο αντικείμενο a και ο μέσος όρος των βαθμολογιών που έχει λάβει το αντικείμενο a.

- **Degree of Similarity with Top Neighbors**

Όπως αναφέρεται και από την ονομασία του το συγκεκριμένο μέτρο υπολογίζει το μέσο βάρος ομοιότητας κάθε χρήστη με τους κορυφαίους K γείτονές του βάσει της έκφρασης που ακολουθεί:

$$\frac{\sum_{i=0}^{N_j} \frac{r_{i,j} - Avgr_i}{NR_i}}{N_j}$$

Εξίσωση 33

- **Rating Deviation from Mean Agreement (RDMA**

Αποτελεί ένα μέτρο που υπολογίζει την απόκλιση της συμφωνίας στις βαθμολογίες του τρέχοντα χρήστη με τους άλλους χρήστες σε ένα σύνολο αντικειμένων συνδυαζόμενο με την αντίστροφη βαθμολογική συχνότητα αυτών των αντικειμένων. Το RDMA υπολογίζεται ως εξής:

Εξίσωση 34

Όπου N_j είναι ο αριθμός των αντικειμένων όπου ο χρήστης j έχει βαθμολογήσει , $r_{i,j}$ είναι η βαθμολογία του χρήστη j στο αντικείμενο i και NR_i είναι ο συνολικός αριθμός βαθμολογιών που έχουν δοθεί στο αντικείμενο i .

4.9.8 Επιθέσεις με μεγάλο αριθμό ψεύτικων προφίλ (shilling attacks)

Συνήθως έχουν σκοπό την προώθηση ή την denote των αντικειμένων του συστήματος. Τα προφίλ που συμμετέχουν στην επίθεση έχουν υψηλό RDMA καθώς ο αριθμητής (η διαφορά της βαθμολογίας από το μέσο όρο) θα είναι υψηλή ενώ ο παρονομαστής (το άθροισμα των βαθμολογιών που έχουν δοθεί στο συγκεκριμένο αντικείμενο) θα είναι χαμηλό. Έτσι συνολικά όλο το κλάσμα θα έχει υψηλή τιμή και οι επιτιθέμενοι ξεχωρίζουν βάσει αυτής της τιμής και το σύστημα δεν περιλαμβάνει τα συγκεκριμένα προφίλ στη διαδικασία παραγωγής προτάσεων για τους υπόλοιπους χρήστες.

4.10 Αντιμετώπιση απειλών

Έρευνες έχουν δείξει ότι οι άνθρωποι έχουν την τάση να συμφωνούν με τη γνώμη των άλλων ακόμα κι αν αυτές είναι λανθασμένες. Αυτή η πρόταση επιβεβαιώνεται από την έρευνα στο πεδίο της ψυχολογίας από τον *Asch (1956)* όπου έκανε ένα απλό πείραμα βάζοντας άτομα –τα υποκείμενα του πειράματος- να πραγματοποιήσουν μικρές ενέργειες σε μία ομάδα. Τα υπόλοιπα άτομα της ομάδας δούλευαν για τον ερευνητή και επίτηδες του έδιναν λανθασμένες απαντήσεις. Ακόμα κι αν οι απαντήσεις ήταν οφθαλμοφανώς λανθασμένες τα υποκείμενα συμφώνησαν με αυτές και έλαβαν λάθος αποφάσεις το 1/3 των επαναλήψεων. Ο *Cosley et al., (2003)* βρήκε ότι οι χρήστες πραγματικά επηρεάζονται από τις προβλέψεις που έχουν παραποιηθεί κακοβούλως και προτάσσονται από το σύστημα. Οι χρήστες δηλαδή τείνουν να βαθμολογούν τα αντικείμενα που τους προτάσσονται κοντά στη βαθμολογία πρόβλεψης ανεξαρτήτως του κατά πόσο τελικά θεωρούν ορθή την πρόταση. Επιπρόσθετα ο αντίκτυπος της επίθεσης μπορεί να πολλαπλασιασθεί καθώς όταν οι χρήστες επηρεαστούν από τις προβλέψεις που έχουν προκύψει μετά την επιτυχή επίθεση θα επηρεάσουν κι αυτοί μέσω της ανάλογης βαθμολογίας τους άλλους χρήστες κι έτσι θα έχουμε διάδοση της πρότασης.

Η άμυνα ενάντια στις επιθέσεις μπορεί να λάβει πολλές μορφές. Κάποιες συνεργατικές μέθοδοι μπορεί να είναι πιο ανθεκτικές σε επιθέσεις από άλλες, ο βαθμός ευρωστία τους ποικίλει ανάλογα τη λογική κατασκευής τους. Οι πρόσφατες ερευνητικές προσπάθειες στόχο έχουν την προστασία των συστημάτων από αυτές τις κακόβουλες επιθέσεις. Από την έρευνα έχουν προκύψει κυρίως δύο κατηγορίες τεχνικών άμυνας και προστασίας αυτές που αυξάνουν την ευρωστία του συστήματος και τεχνικές που στόχο έχουν την εύρεση και παραμεθοριοποίηση των κακόβουλων προφίλ.

Το κόστος μίας επίθεσης όπως αναφέρθηκε και σε προηγούμενο σημείο μπορεί να αποτιμηθεί σε προσπάθεια , χρόνο αλλά και σε χρήματα. Μία αναζήτηση που τελικά δεν καλύπτει όπως θα έπρεπε τις ανάγκες του χρήστη σίγουρα θα τον δυσαρεστήσουν καθώς επένδυσε κόπο και ώρα από το χρόνο του σε αυτό. Αν το αντικείμενο που επέλεξε βάσει του συστήματος τελικά το αγόρασε τότε πλέον ο χρήστης έχει και χρηματικό κόστος. Ένας δυσαρεστημένος χρήστης θα χάσει την εμπιστοσύνη του προς το σύστημα . θα μειωθεί το αίσθημα αξιοπιστίας και θα απευθυνθεί σε εναλλακτικές. Σκοπός λοιπόν του ιδιοκτήτη του συστήματος είναι η αντιμετώπιση των επιθέσεων σε όσον το δυνατόν λιγότερο χρόνο και όσο πιο αποτελεσματικά είναι δυνατόν να γίνει ώστε να έχει ικανοποιημένους χρήστες που νιώθουν ασφάλεια και εμπιστοσύνη ότι οι όποιες επιθέσεις δεν επηρεάζουν τις προβλέψει που λαμβάνουν.

Ως στόχοι για την αντιμετώπιση των κακόβουλων επιθέσεων σε ένα σύστημα τίθενται οι εξής:

- Ελαχιστοποίηση της επιρροής της επίθεσης στη λειτουργία και τη συμπεριφορά του συστήματος
- Μείωση της πιθανότητας να είναι η επίθεση επιτυχής
- Ελαχιστοποίηση των αρνητικών επιρροών που προκύπτουν από τις τεχνικές ανίχνευσης επιθέσεων

Οι παραπάνω στόχοι για την αντιμετώπιση της κατάστασης και της αποτροπής δυσάρεστων αποτελεσμάτων προσεγγίζονται μέσω της προσπάθειας κατασκευής πιο εύρωστων εναλλακτικών μοντέλων από συνεργατικές μεθόδους, τη δημιουργία γειτονιάς με τη χρήση κατάλληλων ιδιοτήτων που να είναι λιγότερο ευάλωτες σε επιθέσεις και μέσω της προσπάθειας ανίχνευσης και καταπολέμησης των επιτιθέμενων πριν αυτοί εξαπολύσουν την επίθεση που σχεδιάζουν.

Από μελέτες που έχουν πραγματοποιηθεί μπορούν να εξαχθούν τα παρακάτω συμπεράσματα:

1. Όσο αυξάνεται ο αριθμός των κακόβουλων προφίλ που λαμβάνουν μέρος σε μία επίθεση τόσο αυξάνεται και ο αντίκτυπος της επίθεσης αλλά μέχρι ενός σημείου.

Μετά από αυτό το σημείο μεγαλώνει το κόστος επίθεσης για τον επιτιθέμενο χωρίς να λαμβάνει το αντίστοιχο κέρδος από τη χειραγώγηση του συστήματος.

2. Η ποιότητα των προβλέψεων είναι υποκειμενική υπόθεση καθώς επειδή ένα προϊόν προωθείται δεν σημαίνει ότι δεν είναι καλό ή ότι δεν θα άρεσε στους χρήστες. Έτσι τυχαίνει μετά την επίθεση να ανέβουν τα good predictions. Ουσιαστικά αυτό που βλέπουμε είναι η διαφοροποίηση μεταξύ του πριν και του μετά της επίθεσης.

5 Πειραματική προσέγγιση

5.1 Εισαγωγή

Στα προηγούμενα κεφάλαια παρουσιάσαμε τις έννοιες της Συνεργατικής Μεθόδου για την παραγωγή υποδείξεων, της εμπιστοσύνης και των κακόβουλων επιθέσεων. Ένα σύστημα Συνεργατικής Μεθόδου πρέπει να είναι ανοικτό στις απόψεις των χρηστών ακόμα κι αν είναι ακραίες ή ιδιαίτερες, ταυτόχρονα όμως οφείλει να βγει οδούς προστασίας από κακόβουλες επιθέσεις που ως στόχο μπορεί να έχουν είτε την ορθή και αποδοτική λειτουργία του συστήματος είτε την προτεινόμενη βαθμολογία συγκεκριμένων αντικειμένων.

Όπως έχει αναφερθεί δύο βασικές μετρικές αξιολόγησης των συστημάτων υπόδειξης είναι η ορθότητα και η ευρωστία. Οι δύο έννοιες είναι διακριτές και δεν ταυτίζονται αν και συχνά σχετίζονται. Έστω μια επίθεση που έχει ως στόχο την αύξηση της προτεινόμενης βαθμολογίας του αντικειμένου Α. Το αντικείμενο Α είναι ποιοτικό και μπορεί να ικανοποιήσει αποδοτικά της ανάγκες των χρηστών του. Μετά την επίθεση το αντικείμενο Α προτείνεται περισσότερο στους χρήστες και αξιολογείται από αυτούς θετικά. Έτσι οι υποδείξεις που αφορούν το Α είναι μεν ορθές αλλά δεν είναι εύρωστες καθώς προέρχονται από την επιτυχή επίθεση της κακόβουλης οντότητας. Αν το αντικείμενο Α ήταν χαμηλής ποιότητας τότε θα μείωνε την ορθότητα του συστήματος καθώς οι χρήστες δεν θα ήταν ικανοποιημένοι από αυτό και θα το αξιολογούσαν χαμηλά όμως και πάλι η επίθεση θα ήταν επιτυχής καθώς στόχος του επιτιθέμενου είναι η αύξηση των υποδείξεων του συγκεκριμένου αντικειμένου (μείωση ευρωστίας). Κατά συνέπεια μια επιτυχής επίθεση μπορεί να αυξήσει ή να μειώσει την ορθότητα των προβλέψεων ενώ σε κάθε περίπτωση μειώνει την ευρωστία.

Για να μπορέσουν να προστατευτούν τα συστήματα υπόδειξης από την επίθεση κακόβουλων οντοτήτων θα πρέπει είτε να μπορούν να αναγνωρίζουν τα κακόβουλα προφίλ είτε να μην τα λαμβάνουν υπόψιν κατά τη διάρκεια παραγωγής των υποδείξεων. Επιθυμία κάθε επιτυχημένου συστήματος υπόδειξης είναι η αύξηση της ορθότητας των προβλέψεων με ταυτόχρονη διατήρηση της ευρωστίας του συστήματος σε υψηλά επίπεδα. Η χρήση της εμπιστοσύνης (βάλε άρθρα) μπορεί να οδηγήσει σε αύξηση της ορθότητας των προβλέψεων παρά την παρουσία θορύβου στα δεδομένα ενώ οι O'Donovan et al υποστηρίζουν ότι αυξάνει και την ευρωστία.

Οι *O'Donovan & Smyth (2006)* ορίζουν την εμπιστοσύνη σε παγκόσμια εμβέλεια (global trust) ως τις επιτυχημένες υποδείξεις κάθε χρήστη στο δίκτυο προς το σύνολο των υποδείξεών του και χρησιμοποιείται στην παραγωγή υποδείξεων μόνο ως παράγοντας για τον υπολογισμό του συντελεστή βαρύτητας της βαθμολογίας κάθε γείτονα ενώ η ομοιότητα χρησιμοποιείται σε μεγαλύτερη κλίμακα. Δεδομένου ότι βρέθηκε μόνο η συγκεκριμένη εργασία στη βιβλιογραφία που συνδέει την εμπιστοσύνη με την ευρωστία και λόγω της χαμηλής κλίμακας χρήσης της έννοιας στην παραγωγή υποδείξεων ακολουθεί πειραματική προσέγγιση που κάνει εκτεταμένη χρήση της εμπιστοσύνης (ορισμένη σε τοπικό και παγκόσμιο επίπεδο) ώστε να αξιολογηθεί η προβλεπτική ικανότητά της ενώ το σύστημα έχει δεχθεί επίθεση (δηλαδή έχει εισαχθεί μη φυσικός θόρυβος στα δεδομένα). Η αξιολόγηση συνεπώς αφορά την εμπιστοσύνη σε όρους ορθότητας και ευρωστίας.

5.2 Δημιουργία υποδείξεων

Κατά τις προηγούμενες δεκαετίες οι προσπάθειες επικεντρώνονταν στη συλλογή πληροφοριών ώστε η λήψη αποφάσεων να στηριχθεί σε ολοκληρωμένα στοιχεία. Πλέον το πρόβλημα της λήψης αποφάσεων έχει μετασχηματιστεί. Η δυσκολία δεν έγκειται στην εύρεση πληροφοριών αλλά στην επιλογή εκείνων που καλύπτουν ιδανικά την ανάγκη των χρηστών. Η εξαντλητική αναζήτηση έχει πολύ υψηλό κόστος που τις περισσότερες φορές είναι απαγορευτικό. Τα εργαλεία εξόρυξης γνώσης είναι απαραίτητα και αποτελούν αναπόσπαστο κομμάτι πολλών ιστοχώρων. Τα συστήματα υπόδειξης και ιδίως τα συστήματα που ενσωματώνουν συνεργατικές μεθόδους (collaborative filtering) στόχο έχουν την ενίσχυση της διαδικασίας αξιολόγησης των εναλλακτικών για την κάλυψη των αναγκών τους. Οι συνεργατικές τεχνικές (CF) αναπαραγάγουν μία κοινή συμπεριφορά των ανθρώπων. Οι άνθρωποι για πράγματα που δεν ξέρουν ή που δεν είναι σίγουροι ζητούν πληροφορίες ή βοήθεια από άλλα άτομα. Έτσι και οι CF τεχνικές κάνουν προτάσεις ή υποδεικνύουν αντικείμενα στηριζόμενα στις γνώμες και τις αξιολογήσεις άλλων χρηστών.

Το πρώτο ερώτημα που προκύπτει είναι πώς επιλέγονται οι χρήστες που το σύστημα θα «ρωτήσει» τη γνώμη τους. Το πιο διαδεδομένο κριτήριο είναι η ομοιότητα των προφίλ. Δηλαδή έστω ένα ηλεκτρονικό κατάστημα πώλησης βιβλίων. Συχνά κάτω από την προβολή κάθε βιβλίου υπάρχουν προτάσεις για άλλα. Στο CF οι προτάσεις αυτές γίνονται βάσει των αξιολογήσεων άλλων χρηστών που μοιάζουν με τον τρέχοντα χρήστη. Για τον υπολογισμό της ομοιότητας μεταξύ των χρηστών χρησιμοποιείται εκτενέστερα το μέτρο Pearson (εξίσωση 35) το οποίο λαμβάνει υπόψη τα κοινά βαθμολογημένα αντικείμενα των προφίλ για τα οποία εξετάζεται η ομοιότητα.

$$sim_{u,v} = \frac{\sum_{i \in I}(r_{u,i}-\overline{r_u})(r_{v,i}-\overline{r_v})}{\sqrt{\sum_{i \in I}(r_{u,i}-\overline{r_u})^2}\sqrt{\sum_{i \in I}(r_{v,i}-\overline{r_v})^2}}$$

Εξίσωση 35

Όπου $r_{u,i}$ $r_{u,i}$ δηλώνει τη βαθμολογία του χρήστη u για το αντικείμενο i και $\overline{r_u}$ είναι ο μέσος όρος των βαθμολογιών που έχει δώσει ο χρήστης u για όλα τα αντικείμενα που έχει βαθμολογήσει. Αντίστοιχα $r_{v,i}$ είναι η βαθμολογία του χρήστη v για το αντικείμενο i και $\overline{r_v}$ είναι ο μέσος όρος των βαθμολογιών που έχει δώσει ο χρήστης u για όλα τα αντικείμενα που έχει βαθμολογήσει.

Αφού έχει υπολογισθεί η ομοιότητα μεταξύ χρηστών στη συνέχεια παράγονται οι προβλέψεις των βαθμολογιών για το χρήστη u που αφορούν αντικείμενα που ο ίδιος δεν έχει ακόμα δει (ουσιαστικά δεν έχει βαθμολογήσει). Ο κλασικός υπολογισμός προβλέψεων με την υλοποίηση συνεργατικής μεθόδου πραγματοποιείται με τη χρήση της εξίσωσης 2 όπου στη θέση του $w_{u,v}$ τοποθετείται η υπολογισθείσα ομοιότητα των χρηστών u και v

$$p_{u,i} = \overline{r_u} + \frac{\sum_{v \in V} w_{u,v}\,(r_{u,i}-\overline{r_u})}{\sum_{v \in V} |w_{u,v}|}$$

Εξίσωση 36

Όπου $p_{u,i}$ είναι η προτεινόμενη βαθμολογία για το αντικείμενο i προς το χρήστη u . V είναι το σύνολο των k όμοιων χρηστών που έχουν επιλεχθεί να δομήσουν τη γειτονιά του τρέχοντα χρήστη. Στους υπολογισμούς των προβλέψεων για κάθε αντικείμενο συμπεριλαμβάνονται μόνο όσοι γείτονες έχουν βαθμολογήσει το αντικείμενο αυτό. Το

$w_{u,v}$ εκτός της ομοιότητας που αποτελεί κλασική εφαρμογή μπορεί να επιλεχθεί οποιοδήποτε άλλο βάρος όπως η εμπιστοσύνη που θα εξηγηθεί στη συνέχεια (εξίσωση 37).

Για τον ορισμό του V χρησιμοποιείται η χρήση του αλγορίθμου k-NN όπου από το σύνολο των χρηστών επιλέγονται οι κορυφαίοι k βάσει με κριτήριο την τιμή της ομοιότητας. Συνήθως τίθεται $k = 25$ χωρίς όμως να αποτελεί απαράβατο κανόνα αυτή η επιλογή.

Τέλος για τη δόμηση της λίστας υποδείξεων κάθε χρήστη λαμβάνονται από το σύνολο των προβλέψεων (Εξίσωση 36) που παράγονται για αυτόν τα αντικείμενα με την υψηλότερη τιμή βαθμολόγησης.

5.3 Εμπιστοσύνη

5.3.1 Προβλέψεις και εμπιστοσύνη

Η χρήση της εμπιστοσύνης έχει αποδειχθεί ότι αυξάνει την αποτελεσματικότητα των συστημάτων υπόδειξης και κυρίως αυτών που χρησιμοποιούν συνεργατικές μεθόδους για την παραγωγή προτάσεων. Ένας χρήστης για να επιλεχθεί προς συμμετοχή από το σύστημα για πρόταση υποδείξεων δεν αρκεί να έχει υψηλή ομοιότητα με τον τρέχοντα χρήστη (δηλαδή τον χρήστη στον οποίο θα υποδειχθούν τα αντικείμενα) αλλά θα πρέπει να είναι και αξιόπιστος, δηλαδή να έχει ιστορικό αξιόλογων υποδείξεων.

Η ασυμμετρία της εμπιστοσύνης είναι επίσης σημαντική διάσταση καθώς αντανακλά ένα συγκεκριμένο τύπο εξατομίκευσης. Για δύο άτομα που εμπλέκονται σε μια σχέση, η εμπιστοσύνη δεν είναι απαραιτήτως αμφίδρομη και ισότιμη. Η ασυμμετρία της εμπιστοσύνης έχει μελετηθεί σε σημαντικό βαθμό αντίθετα με τη συμμετρία, δηλαδή τη χρήση της αμφίδρομης εμπιστοσύνης.

Έτσι η Εξίσωση 36 μπορεί να μετασχηματιστεί στην Εξίσωση 37 η οποία ως συντελεστή βάρους κάθε βαθμολογίας χρησιμοποιεί την εμπιστοσύνη κι όχι στην ομοιότητα.

$$p_{a,i} = \overline{r_a} + \frac{\sum_{u \in R^T} t_{a,u}\left(r_{u,i} - \overline{r_u}\right)}{\sum_{u \in R^T} t_{a,u}}$$

Εξίσωση 37

Η άγνωστη βαθμολογία $p_{a,i}$ για το αντικείμενο *i* και τον τρέχοντα χρήστη *α* υπολογίζεται με βάση τη μέση βαθμολογία $\overline{r_a}$ του χρήστη *α* των αντικειμένων που έχει ήδη

βαθμολογήσει καθώς και για την βαθμολογία $r_{u,i}$ από τους άλλους χρήστες για *u* για το αντικείμενο *i*. Η εξίσωση λαμβάνει επίσης υπόψη την τιμή εμπιστοσύνης $t_{a,u}$. Η στρατηγική αυτή υποστηρίζεται επίσης από το γεγονός ότι η εμπιστοσύνη και ομοιότητας συσχετίζονται *(Massa & Avesani (2007), Golbeck thesis)*.

5.3.2 Υπολογισμός έμμεσης εμπιστοσύνης

Η εμπιστοσύνη αφορά το πόσο αξιόλογο θεωρεί ένας χρήστης έναν άλλο και ένας από τους παράγοντες διαφοροποίησης του ορισμού της είναι το είναι άμεση ή έμμεση. Η εμπιστοσύνη βάσει της βιβλιογραφίας μπορεί να λάβει τιμή με δύο τρόπους:

1. Με ρητή δήλωση. Δηλαδή ο χρήστης Α ερωτάται από το σύστημα «αν εμπιστεύεται το χρήστη Β». Dataset που να παρέχει τέτοια πληροφορία είναι μόνο το epinions. Οι τιμές εμπιστοσύνης στη συγκεκριμένη βάση μπορεί να είναι 0 (δεν εμπιστεύομαι) και 1 (εμπιστεύομαι). Η απουσία τιμής εμπιστοσύνης μεταξύ δύο χρηστών δεν αποτελεί ένδειξη «έλλειψης εμπιστοσύνης». Παρόλα αυτά για δημόσια χρήση δεν δίνονται οι τιμές μη εμπιστοσύνης 0 που αναφέρονται και ως blacklist.

2. Η εμπιστοσύνη μπορεί να υπολογισθεί εμμέσως βάσει των κοινών βαθμολογιών που έχουν δώσει για αντικείμενα δύο χρήστες.

Συγκεκριμένα *O'Donovan & Smyth (2006)* την υπολογίζουν έμμεσα ως το πηλίκο των σωστών προβλέψεων ενός χρήστη μέσα στο δίκτυο προς το σύνολο των προβλέψεων κάθε χρήστη σε αυτό (παγκόσμιας εμβέλειας τιμή, δηλαδή σε κάθε χρήστη αντιστοιχεί μία τιμή εμπιστοσύνης) και υπολογίζεται βηματικά ακολούθως:

$$Correct(i,p,c) \Leftrightarrow |p(i) - c(i)| < e$$

Εξίσωση 38

Όπου $p(i)$ είναι η βαθμολογία που προέκυψε από τη χρήση της εξίσωσης 37 για το αντικείμενο i και αφορά το χρήστη c. Στη γειτονιά που χρησιμοποιήθηκε για τον υπολογισμό της βαθμολογίας για το χρήστη c συμπεριλαμβάνεται και ο χρήστης p. Αν η διαφορά πρόβλεψης $p(i)$ μείων την πραγματική βαθμολογία του χρήστη $c, c(i)$ είναι μικρότερη της τιμής e η οποία ορίζεται εμπειρικά τότε η πρόβλεψη θεωρείται ορθή.

$$T_p(i.c) = Correct(i, p, c)$$

Εξίσωση 39

Η εξίσωση 39 αφορά το σύνολο των προβλέψεων (ανεξαρτήτως ορθότητας) για τους χρήστες c του δικτύου στις οποίες συμμετείχε ο p .

$$RecSet(p) = \{(c_1, i_1), \dots, (c_n, i_n)\}$$

Εξίσωση 40

Ενώ η εξίσωση 40 αφορά το σύνολο των ορθών προβλέψεων για τους χρήστες c του δικτύου στις οποίες συμμετείχε ο χρήστης p .

$$CorrectSet(p) = \{(c_k, i_k) \in RecSet(p) : Correct(i_k, p, c_k)\}$$

Εξίσωση 41

Η εξίσωση 41 υπολογίζει την εμπιστοσύνη που απολαμβάνει ο χρήστης p από το δίκτυο ως πηλίκο των ορθών προβλέψεων του προς το σύνολο των προβλέψεων.

$$Trust^P(p) = \frac{|CorrectSet(p)|}{|RecSet(p)|}$$

Εξίσωση 42

Τέλος βάσει της εξίσωσης 42 μετασχηματίζεται η εξίσωση 35 χρησιμοποιώντας ως βάρος το συνδυασμό ομοιότητας και εμμέσως υπολογιζόμενης εμπιστοσύνης .

$$w_{(c,p,i)} = \frac{2\left(sim_{(c,p)}\right)\left(\left(trust^I_{(p,i)}\right)\right)}{sim_{(c,p)} + \left(trust^I_{(p,i)}\right)}$$

Εξίσωση 43

Δηλαδή προκύπτουν προβλέψεις όπου οι γείτονες επιλέγονται με βάσει την ομοιότητα similarity (Εξίσωση 1) και έχουν κατώφλι την τιμή 0.4 (δηλαδή μόνοι όσοι έχουν ομοιότητα μεγαλύτερη του 0,4 μπορούν να συγκροτήσουν τη γειτονιά του τρέχοντα χρήστη) ενώ ο συντελεστής βάρους για κάθε βαθμολογία αποτελεί ένα συνδυασμό ομοιότητας και εμπιστοσύνης. Δηλαδή το βάρος βασίζεται σημαντικά στο μέτρο της ομοιότητας για τον υπολογισμό του, δημιουργώντας έτσι το ερώτημα αν η εμπιστοσύνη μπορεί να θεωρείται

εύρωστη μετρική μόνο μέσω της μερικής συμμετοχής της στη φόρμουλα υπολογισμού (Εξίσωση 9).

5.4 Κακόβουλες πρακτικές

Αδυναμία των συνεργατικών μεθόδων αποτελεί ο ανοικτός χαρακτήρας τους. Δηλαδή το γεγονός ότι οι υποδείξεις στηρίζονται σε γνώμες χρηστών των οποίων οι βαθμολογίες δεν μπορούν να προεξοφληθούν ως αντικειμενικές. Αυτή τη διαπίστωση προσπαθούν να εκμεταλλευτούν κακόβουλοι χρήστες και εξαπολύουν επιθέσεις έναντι του συστήματος. Σκοπός τους μπορεί να είναι:

- Το σύστημα και η εμπόδιση της εύρυθμης λειτουργίας του
- Η ενίσχυση (push) ή η αποδυνάμωση (nuke) ενός συγκεκριμένου αντικειμένου.

Τις περισσότερες φορές στόχος είναι ο επηρεασμός της βαθμολογίες πρόβλεψης για ένα αντικείμενο παρά όλο το σύστημα. Είναι γνωστό ότι τα συστήματα υπόδειξης παρακινούν και ωθούν σε απόκτηση αγαθών οπότε συχνά οι παραγωγοί θεωρούν ότι μέσω μίας επίθεσης προώθησης αντικειμένου μπορούν να επιτύχουν υψηλότερες πωλήσεις.

5.4.1 Μοντέλα επίθεσης

Τα μοντέλα επίθεσης που θα χρησιμοποιηθούν στο πειραματικό σκέλος της εργασίας είναι η τυχαία επίθεση και η επίθεση μέσου όρου. Η υλοποίησή του βασίστηκε στην περιγραφή της στρατηγικής της κάθε επίθεσης όπως περιγράφονται από τον Burke (2002).

Τα παραπάνω μοντέλα επίθεσης αποτελούν τα πιο συνήθη και αναλύονται στη συνέχεια:

Τυχαία Επίθεση (Random Attack)

Ο αριθμός των αντικειμένων που επιλέγονται για τα προφίλ επίθεσης καθώς και τα αντικείμενα και οι βαθμολογίες τους επιλέγονται τυχαία.

Στην τυχαία επίθεση στόχος είναι η μείωση της συνολικής απόδοσης του συστήματος. Αυτής της κατηγορίας οι επιθέσεις δεν εστιάζουν σε συγκεκριμένους χρήστες ή συγκεκριμένα αντικείμενα αλλά στοχεύουν σε όλους τους χρήστες και τα αντικείμενα ισότιμα σε μία προσπάθεια να περιορίσει τη γενική ορθότητα του συστήματος. Ως ένα

πραγματικό σενάριο θεωρούμε το διαχειριστή ενός συστήματος που είναι ανταγωνιστικός και θέλει να «χτυπήσει» ένα αντίπαλο σύστημα για να προσελκύσει περισσότερους πελάτες. Η στρατηγική επίθεσης είναι καθαρή, ο αριθμός των αντικειμένων που επιλέγονται για τα προφίλ επίθεσης καθώς και τα αντικείμενα και οι βαθμολογίες τους επιλέγονται τυχαία.

Επίθεση Μέσου Όρου (Average attack)

Τα αντικείμενα επιλέγονται τυχαία και σε κάθε ένα από αυτά τίθεται τιμή ίση με το μέσο του αντικειμένου όπως αυτό βαθμολογήθηκε από τους χρήστες του συστήματος.

Σε αυτού του είδους τις επιθέσεις το σύνολο των επιλεγμένων αντικειμένων είναι κενό. Τα αντικείμενα πλήρωσης επιλέγονται τυχαία και σε κάθε ένα από αυτά τίθεται τιμή ίση με το μέσο του αντικειμένου όπως αυτό βαθμολογήθηκε από τους χρήστες του συστήματος. Η αποτελεσματικότητα των συγκεκριμένων επιθέσεων είναι από τις υψηλότερες. Για να οργανωθεί μία average επίθεση είναι απαραίτητη υψηλή γνώση των ιδιοτήτων του συστήματος καθώς χρειάζεται ο μέσος όρος όλων των αντικειμένων πλήρωσης (filler items). Αν και υψηλότερη πιθανότητα να είναι μεγαλύτερη ομοιότητα με περισσότερους χρήστες.

5.5 Μετρικές αξιολόγησης

Στη συνέχεια παρουσιάζονται τρεις από τις πιο χρησιμοποιούμενες και αναγνωρισμένες μετρικές αξιολόγησης των συστημάτων υπόδειξης. Οι μετρικές αυτές θα χρησιμοποιηθούν για την ανάλυση των αποτελεσμάτων της πειραματικής μας προσέγγισης.

Ορθότητα προβλέψεων (accuracy)

Η πλειοψηφία των συστημάτων υπόδειξης βασίζονται σε μια μηχανή παραγωγής προβλέψεων, η οποία προβλέπει τις απόψεις των χρηστών σε σχέση με τα διάφορα αντικείμενα του συστήματος (βαθμολογίες), ή η πιθανότητα χρήσης των αντικειμένων (αγορά). Προϋπόθεση ενός πετυχημένου συστήματος υπόδειξης είναι η παροχή υποδείξεων υψηλής ορθότητας. Δηλαδή τα αντικείμενα που προτείνονται στους χρήστες να τυγχάνουν αποδοχής.

Mean Absolute Error (MAE)

Κλασική μέθοδος για τον υπολογισμό της ορθότητας των προβλέψεων είναι το μέσο απόλυτο λάθος *Mean Absolute Error (MAE)*:

$$MAE = \frac{\sum_{(u,i)} |p_{u,i} - r_{u,i}|}{N}$$

Εξίσωση 44

Όπου $\mathbf{P_{u,i}}$ είναι η πρόβλεψη βαθμολογίας του αντικειμένου i για το χρήστη u ενώ $r_{u,i}$ είναι η βαθμολογία που έχει δώσει. Τέλος N το σύνολο των χρηστών για τους οποιους εκτελέστηκε η διαδικασία πρόβλεψης.

Mean Absolute User Error (MAUE)

Κατά κύριο λόγο η Συνεργατική Μέθοδος λειτουργεί καλά για χρήστες που έχουν ήδη αξιολογήσει αρκετά αντικείμενα ενώ λιγότερο καλά για όσους δεν έχουν αξιολογήσει περιορισμένο αριθμό. Αποτέλεσμα αυτού του γεγονότος είναι οι χρήστες με πολλά αξιολογημένα αντικείμενα να έχουν μικρό μέσο λάθος προβλέψεων ενώ όσοι έχουν αξιολογήσει λίγα να έχουν μεγάλο μέσο λάθος (MAE). Αν έχουμε ένα χρήστη που έχει αξιολογήσει 100 αντικείμενα και 5 άλλους που έχουν αξιολογήσει πολύ λίγα θα έχουμε έναν ευχαριστημένο χρήστη και 5 δυσαρεστημένους αν και η τιμή το μέσου λάθος θα έχει μειωθεί λόγω των πολλών αξιολογήσεων του πρώτου χρηστή. Αυτό φυσικά δεν αντιπροσωπεύει την πραγματική κατάσταση, για αυτό οι *Massa & Aventi (2004)* πρότειναν τη μετρική MAUE (Mean Absolute User Error) η οποία υπολογίζεται ξεχωριστά για κάθε χρήστη και στη συνέχεια υπολογίζεται ο μέσος όρος αυτών. Έτσι κάθε χρήστης έχει το ίδιο συντελεστή βάρους στον υπολογισμό του μέσου λάθους.

Ευρωστία (robustness)

Η ευρωστία *(Groot et all 2000)* αποτελεί το βαθμό κατά τον οποίο ένα σύστημα ή ένα υποσύστημα λειτουργεί ορθά παρουσία θορύβου στα δεδομένα ή κάτω από πιεστικές περιβαλλοντολογικές συνθήκες.

Prediction Shift

Για τη μέτρηση της αποτελεσματικότητας μίας επίθεσης ένα μέτρο που χρησιμοποιείται ευρύτατα είναι η μεταβολή πρόβλεψης (Prediction Shift). Η μεταβολή πρόβλεψης για το αντικείμενο στόχος είναι η διαφορά της μέσης προβλεπόμενης βαθμολογίας του

αντικειμένου πριν και μετά την επίθεση για όλους τους χρήστες. Ως μέση μεταβολή πρόβλεψης ονομάζεται η μέση αλλαγή πρόβλεψης πριν και μετά την επίθεση για όλα τα αντικείμενα στόχους. Βάσει του *(Dellarocas, 2000).* ορίζεται:

Έστω U το σύνολο των χρηστών , I το σύνολο των αντικειμένων του συστήματος και $\Delta_{u,i}$ η μεταβολή πρόβλεψης του χρήστη u για το αντικείμενο i .Δηλαδή

όπου είναι η τιμή πρόβλεψης μετά την επίθεση και $p_{u,i}$ η πρόβλεψη πριν την επίθεση. Η μέση μεταβολή πρόβλεψης υπολογίζεται ως εξής:

$$\Delta_i = \frac{\sum_{u \in U} \Delta_{u,i}}{|U|}$$

Εξίσωση 45

Η πρόβλεψη μεταβολής ενός μοντέλου είναι η μέση πρόβλεψη μεταβολής για όλα τα αντικείμενα στόχους και υπολογίζεται ως εξής:

$$\Delta = \frac{\sum_{i \in I} \Delta_i}{|I|}$$

Εξίσωση 46

5.6 Η προσέγγιση της εργασίας

Σκοπός της εργασίας είναι η αξιολόγηση της εμπιστοσύνης ως μετρική ευρωστίας. Δηλαδή αν μπορεί να χρησιμοποιηθεί στα πλαίσια της συνεργατικής μεθόδου για την προστασία του συστήματος έναντι των επιθέσεων. Στα πλαίσια της πειραματικής προσέγγισης θα χρησιμοποιηθεί η εμπιστοσύνη σε δύο διαφορετικές θέσεις, ως συντελεστής βάρους κατά την παραγωγή των υποδείξεων (Εξίσωση 1) και ως κριτήριο για την επιλογή γειτονιάς στον αλγόριθμο k-NN . Η εμπιστοσύνη διερευνήθηκε ορισμένη έμμεσα, δηλαδή η τιμή που απολαμβάνει κάθε χρήστης υπολογίζεται βάσει των προβλέψεων που έχει κάνει στο παρελθόν και δεν προέρχεται από ρητή κρίση των υπολοίπων χρηστών. Ο ορισμός στηρίζεται στους *O' Donovan et al* αλλά με κάποιες διαφοροποιήσεις που σκοπό έχουν τη μείωση της εξάρτησής της από τη μετρική της ομοιότητας ώστε να στηρίζεται το κατά δυνατόν λιγότερο στην έννοια της ομοιότητας.

Διαφοροποιώντας την μετρική των *O' Donovan & Smyth (2005)* διερευνήθηκαν μετρικές σε τοπικό/παγκόσμιο επίπεδο και αν υπάρχει κέρδος όταν η εμπιστοσύνη είναι αμφίδρομη

(δηλαδή δύο κόμβοι -χρήστες- του συστήματος να έχουν εισερχόμενη και εξερχόμενη ακμή -τιμή- εμπιστοσύνης) ενώ τέλος προτάσσεται μια νέα εμπιστοσύνη για την οποία δεν αρκεί ένας χρήστης να συμμετάσχει σε καλές προβλέψεις αλλά και να έχει προσφέρει θετικά σε αυτές. Οι διαφορετικές μετρικές εμπιστοσύνης αναλύονται στη συνέχεια.

5.6.1 Μετρικές εμπιστοσύνης

localTrust

Μεταβλητή τοπικής εμβέλειας δηλαδή κάθε χρήστης λαμβάνει από διαφορετικούς χρήστες διαφορετική τιμή εμπιστοσύνης που υπολογίζεται ως το άθροισμα των επιτυχημένων προτάσεων που συμμετείχε ο συγκεκριμένος χρήστης (τρέχοντας) προς το σύνολο των προτάσεων έναντι κάθε ενεργού χρήστη. Αποτελεί παραλλαγή των *O' Donovan et al (2005)*.

localContributorTrust

Μεταβλητή τοπικής εμβέλειας παρόμοια με την localTrust με τη διαφοροποίηση ότι δεν αρκεί να έχει συμμετάσχει ο χρήστης σε καλές προβλέψεις αλλά θα πρέπει η συνεισφορά του σε αυτές να είναι θετική. Δηλαδή δεν ανήκει σε γειτονιές που παράγουν ορθές προβλέψεις αλλά θα πρέπει να έχει δώσει για το τρέχον αντικείμενο βαθμολογία που να είναι κοντά σε αυτή του χρήστη για τον οποίο προορίζεται η υπόδειξη. Το όριο του «κοντά» έχει ορισθεί σε e<1.

neighbourTrust

Μεταβλητή τοπικής εμβέλειας , υπολογίζεται σε κάθε γειτονιά που συμμετέχει ένας χρήστης και η τιμή της προέρχεται από το μέσο όρο των βαθμολογιών εμπιστοσύνης που έχει λάβει από κάθε άλλο γείτονα και τον τρέχοντα χρήστη (το χρήστη για τον οποίο έχει δομηθεί η γειτονιά).

globalTrust

Μεταβλητή καθολικής εμβέλειας, κάθε χρήστης λαμβάνει μία τιμή εμπιστοσύνης για τη συνολική συνεισφορά του στο δίκτυο. Ακολουθεί πλήρως τη λογική των *O΄ Donovan et al (2005)*

globalTrustContributor

Μεταβλητή καθολικής εμβέλειας, δηλαδή κάθε χρήστης λαμβάνει μία τιμή εμπιστοσύνης για τη συνολική συνεισφορά του στο δίκτυο. Βασίζεται κι αυτή στους *O΄ Donovan et al (2005)* με τη διαφορά ότι δεν αρκεί να έχει συμμετάσχει σε καλές γειτονιές αλλά και να έχει

συνεισφέρει θετικά στην ορθή πρόβλεψη, δηλαδή η βαθμολογία που έχει συνεισφέρει παραγωγή πρόβλεψης για το συγκεκριμένο αντικείμενο να είναι κοντά στη βαθμολογία του τρέχοντα χρήστη. Το κοντά ορίζεται κι εδώ στο <1, δηλαδή ένας γείτονας θεωρείται ότι έχει επιτυχημένη υπόδειξη όταν η υπόδειξη που συμμετέχει είναι κοντά σε εκείνη του τρέχοντα χρήστη και η συνεισφορά του το ίδιο. Σε αυτό το σημείο να σημειωθεί ότι οι βαθμολογίες έχουν γίνει με αναγωγή ώστε να μην παίζει ρόλο αν ο συγκεκριμένος χρήστης έχει την τάση να βαθμολογεί υψηλά ή χαμηλά. Δηλαδή η βαθμολογία του είναι η βαθμολογία μείων το μέσο όρο των βαθμολογιών του. Δηλαδή αθροίζονται όλες ορθές προβλέψεις που συμμετάσχει κάθε χρήστης στο δίκτυο και διαιρούνται με το σύνολο των προβλέψεων που έχει συμμετάσχει. Με κάθε νέα συμμετοχή σε πρόβλεψη αλλάζει και η τιμή εμπιστοσύνης του δικτύου προς το συγκεκριμένο χρήστη.

mutualTrust

Στη βιβλιογραφία έχει μελετηθεί μόνο για τη μονόδρομη μορφή της , ενώ από την *Golbeck (2005)* έχει συζητηθεί ότι η εμπιστοσύνη μπορεί να είναι μονόδρομη ή και αμφίδρομη. Δηλαδή έστω τους κόμβους Α και Β. Αν ο Α εμπιστεύεται τον Β θα υπάρχει ακμή που να συνδέει τους δύο κόμβους με προσανατολισμό από τον Α στο Β, ενώ η ακμή θα έχει τιμή που θα αντιπροσωπεύει την τιμή εμπιστοσύνης του Α προς τον Β. Η τιμή της ακμής είναι τοπικής εμβέλειας καθώς δείχνει την τιμή του Α προς τον Β και δεν αντιπροσωπεύει την εμπιστοσύνη που δείχνει το δίκτυο προς τον Β. Επίσης η ύπαρξη της συγκεκριμένης σύνδεσης δεν σημαίνει ότι θα πρέπει να υπάρχει και αντίστοιχη ακμή από τον Β στον Α και ούτε είναι απαραίτητο η τιμή της να συμβαδίζει με την τιμή του Α προς τον Β. Η τιμή της εμπιστοσύνης μπορεί να είναι έμμεση ή άμεση.

Σε μια ανθρώπινη επαφή η αμοιβαιότητα ή έστω η αμφίδρομη σχέση εμπιστοσύνης παίζει σημαντικό ρόλο στη εξέλιξη της ενώ στα πλαίσια των συστημάτων υπόδειξης δεν έχει διερευνηθεί καθόλου. Για αυτό θα ορίσουμε την αμοιβαία εμπιστοσύνη (***mutualTrust)*** το μέσο όρο της εμπιστοσύνης που αναπτύσσεται μεταξύ των κόμβων Α και Β (σε περίπτωση που υπάρχει εισερχόμενη και εξερχόμενη ακμή σύνδεσης). Δηλαδή ***mutualTrust=*** $\left(\frac{t_v + t_u}{2}\right)$.

5.6.2 Περιγραφή βημάτων για την παραγωγή υποδείξεων

Συνοπτικά τα βήματα του αλγορίθμου είναι τα εξής:

1. Υπολογισμός της ομοιότητας (εξίσωση 1)

   Υπολογισμός της ομοιότητας μεταξύ όλων των χρηστών βάσει των βαθμολογιών που έχουν δώσει σε ίδια αντικείμενα και αποθήκευση αποτελέσματος. Για κάθε ζευγάρι χρηστών αποθηκεύεται μία τιμή, ενώ για όσους χρήστες δεν υπάρχει αποθηκευμένη εγγραφή στη βάση δεδομένων σημαίνει ότι δεν έχουν μέχρι στιγμής κοινά βαθμολογημένα αντικείμενα.

2. Υπολογισμός εμπιστοσύνη μεταξύ των χρηστών

   Χρησιμοποιώντας την εμπιστοσύνης που απολαμβάνει κάθε χρήστης στο δίκτυο βάσει των επιτυχημένων προβλέψεων που έχει συμμετάσχει στο παρελθόν. Στα πλαίσια της εργασίας έχουν υπολογισθεί διαφορετικές παραλλαγές της έννοιας ώστε με σκοπό την εξαγωγή συμπερασμάτων που σχετίζονται με τα χαρακτηριστικά της (πχ τοπική/ παγκόσμια μεταβλητή)

3. Εύρεση γειτονιάς με τον k-NN βάσει της ομοιότητας και βάσει της εμπιστοσύνης

   Ουσιαστικά λαμβάνονται από το σύνολο του βήματος 1 οι k χρήστες με την υψηλότερη ομοιότητα. Πειραματισμός με k, βιβλιογραφικά χρησιμοποιείται συχνά k=25 ή k=50. Στα πλαίσια της συγκεκριμένης εργασίας επιλέχθηκε γειτονιά 25 ατόμων.

   Σχηματίστηκαν 2 διαφορετικές γειτονιές:

   a. ***Την ομοιότητα (similarity)***

   b. ***Την τοπική εμπιστοσύνη (localContributorTrust)***

4. Υπολογισμός των προβλέψεων (εξίσωση 3)

   Οι υποδείξεις αφορούν αντικείμενα που ο τρέχων χρήστης δεν έχει βαθμολογήσει με συντελεστή βάρους:

a. ***Την τοπική εμπιστοσύνη (localTrust)***

b. ***Την τοπική εμπιστοσύνη (localContributorTrust)***

c. ***Την τοπική εμπιστοσύνη που υπολογίζεται από τους γείτονες (neighbourTrust***)

d. ***Την παγκόσμια εμπιστοσύνη (globalTrust)***

e. ***Την συμμετρική εμπιστοσύνη (mutualTrust)***

5.6.3 Διαστάσεις αξιολόγησης αλγορίθμου

- Ορθότητα (accuracy)

Χρήση MAE και MAUE για σύγκριση αποτελεσμάτων 5a και 5b προς την ορθότητα για να δούμε αν και κατά πόσο αποτελεσματική είναι η συμμετρική εμπιστοσύνη.

- Ευρωστία (robustness)

Για την ευρωστία θα χρησιμοποιήσουμε την μετρική Prediction Shift για να ελεγχθεί κατά πόσο οι επιθέσεις είναι αποτελεσματικές με τον αλγόριθμο που χρησιμοποιεί την άμεση εμπιστοσύνη 5a και την συμμετρική 5b.

5.6.4 Περιγραφή επιθέσεων

Στο σύστημα θα γίνουν δύο διαφορετικές επιθέσεις διαφορετικών μοντέλων Random Attack και Average attack *Burke (2002)* με σκοπό να επηρεάσουν την ομαλή λειτουργία του. Τόσο στην τυχαία επίθεση όσο και στην μέσης τιμής ο αριθμός των κακόβουλων προφίλ αφορά το 5% των χρηστών και δεν επιλέχθηκε τυχαία. Βάσει της βιβλιογραφίας αποτελεί ένα ποσοστό ικανό να θεωρηθεί μεν επικίνδυνο για το σύστημα χωρίς όμως να είναι προφανής η ύπαρξή τους.

Στο σύστημά μας υπάρχουν 300 γνήσιοι χρήστες και 15 κακόβουλοι κάθε επίθεσης. Το μέγεθος του κάθε κακόβουλου προφίλ επιλέχθηκε να είναι ίσο με το μέγεθος του μέσου προφίλ των καλόβουλων χρηστών για να μην ξεχωρίζει στο dataset. Δηλαδή κάθε κακόβουλο προφίλ έχει βαθμολογήσει 20 διαφορετικά αντικείμενα από αυτά που υπάρχουν στη βάση.

Η εισαγωγή των κακόβουλων προφίλ σε ένα πραγματικό σύστημα γίνεται μέσω της διεπαφής , δηλαδή η κακόβουλη οντότητα εγγράφεται πολλαπλώς στο σύστημα και βαθμολογεί τα αντικείμενα που επιθυμεί. Στα πλαίσια των ερευνητικών εργασιών η εισαγωγή των κακόβουλων βαθμολογιών γίνεται κατευθείαν στη βάση δεδομένων.

Στην Τυχαία επίθεση η επιλογή των αντικειμένων και η επιλογή της βαθμολογίας γίνεται τυχαία ενώ στην επίθεση Μέσου Όρου η επιλογή των αντικειμένων γίνεται τυχαία αλλά η επιλογή της βαθμολόγησης είναι ίση με τη μέση βαθμολογία του αντικειμένου στη βάση και βαθμολόγηση του αντικειμένου στόχου με την τιμή που επιλέγει ο επιτιθέμενος. Στα πλαίσια του πειράματος το αντικείμενο στόχος είναι το αντικείμενο 1. Και για τις 2 περιπτώσεις χρησιμοποιήθηκε τυχαία γεννήτρια δεδομένων όπως αυτή υλοποιείται από την Java 1.8.

Η αξιολόγηση των συστήματος για την τυχαία επίθεση πραγματοποιείται με τη χρήση του μέσου απόλυτου λάθους (MAE) και του μέσου απόλυτου λάθους ανά χρήστη (MAUE). Αντίστοιχα για την επίθεση μέσου όρου πραγματοποιούμε αξιολόγηση με τα μέτρα MAE και MAUE όπως και στην τυχαία αλλά επειδή έχουμε στην μέσου όρου αντικείμενο στόχο μπορούμε να χρησιμοποιήσουμε και το μέτρο του Prediction Shift (μέτρο ευρωστίας) για να δούμε το βαθμό επιτυχίας μείωσης ή αύξησης της βαθμολογίας του αντικειμένου στόχου. Κάτι τέτοιο στις τυχαίες επιθέσεις δεν μπορεί να πραγματοποιηθεί καθώς δεν έχουν συγκεκριμένα αντικείμενα που υπάρχουν σε κάθε κακόβουλο προφίλ για να δούμε το βαθμό επηρεασμού των αντικειμένων αυτών μετά την προσθήκη των κακόβουλων προφίλ στο σύστημα.

5.7 Περιγραφή συστήματος

Στην ενότητα αυτή θα περιγραφούν τα δομικά στοιχεία του συστήματος που υλοποιήθηκαν στα πλαίσια της εργασίας.

5.7.1 Βάση δεδομένων

Το dataset πάνω στο οποίο στηρίχθηκε η πειραματική αξιολόγηση προέρχεται από το epinions.com. Το epinions είναι ένας ιστοχώρος [3]στον οποίο οι χρήστες μπορούν να αξιολογήσουν διαφόρων κατηγοριών προϊόντα όπως ηλεκτρονικές συσκευές, υλικό υπολογιστών, ρολόγια κά. Σύμφωνα με την περιγραφή που είναι αναρτημένη στο site «το Epinions βοηθά τους ανθρώπους να κάνουν συνειδητές αποφάσεις αγοράς. Πρόκειται για

[3] www.epinions.com

μια κορυφαία πλατφόρμα που περιλαμβάνει αξιολογήσεις καταναλωτών για το Web και μια αξιόπιστη πηγή για την πολύτιμη γνώση των καταναλωτικών αναγκών, προσφέρει αμερόληπτες συμβουλές, σε βάθος αξιολογήσεις των προϊόντων και εξατομικευμένες συστάσεις.»

Το dataset από το συγκεκριμένο σύστημα έχει χρησιμοποιηθεί σε πάρα πολλές αντίστοιχες εργασίες όπου γίνονται πειραματικές προσεγγίσεις για τη βελτίωση των υποδείξεων. Η κλίμακα της βαθμολογίας είναι από 1 έως 5, με το 5 να αποτελεί την υψηλότερη βαθμολογία που μπορεί ένας χρήστης να αξιολογήσει ένα αντικείμενο. Οι ερευνητικές προσπάθειες σχετίζονται με της ορθότητας των προβλέψεων, την κάλυψης και την ευρωστία και αποτελεί μια κλασική επιλογή για πειράματα του συγκεκριμένου χώρου.

Ενδιαφέρον χαρακτηριστικό του συγκεκριμένου συστήματος αποτελεί το γεγονός ότι οι χρήστες μπορούν να δηλώσουν ρητά την εμπιστοσύνη τους προς τους άλλους χρήστες δημιουργώντας έτσι ο καθένας το δικό του δίκτυο εμπιστοσύνης (web of trust) ενώ παράλληλα μπορούν να δηλώσουν τους χρήστες που δεν θεωρούν έμπιστους τοποθετώντας τους σε λίστα αποκλεισμού (*blocklist*). Για αυτό το λόγο έχει σχετιστεί πλήρως με πειράματα που εμπλέκουν την άμεση εμπιστοσύνη (δηλαδή τη ρητή δήλωση εμπιστοσύνης ενός χρήστη από έναν άλλο ή το αντίθετο –έλλειψη εμπιστοσύνης-).

Βάσει των πληροφοριών που είναι αναρτημένες στο epinions.com το συνολικό dataset συγκροτείται από 139.738 διαφορετικά αντικείμενα, 49.290 χρήστες που έχουν υποβάλλει 664.824 κριτικές και 487.181 δηλώσεις εμπιστοσύνης. Το dataset του epinions αποτελείται από δύο βασικά αρχεία. Το ένα αφορά τις βαθμολογίες και το άλλο τις δηλώσεις εμπιστοσύνης. Στα πλαίσια της εργασίας θα χρησιμοποιηθεί το πρώτο αρχείο που περιλαμβάνει τις βαθμολογίες που έχουν δηλώσει οι χρήστες για συγκεκριμένα αντικείμενα. Το αρχείο αυτό ονομάζεται *Rating data* : και κάθε γραμμή του αρχείου έχει την ακόλουθη μορφή: user_id item_id rating_value. Παραδείγματος χάριν «23 387 5» το οποίο σημαίνει ότι ο χρήστης με κωδικό 23 έχει βαθμολογήσει το αντικείμενο 387 με βαθμό 5.

5.7.2 Πίνακες βάσης δεδομένων

Για την καλύτερη κατανόηση των βημάτων θα χρησιμοποιήσουμε τα βήματα της ενότητας 1.6.2 και θα τα αντιστοιχίσουμε σε πίνακες της βάσης.

Όπως ειπώθηκε, το dataset που χρησιμοποιήθηκε για τα πειράματα είναι το Epinions, όμως λόγω του μεγάλου όγκου πληροφορίας αναγκαστήκαμε να μειώσουμε το μέγεθος πάνω στο οποίο εκτελέστηκαν τα πειράματα. Για να αντιμετωπίσουμε τα προβλήματα που προκύπτουν από την έλλειψη επεξεργαστικής ισχύος και διαθέσιμης μνήμης του διαθέσιμου υπολογιστικού συστήματος, αποφασίστηκε να μειωθεί ο αριθμός των βαθμολογιών και χρηστών που το αποτελούν.

Συγκεκριμένα κάθε πίνακας βαθμολογιών που χρησιμοποιήθηκε από το Epinions αφορούσε τους 300 πρώτους χρήστες και 300 διαφορετικά αντικείμενα, δηλαδή περίπου 5000 εγγραφές.

Γνήσιοι Χρήστες

1. Υπολογισμός της ομοιότητας (εξίσωση 1)

Υπολογισμός της ομοιότητας μεταξύ όλων των χρηστών βάσει των βαθμολογιών που έχουν δώσει σε ίδια αντικείμενα. Για το σκοπό αυτό χρησιμοποιούμε τον πίνακα «ratingsdensetocalculate» με 5253 εγγραφές ο ποίος αποτελεί υποσύνολο του epinions. Αφορά βαθμολογίες των 299 πρώτων χρηστών (κωδικοί 1 έως και 299) και περιέχει βαθμολογίες αντικειμένων με κωδικούς από 231 έως και 1047.

Ο πίνακας έχει την παρακάτω δομή:

Πεδίο	**Τύπος**	**Περιγραφή**
user_id	int(3)	Ο κωδικός του χρήστη που βαθμολογεί
item_id	int(4)	Ο κωδικός του αντικειμένου που βαθμολογείται
ratings	int(1)	Η βαθμολογία που λαμβάνει το αντικείμενο από το χρήστη

Πίνακας 7 Δομή πίνακα ratingsdensetocalculate

Για κάθε ζευγάρι χρήστη – αντικειμένου αποθηκεύεται μία τιμή και ο συνδυασμός τους αποτελεί το πρωτεύον κλειδί κάθε εγγραφής του πίνακα.

Ο παραπάνω πίνακας χρησιμοποιείται για τον υπολογισμό της ομοιότητας. Δηλαδή πραγματοποιούνται όλοι οι δυνατοί συνδυασμοί χρηστών και υπολογίζεται κατά πόσο μοιάζουν. Ο ομοιότητά τους σχετίζεται με το πόσο κοντά είναι οι βαθμολογίες που έχουν δώσει σε κοινά αντικείμενα. Το αποτέλεσμα της διαδικασίας αποθηκεύεται στον πίνακα «ratingsdensetocalculatesimilarity». Η έλλειψη εγγραφής ομοιότητας δύο χρηστών σημαίνει ότι δεν έχουν μέχρι στιγμής κοινά βαθμολογημένα αντικείμενα.

Ο πίνακας έχει την παρακάτω δομή:

Πεδίο	**Τύπος**	**Περιγραφή**
user_1	int(11)	Ο πρώτος από το ζευγάρι χρηστών για το οποίο υπολογίζεται η ομοιότητα
user_2	int(11)	Ο δεύτερος από το ζευγάρι χρηστών για το οποίο υπολογίζεται η ομοιότητα
sim	double	Η τιμή της ομοιότητας όπως υπολογίστηκε
user_1_avg	double	Η μέση βαθμολογία του πρώτου χρήστη στο σύστημα
user_2_avg	double	Η μέση βαθμολογία του δεύτερου χρήστη στο σύστημα

Πίνακας 8 Δομή πίνακα ratingsdensetocalculatesimilarity

Για κάθε ζευγάρι χρηστών αποθηκεύεται μία τιμή ομοιότητας και ο συνδυασμός τους αποτελεί το πρωτεύον κλειδί κάθε εγγραφής του πίνακα. Η μέση τιμή της βαθμολογίας κάθε χρήστη είναι μια μεταβλητή που χρειάζεται συχνά στους υπολογισμούς για αυτό και η αποθηκεύεται αντί να υπολογίζεται κάθε φορά για κάθε χρήστη από τον πίνακα με τις βαθμολογίες «ratingsdensetocalculate».

2. Υπολογισμός εμπιστοσύνη μεταξύ των χρηστών

Υπολογισμός των διαφόρων μετρικών της εμπιστοσύνης που απολαμβάνει κάθε χρήστης στο δίκτυο βάσει των επιτυχημένων προβλέψεων που έχει συμμετάσχει στο παρελθόν (έμμεση εμπιστοσύνη). Για τον υπολογισμό των διαφορετικών μετρικών θα χρησιμοποιηθεί

ο πίνακας «ratingsdensetocalculate» ο οποίος περιέχει τις βαθμολογίες από το epinions και χρησιμοποιήθηκε και για τον υπολογισμό της ομοιότητας.

Για κάθε χρήστη του συστήματος επιλέγουμε από τον πίνακα ratingsdensetocalculatesimilarity τους 25 χρήστες με τη μεγαλύτερη ομοιότητα (η τιμή της ομοιότητας θα πρέπει να είναι πάνω του 0,4). Η τιμή 0,4 είναι «γνωστή» τιμή κατωφλίου για την ομοιότητα όταν χρησιμοποιείται στον αλγόριθμο k-NN για την επιλογή των κ (εδώ 25) κορυφαίων χρηστών που θα αποτελέσουν τη γειτονιά του τρέχοντα χρήστη (του χρήστη για τον οποίο θέλουμε τους 25 άλλους χρήστες που έχουν πιο όμοια γούστα με εκείνον).

Στη συνέχεια , κι αφού έχουμε δομήσει τη γειτονιά του τρέχοντα χρήστη βάσει της ομοιότητας παράγουμε προβλέψεις για αυτόν. Δηλαδή χρησιμοποιώντας την εξίσωση 37 και ως βάρος την ομοιότητα από τον πίνακα ratingsdensetocalculatesimilarity παράγουμε μια λίστα με αντικείμενα και βαθμολογίες όπως προέκυψαν από τις βαθμολογίες των γειτόνων. Συγκρίνοντας την προτεινόμενη βαθμολογία και τη βαθμολογία που έχει δώσει ο χρήστης προκύπτει ο βαθμός ορθότητας της πρόβλεψης βάσει του μεγέθους της διαφοράς τους και υπολογίζουμε τις διαφορετικές μετρικές εμπιστοσύνης. Οι μετρικές αυτές και κάποια άλλα στατιστικά στοιχεία που είναι σημαντικά για την εξαγωγή χρήσιμων συμπερασμάτων αποθηκεύονται στους πίνακες ratingsdensetocalculatetrust και ratingsdensetocalculatestatistics

Πεδίο	**Τύπος**	**Περιγραφή**
truster	int(11)	Ο τρέχων χρήστης για τον οποίο έχει σχηματισθεί η γειτονιά
trusted	int(11)	Γείτονας του τρέχοντα , τον οποίο αφορά η τιμή της εμπιστοσύνης που θα δώσει ο truster
trust	double	Η τιμή της τοπικής έμμεσης εμπιστοσύνης του truster προς τον trusted. Είναι το κλάσμα των συνολικών επιτυχημένων προβλέψεων του trusted προς το σύνολο των προβλέψεων του trusted προς τον truster
trustContributor	double	Η τιμή της τοπικής έμμεσης εμπιστοσύνης του truster προς τον trusted. Είναι το κλάσμα των συνολικών επιτυχημένων προβλέψεων του trusted προς το σύνολο των προβλέψεων του

Πεδίο	Τύπος	Περιγραφή
		trusted προς τον truster. Με τη διαφορά ότι επιτυχημένη πρόβλεψη αποτελεί η πρόβλεψη κατά την οποία η γειτονιά έχει φτάσει πολύ κοντά στην πραγματική βαθμολογία ενός αντικειμένου σε αυτή του truster αλλά έχει συνεισφέρει σε αυτή την πρόβλεψη με βαθμολογία που δεν υπερβαίνει το <1 από την πραγματική τιμή το χρήστη.

Πίνακας 9 Δομή πίνακα ratingsdensetocalculatetrust

Για κάθε ζευγάρι χρηστών (truster-trusted) αποθηκεύεται μία τιμή εγγραφή εμπιστοσύνης και ο συνδυασμός τους αποτελεί το πρωτεύον κλειδί του πίνακα

Ο δεύτερος πίνακας στον οποίο αποθηκεύονται πληροφορίες είναι ο ratingsdensetocalculatestatistics

Πεδίο	Τύπος	Περιγραφή
trusted	int(11)	Ο τρέχων χρήστης για τον οποίο έχει σχηματισθεί η γειτονιά
truster	int(11)	Γείτονας του τρέχοντα , τον οποίο αφορά η τιμή της εμπιστοσύνης που θα δώσει ο truster
correct	int(11)	Αριθμός ορθών προβλέψεων (Ο' Donovan) δηλαδή πόσες από τις προβλέψεις της γειτονιάς απείχαν κάτω από το κατώφλι e=1 από την πραγματική βαθμολογία του τρέχοντα χρήστη (ουσιαστικά του truster)
correctContributor	int(11)	Αριθμός ορθών προβλέψεων δηλαδή πόσες από τις προβλέψεις της γειτονιάς απείχαν κάτω από το κατώφλι e=1 από την πραγματική βαθμολογία του τρέχοντα χρήστη (ουσιαστικά του truster) και ταυτόχρονα ο trusted συνείσφερε με βαθμολογία που

Πεδίο	Τύπος	Περιγραφή
		να είναι κάτω από το κατώφλι e1=1 από τη βαθμολογία του τρέχοντα.
allParticipations	int(11)	Σύνολο προβλέψεων στις οποίες συμμετείχε το trusted ανεξαρτήτων ορθότητας
trustGlobal	double	Η τιμή της παγκόσμιας έμμεσης εμπιστοσύνης του truster προς τον trusted. Είναι το κλάσμα των συνολικών επιτυχημένων προβλέψεων του trusted προς το σύνολο των προβλέψεων του trusted προς τον truster
trustContributorGlobal	double	Η τιμή της τοπικής έμμεσης εμπιστοσύνης του truster προς τον trusted. Είναι το κλάσμα των συνολικών επιτυχημένων προβλέψεων του trusted προς το σύνολο των προβλέψεων του trusted προς τον truster. Με τη διαφορά ότι επιτυχημένη πρόβλεψη αποτελεί η πρόβλεψη κατά την οποία η γειτονιά έχει φτάσει πολύ κοντά στην πραγματική βαθμολογία ενός αντικειμένου σε αυτή του truster αλλά έχει συνεισφέρει σε αυτή την πρόβλεψη με βαθμολογία που δεν υπερβαίνει το <1 από την πραγματική τιμή το χρήστη.

Πίνακας 10 Δομή πίνακα ratingsdensetocalculatestatistics

3. Εύρεση γειτονιάς με τον k-NN χρησιμοποιώντας ως κριτήρια επιλογής την ομοιότητα και την εμπιστοσύνη

Ουσιαστικά λαμβάνονται υπόψιν τα αποτελέσματα των βημάτων 1 και 2 για τους k χρήστες με την υψηλότερη τιμή για τη μετρική που έχει τεθεί ως κριτήριο επιλογής (ομοιότητα ή εμπιστοσύνη). Στα πλαίσια της συγκεκριμένης εργασίας επιλέχθηκε γειτονιά 25 ατόμων.

Σχηματίστηκαν 2 διαφορετικές γειτονιές:

a. ***Την ομοιότητα (similarity)***

   (ελάχιστη αποδεκτή τιμή ομοιότητας >0.4 με τον τρέχοντα χρήστη για να συμμετάσχει σε γειτονιά, sim E [-1,1])

b. ***Την τοπική εμπιστοσύνη (localContributorTrust)***

   (ελάχιστη αποδεκτή τιμή εμπιστοσύνης >0.4 από τον τρέχοντα χρήστη για να συμμετάσχει σε γειτονιά localContributorTrust E [0,1])

4. Υπολογισμός των προβλέψεων (εξίσωση 3)

Για την παραγωγή υποδείξεων θα χρησιμοποιήσουμε ένα άλλο κομμάτι του epinions το οποίο αφορά πάλι τους χρήστες με κωδικό 1-299 και περιέχει βαθμολογίες των αντικειμένων 1-230 και που ονομάσαμε ratingsdensetotest_test. Δηλαδή οι εγγραφές βαθμολόγησης αφορούν τους ίδιους χρήστες με αυτούς του πίνακα ratingsdensetocalculate αλλά έχουν αξιολογηθεί άλλα αντικείμενα.

Ο πίνακας έχει την παρακάτω δομή:

Πεδίο	**Τύπος**	**Περιγραφή**
user_id	int(3)	Ο κωδικός του χρήστη που βαθμολογεί
item_id	int(4)	Ο κωδικός του αντικειμένου που βαθμολογείται
ratings	int(1)	Η βαθμολογία που λαμβάνει το αντικείμενο από το χρήστη

Πίνακας 11 Δομή πίνακα ratingsdensetocalculate

Για κάθε χρήστη βλέπουμε ποια αντικείμενα έχει βαθμολογήσει και υπολογίζουμε την τιμή που προτείνουν οι γείτονές του σαν να μην το είχε βαθμολογήσει. Συγκρίνοντας την προτεινόμενη τιμή και αυτή που πραγματικά έχει δώσει ο χρήστης προκύπτει η επιτυχία ή όχι της πρόβλεψης.

Εκτός από τις 2 διαφορετικές γειτονιές που συντάσσονται , υπολογίζονται προβλέψεις με τη χρήση διαφορετικών μετρικών εμπιστοσύνης που αναλύονται στην αντίστοιχη ενότητα. Κάποιες από αυτές τις μετρικές υπολογίζονται on- the –fly εκείνη τη στιγμή πχ neighbourTrust κι άλλες προέρχονται από τη βάση όπου ο υπολογισμός τους έγινε σε προηγούμενη βήμα όπως η localTrust.

a. ***Την τοπική εμπιστοσύνη (localTrust)→από τον πίνακα***
b. ***Την τοπική εμπιστοσύνη (localContributorTrust))→από τον πίνακα***
c. ***Την τοπική εμπιστοσύνη που υπολογίζεται από τους γείτονες (neighbourTrust)→ υπολογίζεται ως ο μέσος όρος των τιμών localContributorTrust των γειτόνων του τρέχοντα χρήστη***
d. ***Την παγκόσμια εμπιστοσύνη (globalTrust))→από τον πίνακα***
e. ***Την συμμετρική εμπιστοσύνη (mutualTrust))→ υπολογίζεται ως ο μέσος όρος της τιμής localContributorTrust του τρέχοντα χρήστη προς τον ενεργό και του ενεργού χρήστη προς τον τρέχοντα***

Τα αποτελέσματα κάθε πειράματος (κάθε πείραμα ορίζεται από την επιλογή του κριτηρίου γειτονιάς κα συντελεστή βάρους της βαθμολογίας κάθε επιλεγμένου γείτονα) αποθηκεύεται στον πίνακα «ratingsdensetocalculatepredictions».

Πεδίο	**Τύπος**	**Περιγραφή**
user_id	int(11)	Ο χρήστης στον οποίο προτάσσεται το αντικείμενο
item_id	int(11)	Το αντικείμενο που προτάσσεται
prediction	double	Η προτεινόμενη βαθμολογία όπως προέκυψε από τη διαδικασία υπόδειξης
error	double	Η απόσταση μεταξύ της πραγματικής βαθμολογίας του χρήστη και της τιμής της υπόδειξης που υπολογίσθηκε για εκείνον.

Πίνακας 12 Δομή πίνακα ratingsdensetocalculatepredictions

Σε κάθε χρήστη γίνεται μοναδική πρόταση για κάθε αντικείμενο από τη γειτονιάς του. Έτσι ο συνδυασμός των πεδίων user_id και item_id αποτελούν το πρωτεύον κλειδί του πίνακα.

Average Επίθεση

Τα βήματα είναι ακριβώς της ίδια λογικής με αυτή που ακολουθήθηκε στην επεξήγηση του πώς υλοποιείται η παραγωγή υποδείξεων για τους γνήσιους χρήστες. Η μόνο διαφοροποίηση έγκειται στην προσθήκη κακόβουλων χρηστών στους δύο πίνακες που προέρχονται από το epinions. Έχουμε θέσει ως παραδοχή ότι όλοι χρήστες και κατ΄ επέκταση οι βαθμολογήσεις που προέρχονται από αυτό το dataset αποτελούνται από γνήσιους χρήστες που δεν έχουν κακόβουλες διαθέσεις κατά την αξιολόγηση των αντικειμένων. Ο φυσικός θόρυβος στα δεδομένα είναι κάτι φυσιολογικό και δεν αποτελεί πρόβλημα.

Οι πίνακες που δημιουργούμε είναι πλήρως αντιστοιχούμενοι με αυτούς των γνησίων και είναι οι εξής:

- Ratingsdensetocalculate_average περιλαμβάνει τιμές από το dataset του epinions με 5253 εγγραφές. Αφορά βαθμολογίες των 299 πρώτων χρηστών (κωδικοί 1 έως και 299) και περιέχει βαθμολογίες αντικειμένων με κωδικούς από 231 έως και 1047. Οι συγκεκριμένες εγγραφές είναι κοινές με τον αντίστοιχο πίνακα Ratingsdensetocalculate που αφορά των υπολογισμό για πριν την επίθεση αλλά σε αυτόν τον πίνακα προσθέτουμε και κακόβουλες βαθμολογίες ώστε να υπολογισθούν ομοιότητα και εμπιστοσύνη παρουσία κακόβουλου θορύβου. Τα κακόβουλα προφίλ επιλέχθηκαν να είναι 15 δηλαδή το 5% των γνησίων χρηστών, παρομοιάζοντας μία πραγματική επίθεση (ούτε πολύ μικρή για να μην έχει αντίκτυπο αλλά ούτε και πολύ μεγάλη ώστε να είναι ευδιάκριτη η κακόβουλη παρουσία στα δεδομένα). Κάθε ένα από τα κακόβουλα προφίλ βαθμολόγησαν 18 διαφορετικά αντικείμενα που επιλέχθηκαν τυχαία με τη χρήστη γεννήτριας τυχαίων αριθμών. Η βαθμολογία τους είναι ίση με το μέσο όρο των βαθμολογιών που έχουν λάβει στο παρελθόν. Σκοπός της συγκεκριμένης επίθεσης και κυρίως με τη βαθμολόγησης το μέσο όρο γίνεται προσπάθεια τα κακόβουλα προφίλ να μοιάζουν όσο το δυνατόν σε περισσότερους χρήστες ώστε να μπορέσουν να εισέλθουν σε όσο το δυνατόν περισσότερες γειτονιές και να συμμετάσχουν στην παραγωγή υποδείξεων για τους υπόλοιπους χρήστες.

- Ratingsdensetocalculatesimilarity_average στον πίνακα αυτόν αποθηκεύονται οι τιμές ομοιότητας για τους χρήστες του συστήματος. Αφού η ομοιότητα

υπολογίσθηκε βάσει του πίνακα Ratingsdensetocalculate_average στον οποίο περιέχονται εγγραφές που αφορούν βαθμολογίες τόσο γνησίων όσο και των κακόβουλων χρηστών της επίθεσης μέσου όρου που εισάγαμε.

- ratingsdensetocalculatetrust_average στον συγκεκριμένο πίνακα αποθηκεύονται οι τιμές που δίνουν οι χρήστες (γνήσιοι και κακόβουλοι) για τους άλλους χρήστες του συστήματος (γνησίους και κακόβουλους)

- ratingsdensetocalculatestatistics_average πίνακας που αποθηκεύονται τα στατιστικά στοιχεία που αφορούν την εμπιστοσύνη όλων των χρηστών του συστήματος

- ratingsdensetotest_test_average ο πίνακας βάσει του οποίου θα γίνουν οι προβλέψεις για την υπόδειξη αντικειμένων. Σε αυτό τον πίνακα όπως και στον Ratingsdensetocalculate_average θα εισάγουμε κακόβουλους χρήστες. Οι κωδικοί των κακόβουλων χρηστών θα είναι ίδιοι με αυτούς του πίνακα ratingsdensetotest_test _average καθώς αποτελούν συνέχιση των συγκεκριμένων προφίλ. Οπότε κι εδώ έχουμε 15 κακόβουλους που όμως βαθμολογούν 20 τυχαία αντικείμενα στο μέσο όρο που λαμβάνουν στο σύστημα (επίθεση μέσου όρου) ενώ το αντικείμενο στόχος, εδώ έχουμε επιλέξει το αντικείμενο 1 θα βαθμολογηθεί στη μέγιστη δυνατή τιμή δηλαδή το 5 από όλους τους κακόβουλους με σκοπό την αύξηση της προτεινόμενης βαθμολόγησης του αντικειμένου και την αύξηση της συχνότητας υπόδειξης του αντικειμένου στους χρήστες του συστήματος που δεν το έχουν βαθμολογήσει ακόμα.

- Ratingsdensetocalculatepredictions_average στο συγκεκριμένο πίνακα αποθηκεύονται οι τελικές προβλέψεις που γίνονται για κάθε χρήστη του συστήματος. Διαφέρουν από τον πίνακα ratingsdensetocalculatepredictions αφού πλέον υπάρχουν οι κακόβουλοι χρήστες στο σύστημα και έχουν συμπεριληφθεί στη διαδικασία παραγωγής των προτάσεων.

Random Επίθεση

Τα βήματα είναι ακριβώς της ίδια λογικής με αυτή που ακολουθήθηκε στην επεξήγηση του πώς υλοποιείται η παραγωγή υποδείξεων για τους κακόβουλους χρήστες την επίθεση μέσου όρου με τη διαφορά ότι στην τυχαία επίθεση η βαθμολόγηση δεν γίνεται στο μέσο όρο της πραγματικής βαθμολογίας των αντικειμένων αλλά τυχαία όπως είναι και η επιλογή των αντικειμένων που θα αξιολογηθούν από τα κακόβουλα προφίλ. Κι εδώ ισχύει η παραδοχή ότι όλοι χρήστες και κατ΄ επέκταση οι βαθμολογήσεις που προέρχονται από αυτό το dataset αποτελούνται από γνήσιους χρήστες που δεν έχουν κακόβουλες διαθέσεις κατά την αξιολόγηση των αντικειμένων. Ο φυσικός θόρυβος στα δεδομένα είναι κάτι φυσιολογικό και δεν αποτελεί πρόβλημα.

Οι πίνακες που δημιουργούμε είναι πλήρως αντιστοιχούμενοι με αυτούς των γνησίων και είναι οι εξής:

- Ratingsdensetocalculate_Random περιλαμβάνει τιμές από το dataset του epinions με 5253 εγγραφές. Αφορά βαθμολογίες των 299 πρώτων χρηστών (κωδικοί 1 έως και 299) και περιέχει βαθμολογίες αντικειμένων με κωδικούς από 231 έως και 1047. Οι συγκεκριμένες εγγραφές είναι κοινές με τον αντίστοιχο πίνακα Ratingsdensetocalculate που αφορά των υπολογισμό για πριν την επίθεση αλλά σε αυτόν τον πίνακα προσθέτουμε και κακόβουλες βαθμολογίες ώστε να υπολογισθούν ομοιότητα και εμπιστοσύνη παρουσία κακόβουλου θορύβου. Τα κακόβουλα προφίλ επιλέχθηκαν να είναι 15 δηλαδή το 5% των γνησίων χρηστών, παρομοιάζοντας μία πραγματική επίθεση (ούτε πολύ μικρή για να μην έχει αντίκτυπο αλλά ούτε και πολύ μεγάλη ώστε να είναι ευδιάκριτη η κακόβουλη παρουσία στα δεδομένα). Κάθε ένα από τα κακόβουλα προφίλ βαθμολόγησαν 18 διαφορετικά αντικείμενα που επιλέχθηκαν τυχαία με τη χρήστη γεννήτριας τυχαίων αριθμών. Η βαθμολογία είναι κι αυτή τυχαία και κυμαίνεται σε όλο το διάστημα της κλίμακας βαθμολόγησης. Επειδή οι βαθμολογίες των αντικειμένων επιλέγεται τυχαία και δεν υπάρχει λογική πίσω από την επιλογή, οι τυχαίες επιθέσεις δεν είναι επιτυχείς όπως αυτές του μέσου όρου στη δημιουργία προφίλ που συσχετίζονται ισχυρά με τους υπόλοιπους χρήστες του συστήματος καθώς σκοπός τους δεν είναι να μπουν σε πολλές γειτονιές και να προωθήσουν συγκεκριμένο αντικείμενο αλλά να μειώσουν την ορθότητα των προβλέψεων εισάγοντας ουσιαστικά «όχι-έξυπνο» θόρυβο στα δεδομένα.

- Ratingsdensetocalculatesimilarity_Random στον πίνακα αυτόν αποθηκεύονται οι τιμές ομοιότητας για τους χρήστες του συστήματος. Αφού η ομοιότητα υπολογίσθηκε βάσει του πίνακα Ratingsdensetocalculate_Random οποίος περιλαμβάνει βαθμολογίες από τους γνησίους χρήστες του Opinions αλλά και των κακόβουλων που εισάγαμε εμείς.

- ratingsdensetocalculatetrust_Random στον συγκεκριμένο πίνακα αποθηκεύονται οι τιμές που δίνουν οι χρήστες (γνήσιοι και κακόβουλοι) για τους άλλους χρήστες του συστήματος (γνησίους και κακόβουλους)

- ratingsdensetocalculatestatistics_Random πίνακας που αποθηκεύονται τα στατιστικά στοιχεία που αφορούν την εμπιστοσύνη όλων των χρηστών του συστήματος

- ratingsdensetotest_test_Random ο πίνακας βάσει του οποίου θα γίνουν οι προβλέψεις για την υπόδειξη αντικειμένων. Σε αυτό τον πίνακα όπως και στον Ratingsdensetocalculate_ Random θα εισάγουμε κακόβουλους χρήστες. Οι κωδικοί των κακόβουλων χρηστών θα είναι ίδιοι με αυτούς του πίνακα ratingsdensetotest_test _Random καθώς αποτελούν συνέχιση των συγκεκριμένων προφίλ. Οπότε κι εδώ έχουμε 15 κακόβουλους που όμως βαθμολογούν 20 τυχαία αντικείμενα στο μέσο όρο που λαμβάνουν στο σύστημα (επίθεση μέσου όρου) ενώ το αντικείμενο στόχος, εδώ έχουμε επιλέξει το αντικείμενο 1 θα βαθμολογηθεί στη μέγιστη δυνατή τιμή δηλαδή το 5 από όλους τους κακόβουλους με σκοπό την αύξηση της προτεινόμενης βαθμολόγησης του αντικειμένου και την αύξηση της συχνότητας υπόδειξης του αντικειμένου στους χρήστες του συστήματος που δεν το έχουν βαθμολογήσει ακόμα.

- Ratingsdensetocalculatepredictions_Random στο συγκεκριμένο πίνακα αποθηκεύονται οι τελικές προβλέψεις που γίνονται για κάθε χρήστη του

συστήματος. Διαφέρουν από τον πίνακα ratingsdensetocalculatepredictions αφού πλέον υπάρχουν οι κακόβουλοι χρήστες στο σύστημα και έχουν συμπεριληφθεί στη διαδικασία παραγωγής των προτάσεων.

5.8 Αποτελέσματα

Στην τελευταία ενότητα αυτού του κεφαλαίου θα παρουσιασθούν τα αριθμητικά αποτελέσματα των πειραμάτων. Σκοπός να συγκριθούν οι δυνάμεις της ομοιότητας και της έμμεσης εμπιστοσύνης σε επίπεδο α) κριτηρίου για την επιλογή γειτόνων και β) συντελεστή βαρύτητας για κάθε βαθμολογία που συνεισφέρει κάθε γείτονας παρουσία θορύβου στα δεδομένα. Ο θόρυβος παρήχθη τεχνικώς δημιουργώντας δύο ανεξάρτητες επιθέσεις μεγέθους 5% (δηλαδή προστέθηκαν στο αρχικό dataset αριθμός κακόβουλων όσο το 5% των γνησίων) ακολουθώντας την στρατηγική επίθεσης μέσου όρου και τυχαίας επίθεσης.

Τα διαφορετικά κριτήρια για την επιλογή των γειτόνων στον αλγόριθμο συσταδοποίησης k-nn είναι η ομοιότητα όπως ορίζεται από τον Resnick και η εμπιστοσύνη localContributor.

Οι διαφορετικοί συντελεστές βάρους είναι οι local, localContributor, neigh, sim, global, globalContributor.

Οι συγκεντρωτικοί πίνακες που ακολουθούν συγκεντρώνουν τις τιμές που προέκυψαν και αφορούν την ορθότητα των προβλέψεων στο dataset πριν και μετά τις επιθέσεις.

Η «Κατηγορία» αφορά την περιγραφή των χρηστών του dataset. Genuine σημαίνει ότι έχουν με το αρχικό χωρίς την εισαγωγή κακόβουλων χρηστών, average σημαίνει ότι έχει γίνει προσθήκη κακόβουλων χρηστών βάσει της στρατηγικής μέσου όρου και τέλος Random ότι έχουν εισαχθεί το αρχικό dataset κακόβουλοι χρήστες βάσει της τυχαίας στρατηγικής επίθεσης,

Η «Γειτονιά» αφορά το κριτήριο επιλογής του k-nn και μπορεί να είναι η ομοιότητα ή η έμμεση μετρική εμπιστοσύνης LOCALCONTRIBUTOR που προτάθηκε σε αυτή την εργασία.

Το «Βάρος» αφορά το συντελεστή βαρύτητας των βαθμολογιών των χρηστών που χρησιμοποιήθηκε για τον υπολογισμό των προβλέψεων.

Τέλος ακολουθούν οι τιμές mae και maue για την αξιολόγηση κάθε συνδυασμού κατηγορίας χρηστών – κριτηρίου δόμησης γειτονιάς και συντελεστή βαρύτητας.

Κατηγορία	Γειτονιά	Βάρος	mae	meue
average	LOCALCONTRIBUTOR	GLOBALCONTRIBUTOR	0,76649	0,8021
average	LOCALCONTRIBUTOR	SIM	0,76699	0,8009
average	LOCALCONTRIBUTOR	LOCALCONTRIBUTOR	0,76705	0,7994
average	LOCALCONTRIBUTOR	NEIGH	0,76757	0,8025
average	LOCALCONTRIBUTOR	GLOBAL	0,76757	0,8027
average	LOCALCONTRIBUTOR	LOCAL	0,76873	0,8036
genuine	LOCALCONTRIBUTOR	SIM	0,78366	0,8193
genuine	LOCALCONTRIBUTOR	GLOBALCONTRIBUTOR	0,78428	0,8207
genuine	LOCALCONTRIBUTOR	GLOBAL	0,7843	0,8204
genuine	LOCALCONTRIBUTOR	NEIGH	0,78482	0,8205
genuine	LOCALCONTRIBUTOR	LOCALCONTRIBUTOR	0,78499	0,8188
genuine	LOCALCONTRIBUTOR	LOCAL	0,78583	0,8222
Random	LOCALCONTRIBUTOR	SIM	0,80477	0,8391
Random	LOCALCONTRIBUTOR	GLOBALCONTRIBUTOR	0,80512	0,8407
Random	LOCALCONTRIBUTOR	GLOBAL	0,80526	0,8407
Random	LOCALCONTRIBUTOR	NEIGH	0,80609	0,8414
Random	LOCALCONTRIBUTOR	CONTRIBUTOR	0,8061	0,8392
Random	LOCALCONTRIBUTOR	LOCAL	0,80699	0,8426
average	SIM	GLOBAL	0,8171	0,8571
average	SIM	SIM	0,8172	0,8568
average	SIM	GLOBALCONTRIBUTOR	0,8174	0,8586
genuine	SIM	SIM	0,8238	0,8637
genuine	SIM	GLOBAL	0,8252	0,8641
genuine	SIM	GLOBALCONTRIBUTOR	0,826	0,8647
average	SIM	LOCALCONTRIBUTOR	0,8336	0,8689
average	SIM	NEIGH	0,8348	0,8775
average	SIM	LOCAL	0,8366	0,877
genuine	SIM	NEIGH	0,8471	0,8817
genuine	SIM	CONTRIBUTOR	0,8476	0,8754
genuine	SIM	LOCAL	0,8497	0,8824
Random	SIM	SIM	0,8721	0,8955
Random	SIM	GLOBALCONTRIBUTOR	0,8725	0,8964
Random	SIM	GLOBAL	0,8727	0,8963
Random	SIM	NEIGH	0,8936	0,9154
Random	SIM	LOCAL	0,8974	0,9159
Random	SIM	LOCALCONTRIBUTOR	0,8993	0,9105

Πίνακας 13 Ταξινόμηση κατά mae

Κατηγορία	Γειτονιά	Βάρος	mae	meue
average	LOCALCONTRIBUTOR	LOCALCONTRIBUTOR	0,76705	0,7994
average	LOCALCONTRIBUTOR	SIM	0,76699	0,8009
average	LOCALCONTRIBUTOR	GLOBALCONTRIBUTOR	0,76649	0,8021
average	LOCALCONTRIBUTOR	NEIGH	0,76757	0,8025
average	LOCALCONTRIBUTOR	GLOBAL	0,76757	0,8027
average	LOCALCONTRIBUTOR	LOCAL	0,76873	0,8036
genuine	LOCALCONTRIBUTOR	LOCALCONTRIBUTOR	0,78499	0,8188
genuine	LOCALCONTRIBUTOR	SIM	0,78366	0,8193
genuine	LOCALCONTRIBUTOR	GLOBAL	0,7843	0,8204
genuine	LOCALCONTRIBUTOR	NEIGH	0,78482	0,8205
genuine	LOCALCONTRIBUTOR	GLOBALCONTRIBUTOR	0,78428	0,8207
genuine	LOCALCONTRIBUTOR	LOCAL	0,78583	0,8222
Random	LOCALCONTRIBUTOR	SIM	0,80477	0,8391
Random	LOCALCONTRIBUTOR	LOCALCONTRIBUTOR	0,8061	0,8392
Random	LOCALCONTRIBUTOR	GLOBALCONTRIBUTOR	0,80512	0,8407
Random	LOCALCONTRIBUTOR	GLOBAL	0,80526	0,8407
Random	LOCALCONTRIBUTOR	NEIGH	0,80609	0,8414
Random	LOCALCONTRIBUTOR	LOCAL	0,80699	0,8426
average	SIM	SIM	0,8172	0,8568
average	SIM	GLOBAL	0,8171	0,8571
average	SIM	GLOBALLOCALCONTRIBUTOR	0,8174	0,8586
genuine	SIM	SIM	0,8238	0,8637
genuine	SIM	GLOBAL	0,8252	0,8641
genuine	SIM	GLOBALCONTRIBUTOR	0,826	0,8647
average	SIM	LOCALCONTRIBUTOR	0,8336	0,8689
genuine	SIM	LOCALCONTRIBUTOR	0,8476	0,8754
average	SIM	LOCAL	0,8366	0,877
average	SIM	NEIGH	0,8348	0,8775
genuine	SIM	NEIGH	0,8471	0,8817
genuine	SIM	LOCAL	0,8497	0,8824
Random	SIM	SIM	0,8721	0,8955
Random	SIM	GLOBAL	0,8727	0,8963
Random	SIM	GLOBALCONTRIBUTOR	0,8725	0,8964
Random	SIM	LOCALCONTRIBUTOR	0,8993	0,9105
Random	SIM	NEIGH	0,8936	0,9154
Random	SIM	LOCAL	0,8974	0,9159

Πίνακας 14 Ταξινόμηση κατά maue

- Και από τους δύο πίνακες (Πίνακας 13& Πίνακας 14) είναι εμφανές ότι όταν η γειτονιά επιλέγεται με βάσει την εμπιστοσύνη έχει πολύ καλύτερα αποτελέσματα σε όρους ορθότητας από όταν το κριτήριο επιλογής των γειτόνων είναι η ομοιότητα. Δηλαδή ανεξαρτήτως αν έχει δεχθεί το dataset επίθεση και ανεξαρτήτως του συντελεστή βαρύτητας που έχει χρησιμοποιηθεί όταν επιλέγονται αξιόπιστοι γείτονες και όχι όμοιοι γείτονες τότε οι υποδείξεις είναι πάντα καλύτερες. Αυτή η παρατήρηση αφορά και τις δύο μετρικές αξιολόγησης. Δηλαδή με τη χρήση αξιόπιστων χρηστών μειώνεται τόσο το μέσο λάθος (mae) όσο και το μέσο λάθος ανά χρήστη (maue).

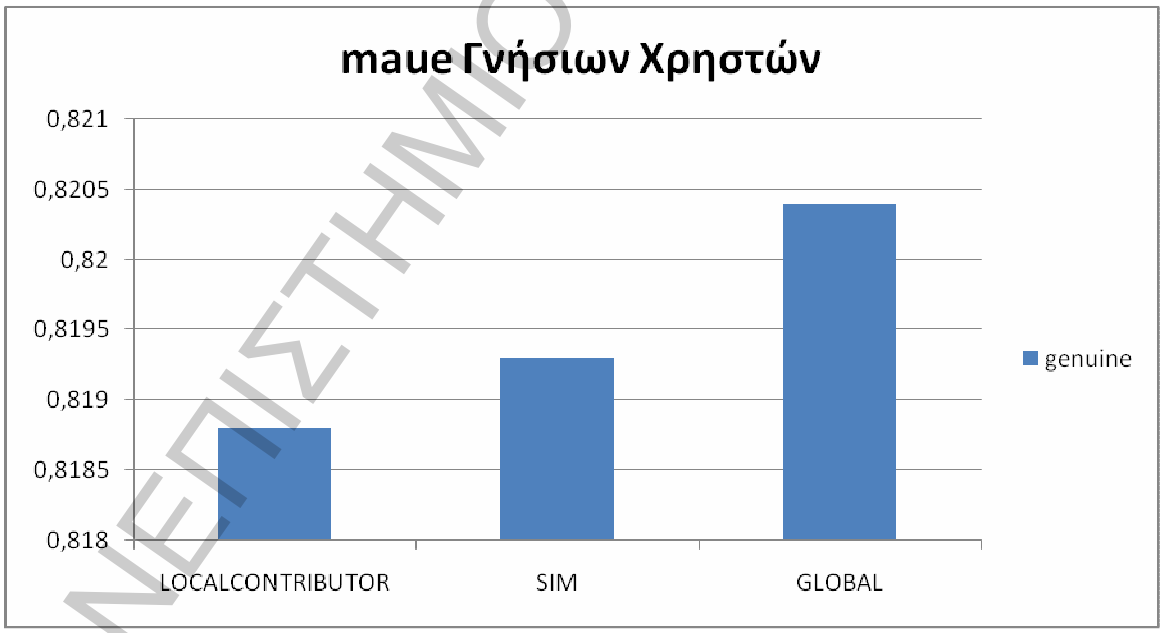
- Οι γειτονιές που δομούνται βάσει της εμπιστοσύνης φαίνεται ότι δίνουν καλύτερα αποτελέσματα τόσο πριν όσο και μετά την εκδήλωση επιθέσεων σε σχέση με την εμπιστοσύνη, έτσι συμπεραίνουμε ότι μπορεί να χρησιμοποιηθεί ως κριτήριο επιλογής γειτόνων σε κάθε περίπτωση.

- Από τους δύο πίνακες είναι ορατή η δράση που έχουν οι δύο διαφορετικού επιθέσεις στο dataset. Θεωρώντας ως σημείο εκκίνησής μας τα αποτελέσματα των μετρικών στο dataset που περιέχει μόνο γνήσιους χρήστες (Κατηγορία= genuine) παρατηρούμε ότι η ορθότητα των προβλέψεων πέφτει σημαντικά μετά την εκδήλωση της τυχαίας επίθεσης ενώ αυξάνεται μετά την επίθεση μέσου όρου. Αυτές οι παρατηρήσεις αφορούν όλους τους συνδυασμούς γειτονιάς –βάρους.

  Κάτι τέτοιο δικαιολογείται από το γεγονός ότι η τυχαία επίθεση εξαπολύεται χωρίς να υπάρχει «λογική» στρατηγική που να επιλέγει τα αντικείμενα ή έστω τις βαθμολογίες τους , συνεπώς δημιουργούν αρνητικές επιπτώσεις στην ποιότητα των προβλέψεων. Αντίθετα οι επιθέσεις μέσου όρου επειδή βαθμολογούν κοντά στο πραγματικό μέσο όρο που λαμβάνουν τα αντικείμενα στο σύστημα είναι πολύ πιθανόν να βοηθήσουν στη δημιουργία προφίλ που σχετίζονται θετικά με πολλούς είτε γιατί μοιάζουν στις βαθμολογίες είτε γιατί προτείνουν αντικείμενα με μια βαθμολογία πολύ κοντά στη γνώμη του μέσου χρήστη.

- Από τους δύο πίνακες είναι ορατή η δράση που έχουν οι δύο διαφορετικού επιθέσεις στο dataset. Θεωρώντας ως σημείο εκκίνησής μας τα αποτελέσματα των μετρικών στο dataset που περιέχει μόνο γνήσιους χρήστες (Κατηγορία= genuine) παρατηρούμε ότι η ορθότητα των προβλέψεων πέφτει σημαντικά μετά την εκδήλωση της τυχαίας επίθεσης

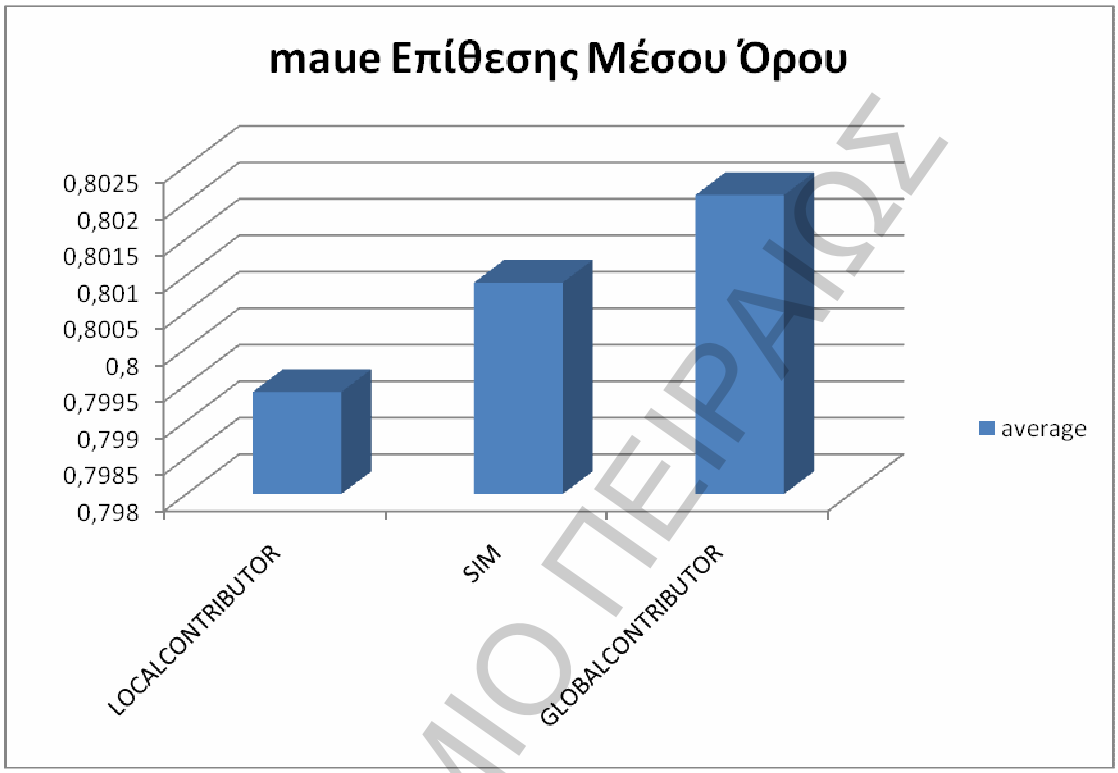
- Καλύτερος συνδυασμός είναι η επιλογή γειτονιάς με localContributor και συντελεστής βάρους πάλι τη localContributor ενώ και με την ομοιότητα δίνει πολύ καλά αποτελέσματα.



Εικόνα 5 Βάρη σε σχέση με maue (dataset χωρίς επίθεση)

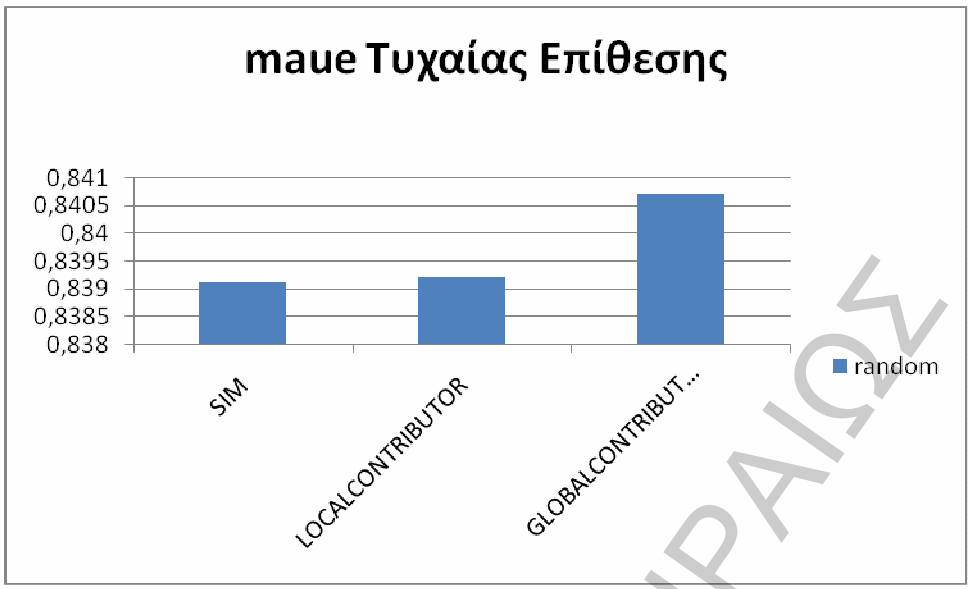
- Στην Εικόνα 5 εμφανίζονται οι τρεις συντελεστές βάρους που έδωσαν το χαμηλότερο maue στο dataset πριν ακόμα εκδηλωθεί κάποια από τις επιθέσεις (ύπαρξη μόνο γνήσιων χρηστών Και στις 3 περιπτώσεις η γειτονιά έχει δομηθεί βάσει της εμπιστοσύνης (έχει αναφερθεί παραπάνω ότι είχε τα καλύτερα αποτελέσματα για όλες τις περιπτώσεις και εννοείται για όλες τις περιπτώσεις που ακολουθούν). Συγκρίνοντας τις αποδόσεις φαίνεται ότι η μετρική localContributor

είχε το χαμηλότερο μέσο λάθος ανά χρήστη και ακολουθούν η ομοιότητα και η global.



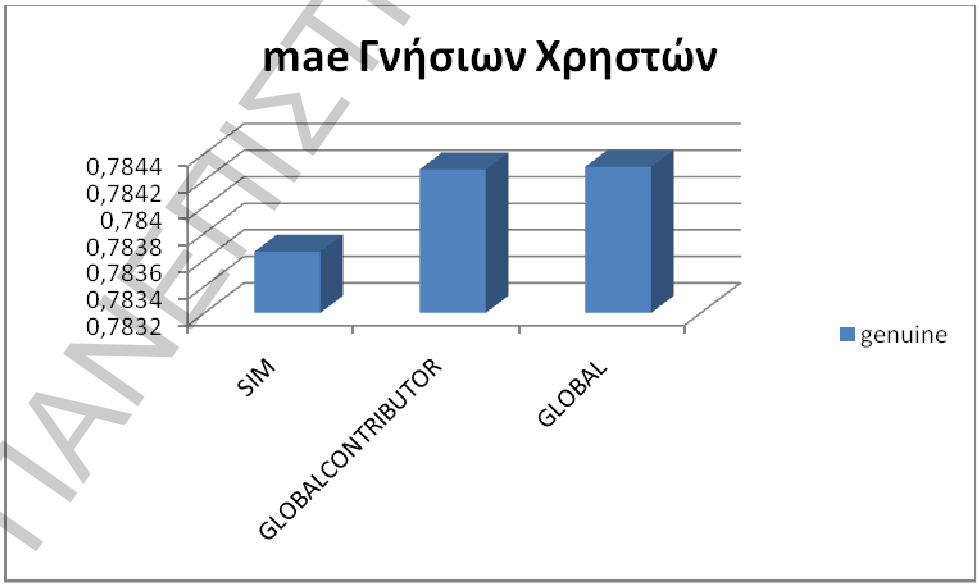
Εικόνα 6 Βάρη σε σχέση με maue (dataset με εκδηλωμένη επίθεση μέσου όρου)

- Στην Εικόνα 6 φαίνονται οι τρεις συντελεστές βάρους που έδωσαν το χαμηλότερο maue στα πλαίσια της επίθεσης μέσου όρου. Συγκρίνοντας τις αποδόσεις φαίνεται ότι η μετρική localContributor είχε το χαμηλότερο μέσο λάθος ανά χρήστη ενώ την ακολουθούν η ομοιότητα και η globalContributor.



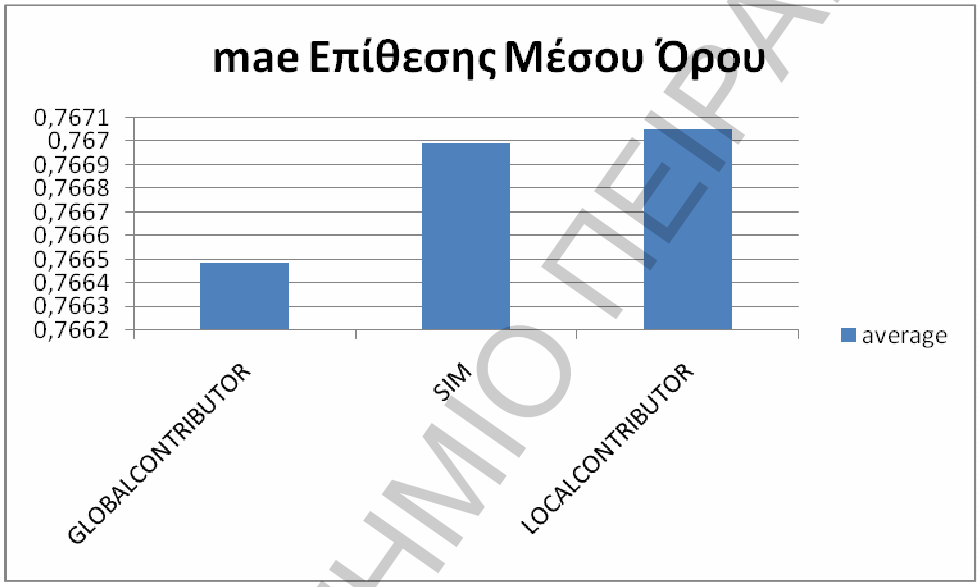
Εικόνα 7 Βάρη σε σχέση με maue (dataset με εκδηλωμένη τυχαία επίθεση)

- Στην Εικόνα 7 εμφανίζονται οι τρεις συντελεστές βάρους που έδωσαν το χαμηλότερο maue ενώ είχε εκδηλωθεί η τυχαία επίθεση στο αρχικό dataset . Συγκρίνοντας τις αποδόσεις φαίνεται ότι η μετρική της ομοιότητας είχε την καλύτερη ενώ την ακολουθεί με μικρή διαφορά η localContributor και τέλος η globalContributor.



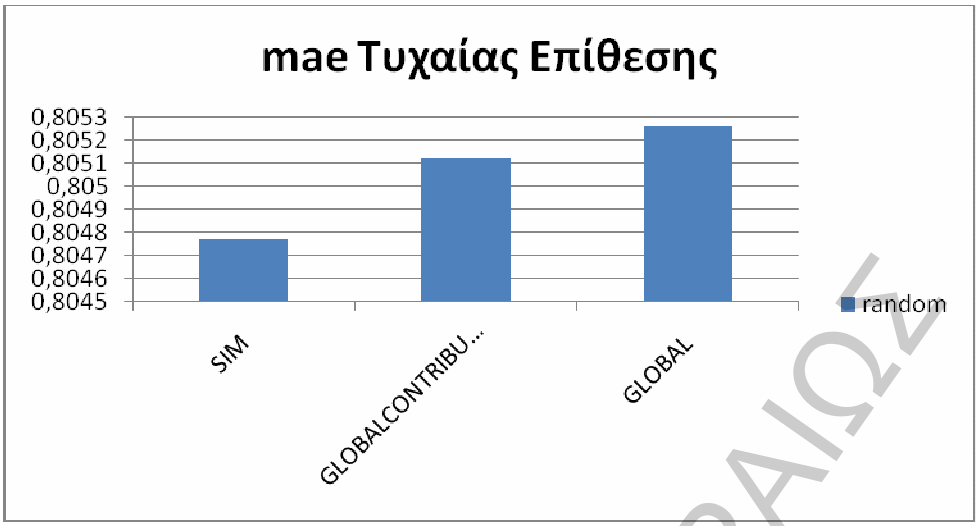
Εικόνα 8 Βάρη σε σχέση με mae (dataset χωρίς επίθεση)

- Στην Εικόνα 8 φαίνονται οι τρεις συντελεστές βάρους που έδωσαν το χαμηλότερο mae στα πλαίσια της αξιολόγησης των διαφόρων μετρικών στο dataset πριν την εκδήλωση οποιασδήποτε επίθεσης. Συγκρίνοντας τις αποδόσεις φαίνεται ότι η ομοιότητα έδωσε το χαμηλότερο μέσο λάθος ακολουθούμενη από τις μετρικές εμπιστοσύνης globalContributor και global.



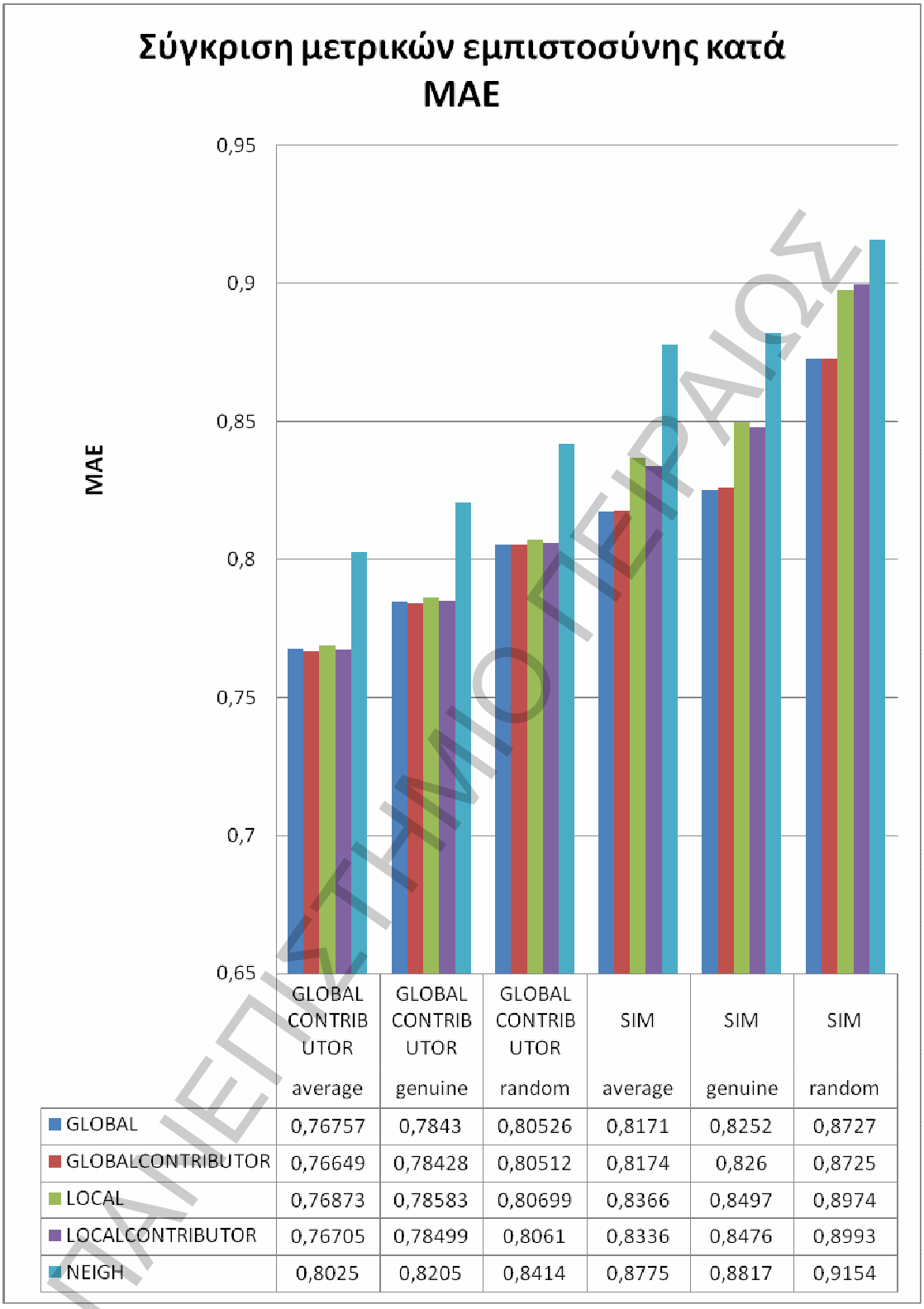
Εικόνα 9 Βάρη σε σχέση με mae (dataset με εκδηλωμένη επίθεση μέσου όρου)

- Στην Εικόνα 9 φαίνονται οι τρεις συντελεστές βάρους που έδωσαν το χαμηλότερο mae στα πλαίσια της αξιολόγησης των διαφόρων μετρικών στο dataset μετά την εκδήλωση της επίθεσης μέσου όρου. Συγκρίνοντας τις αποδόσεις φαίνεται ότι η μετρική globalContributor είχε την καλύτερη ακολουθούμενη από την ομοιότητα και τη localContributor.



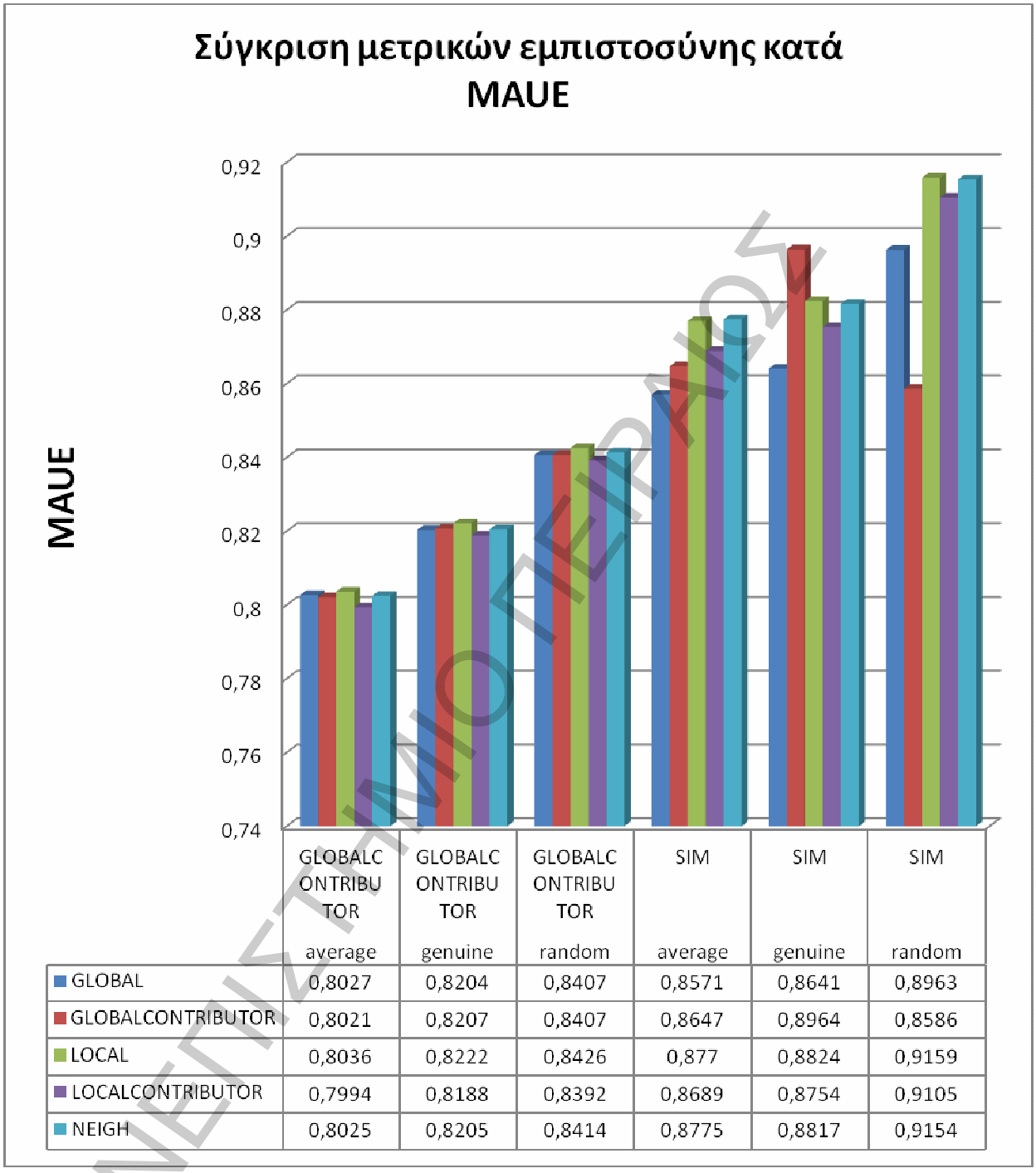
Εικόνα 10 Βάρη σε σχέση με maue (dataset με εκδηλωμένη τυχαία επίθεση)

- Στην Εικόνα 10 παρουσιάζονται οι τρεις συντελεστές βάρους που έδωσαν το χαμηλότερο mae στα πλαίσια της αξιολόγησης των διαφόρων μετρικών του dataset μετά την εκδήλωση της τυχαίας επίθεσης. Συγκρίνοντας τις αποδόσεις φαίνεται ότι η ομοιότητα έδωσε το χαμηλότερο μέσο λάθος ακολουθούμενη από τις μετρικές εμπιστοσύνης globalContributor και global.



	GLOBAL CONTRIBUTOR average	GLOBAL CONTRIBUTOR genuine	GLOBAL CONTRIBUTOR random	SIM average	SIM genuine	SIM random
GLOBAL	0,76757	0,7843	0,80526	0,8171	0,8252	0,8727
GLOBALCONTRIBUTOR	0,76649	0,78428	0,80512	0,8174	0,826	0,8725
LOCAL	0,76873	0,78583	0,80699	0,8366	0,8497	0,8974
LOCALCONTRIBUTOR	0,76705	0,78499	0,8061	0,8336	0,8476	0,8993
NEIGH	0,8025	0,8205	0,8414	0,8775	0,8817	0,9154

Εικόνα 11 Σύγκριση μετρικών εμπιστοσύνης ως προς mae



	GLOBALCONTRIBUTOR average	GLOBALCONTRIBUTOR genuine	GLOBALCONTRIBUTOR random	SIM average	SIM genuine	SIM random
GLOBAL	0,8027	0,8204	0,8407	0,8571	0,8641	0,8963
GLOBALCONTRIBUTOR	0,8021	0,8207	0,8407	0,8647	0,8964	0,8586
LOCAL	0,8036	0,8222	0,8426	0,877	0,8824	0,9159
LOCALCONTRIBUTOR	0,7994	0,8188	0,8392	0,8689	0,8754	0,9105
NEIGH	0,8025	0,8205	0,8414	0,8775	0,8817	0,9154

Εικόνα 12 Σύγκριση μετρικών εμπιστοσύνης ως προς maue

- Αξιολογώντας τις διαφορετικές μετρικές της εμπιστοσύνης (Εικόνες 11 και 12) είναι φανερό ότι αυτές που δεν αρκούνται στις καλές υποδείξεις των γειτόνων αλλά εμπεριέχουν και τον έλεγχο του αν ο κάθε γείτονας προσέφερε θετικά στο να είναι επιτυχημένη η υπόδειξη επιφέρουν καλύτερα αποτελέσματα για το μέσο λάθος και το μέσο λάθος ανά χρήστη.

- Βάσει των Εικόνων 11 και 12 βλέπουμε ότι όσον αφορά το mae οι μετρικές με παγκόσμια ισχύ έχουν προβάδισμα στις αποδόσεις ως προς mae ενώ όσον αφορά το maue προβάδισμα απόδοσης (δηλαδή χαμηλότερες τιμές) έχουν οι τοπικές μεταβλητές**.** Αυτό πιθανολογείται πως οφείλεται στον ορισμό τους, δηλαδή το mae αφορά το μέσο λάθος όλων των προβλέψεων οπότε οι παγκόσμιες μεταβλητές καταφέρνουν κι εξισορροπούν τις αποκλίσεις ενώ το maue αφορά το μέσο λάθος ανά χρήστη οπότε και παρουσιάζει καλύτερα αποτελέσματα με τοπικές μεταβλητές που αφορούν κάθε ενεργό χρήστη ξεχωριστά. Τέλος η μετρική που αφορά την εμπιστοσύνη όπως αυτή προκύπτει από κάθε γειτονιά παρουσίασε σχεδόν πάντα τα χειρότερα αποτελέσματα καθιστώντας τη , τη λιγότερο αξιόπιστη μετρική όσον αφορά την ορθότητα των προβλέψεων.

- Σύμφωνα με τους *O' Donoval & Smyth (2005)* η εμπιστοσύνη ορίζεται ως παγκόσμια μεταβλητή και είναι το κλάσμα των σωστών προβλέψεων στις οποίες συμμετείχε ο χρήστης προς το σύνολο των προβλέψεων που συμμετείχε. Στα πλαίσια της εργασίας προτάθηκαν δύο παραλλαγές αυτής. Η localContributor και η globalContributor . Και στις δύο μετρικές για να θεωρηθεί μια πρόβλεψη επιτυχημένη δεν αρκεί η γειτονιά να έχει προσφέρει μια καλή πρόταση, αλλά θα πρέπει και ο ενεργός χρήστης να έχει συνεισφέρει σε αυτή με βαθμολογία που να είναι κοντά σε αυτή του χρήστη. Αυτή είναι η βασική διαφοροποίηση με τους *O' Donoval & Smyth (2005).* Η globalContributor είναι η παγκόσμια μετρική που χρησιμοποιεί τη σκέψη ότι δεν αρκεί να ανήκεις σε μια καλή γειτονιά πρέπει να είσαι κι εσύ καλός γείτονας ενώ η localContributor είναι τοπικής εμβέλειας. Δηλαδή η localContributor δεν υπολογίζεται ως το κλάσμα των προβλέψεων στις οποίες συνέφερε θετικά στο δίκτυο προς το σύνολο τους αλλά ως το κλάσμα των προβλέψεων στις οποίες συνέφερε θετικά προς κάθε άλλο χρήστη προς το σύνολο των προβλέψεων προς αυτό το χρήστη. Για κάθε χρήστη υπάρχει μία μόνο τιμή globalContributor στο δίκτυο ενώ για κάθε χρήστη υπάρχει διαφορετική τιμή localContributor από κάθε άλλο χρήστη (χωρίς να είναι απαραίτητο ότι κάθε χρήστης θα έχει τιμή για κάθε άλλο χρήστη, η ύπαρξη τοπικής τιμής εξαρτάται από τις γειτονιές που δομούνται)

Οι localContributor και globalContributor έχουν καλύτερες αποδόσεις από αυτή global ενώ μαζί με την sim (ομοιότητα) αποτελούν τις καλύτερες επιλογές συντελεστή βαρών.

neighbourhood	weight	Prediction Shift
trustContributor	globalContributor	0,000961032
trustContributor	Contributor	0,000699696
trustContributor	sim	0,000321099
trustContributor	local	0,00028542
trustContributor	neigh	0,000205703
trustContributor	global	0,000185344
sim	globalContributor	0,003842996
sim	sim	0,004087594
sim	global	0,005264694
sim	local	0,008210494
sim	Contributor	0,009883091
sim	neigh	0,01097118

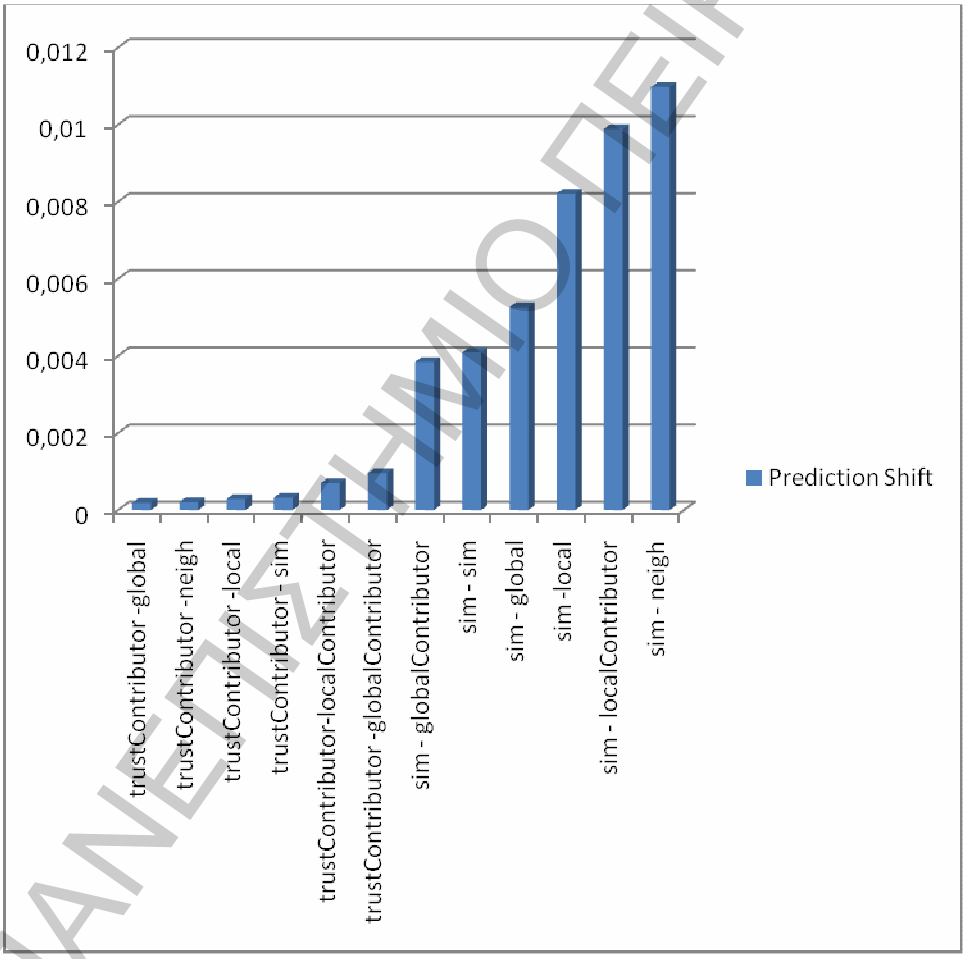
Πίνακας 15 Ταξινόμηση κατά Prediction Shift

- Αναλύοντας τα αποτελέσματα ευρωστίας μέσω της μετρικής Prediction Shift, φαίνεται ότι η επιλογή της γειτονιάς είναι πιο σημαντική από την επιλογή συντελεστή βαρύτητας για κάθε βαθμολογία αφού όπως φαίνεται στον Πίνακα 15 **σε** όλες τις περιπτώσεις - όπως και κατά την αξιολόγηση της ορθότητας- όλοι οι συνδυασμοί που έχουν ως κριτήριο επιλογής γειτόνων την εμπιστοσύνη έχουν πολύ καλύτερη απόδοση από όλους τους συνδυασμούς με κριτήριο την ομοιότητα. Έτσι η εμπιστοσύνη φαίνεται να προσφέρει σημαντική δυναμική ευρωστίας στο σύστημα. Δηλαδή η επιλογή γειτόνων που εμπιστεύεσαι αντί για γείτονες που μοιάζουν με εσένα μπορεί να σε ωφελήσει όχι μόνο γιατί οι προτάσεις που προκύπτουν είναι λιγότερο μεροληπτικές αλλά και γιατί όπως είδαμε στους προηγούμενους πίνακες ταυτόχρονα υπάρχει θετική επίδραση στο επίπεδο ορθότητά τους.

- Όσον αφορά το συντελεστή βαρύτητας τα καλύτερα αποτελέσματα προέρχονται από τη μετρική global και ακολουθούν με μικρή διαφορά οι τοπικές μεταβλητές neigh που υπολογίζεται από τις γνώμες της εκάστοτε γειτονιάς που συμμετέχει για

κάθε τρέχοντα χρήστη και τη local που είναι αντίστοιχη της global αλλά το κλάσμα αφορά τις προβλέψεις που γίνονται προς τον τρέχοντα χρήστη και όχι προς όλες τις προβλέψεις του ενεργού στο δίκτυο. (τοπική μεταβλητή)

Στην Εικόνα 11 ακολουθεί ραβδόγραμμα που εμφανίζει τις τιμές τις μετρικής Prediction Shift όπως υπολογίστηκαν μετά την επίθεση μέσου όρου (στην τυχαία επίθεση δεν μπορεί να χρησιμοποιηθεί αυτή η μετρική καθώς δεν υπάρχει συγκεκριμένο αντικείμενο στόχος). Όπως φαίνεται όταν η επιλογή των γειτόνων πραγματοποιείται με τη μετρική της εμπιστοσύνης αυξάνεται η ευρωστία του συστήματος.



Εικόνα 11 Prediction Shift

6 Σύνοψη εργασίας

Ο σύγχρονος καταναλωτής αφιερώνει πολύ χρόνο στο Διαδίκτυο αναζητώντας αντικείμενα που θα ικανοποιήσουν τις προσωπικές ανάγκες του. Εκτός του βαθμού ικανοποίησης της κάθε ανάγκης, σημαντικοί παράγοντες που λαμβάνονται υπόψιν είναι ο χρόνος που διέθεσε ο χρήστης για να βρει το επιθυμητό αντικείμενο αλλά και το κόστος απόκτησής του. Δεδομένης της υπέρμετρης πληροφορίας που διακινείται στο Διαδίκτυο και του λίγου χρόνου που θέλει ο χρήστης να διαθέσει ώστε να ανακαλύψει τα σωστά αντικείμενα γεννήθηκε η ανάγκη ύπαρξής των Συστημάτων Υπόδειξης. Δηλαδή συστημάτων που στηρίζονται σε διαφορετικά μοντέλα προτείνουν στο χρήστη αντικείμενα βάσει των κριτηρίων που θέτει. Το πιο διαδεδομένο μοντέλο που έχει χρησιμοποιηθεί γενικότερα στη βιβλιογραφία αλλά στα πλαίσια της παρούσας διπλωματικής εργασίας είναι η Συνεργατική Μέθοδος. Η Συνεργατική Μέθοδος προσφέρει το μεγάλο πλεονέκτημα ότι μπορεί να παράξει υποδείξεις για όλους τους χρήστες ακόμα και για αυτούς με πιο περίεργα γούστα ενώ το μεγάλο της μειονέκτημα είναι ότι λόγω της ανοικτής φύσης της -δηλαδή ότι κάθε χρήστης μπορεί να βαθμολογήσει κάθε αντικείμενο όπως επιθυμεί- είναι ευάλωτη σε κακόβουλες επιθέσεις.

Η Συνεργατική Μέθοδος προτείνει αντικείμενα στους χρήστες βάσει των προηγούμενων επιλογών τους. Βασική έννοια σε αυτό το μοντέλο είναι η ομοιότητα, δηλαδή κατά πόσο δύο χρήστες του ίδιου συστήματος βαθμολογούν παρόμοια ίδια αντικείμενα. Έχοντας υπολογίσει για κάθε χρήστη το βαθμό ομοιότητάς του με τους άλλους χρήστες δημιουργείται μια γειτονιά για κάθε χρήστη (τον τρέχοντα χρήστη) αποτελούμενη από εκείνους που έχουν την υψηλότερη ομοιότητα με αυτόν. Εν συνεχεία οι γείτονες προτείνουν αντικείμενα που έχουν βαθμολογήσει στον τρέχοντα χρήστη (τα αντικείμενα αυτά δεν πρέπει να έχουν βαθμολογηθεί από τον τρέχοντα). Από τη λίστα με τα προτεινόμενα από τους γείτονες αντικείμενα τελικά υποδεικνύονται στον τρέχοντα αυτά με την υψηλότερη βαθμολογία. Για την παραγωγή της τελικής βαθμολογίας κάθε αντικειμένου χρησιμοποιείται συνηθέστερα η εξίσωση (25) όπου η βαθμολογία του κάθε γείτονα για το αντικείμενο προς υπόδειξη πολλαπλασιάζεται με ένα βάρος , συνηθέστερα την ομοιότητα. Σημαντικό μέτρο αξιολόγησης των μοντέλων υπόδειξης είναι η ορθότητα, δηλαδή το κατά πόσο οι βαθμολογίες που προτείνονται για κάθε αντικείμενο από την γειτονιά του τρέχοντα χρήστη είναι κοντά στην βαθμολογία που θα δώσει εκείνος. Τα πιο συνήθη μέτρα ορθότητας είναι το MAE (Mean Average Error) και το MAUE (Mean Average User Error) που μετρούν τη μέση απόλυτη απόκλιση της βαθμολογίας κάθε υπόδειξης από τη βαθμολογία που έδωσε ο τρέχοντας χρήστης, και αντίστοιχα τη μέση απόλυτη απόκλιση των

προτεινόμενων βαθμολογιών και των βαθμολογιών που έδωσαν οι τρέχοντες χρήστες ανά χρήστη.

Στην προσπάθεια να βελτιωθεί η ορθότητα των προβλέψεων προτάθηκε η εμπιστοσύνη στα Συστήματα Υπόδειξης, μεταφέροντας στο πεδίο μια έννοια που διέπει τις ανθρώπινες σχέσεις και έχει μελετηθεί από άλλες επιστήμες όπως η ψυχολογία, η κοινωνιολογία , η φιλοσοφία κά. «Η εμπιστοσύνη είναι ένα στοίχημα για τις μελλοντικές ενδεχόμενες ενέργειες των άλλων» σύμφωνα με το *Sztompka (1999)* ενώ η Golbeck στη διδακτορική της διατριβή βασίζεται σε ορισμούς που στηρίζονται κυρίως στην κοινωνιολογία και ορίζει ότι ο χρήστης Α εμπιστεύεται το χρήστη Β αν δεσμεύεται σε μια ενέργεια με βάση την πεποίθηση ότι οι μελλοντικές δράσεις του Β θα οδηγήσουν σε ένα καλό αποτέλεσμα. Η εμπιστοσύνη μπορεί να είναι έμμεση όταν είναι υπολογιζόμενο μέτρο O'Donovan et al ή άμεση όταν δηλώνει ρητά ο χρήστης Α ότι εμπιστεύεται ή δεν εμπιστεύεται το χρήστη Β (epinions.com). Ανεξαρτήτως του αν είναι έμμεση ή άμεση έχει αποδειχθεί μέσω της έρευνας ότι η χρήση της βελτιώνει την ορθότητα των προβλέψεων *(Golbeck 2004, O' Donovan & Smyth 2005)*

Όμως για να είναι επιτυχημένο ένα σύστημα δεν αρκεί να προτάσσει ορθές υποδείξεις αλλά πρέπει να αντιμετωπίζει αποτελεσματικά και τις επιθέσεις που δέχεται. Τα Συστήματα Υπόδειξης που στηρίζονται στη Συνεργατική Μέθοδο χρησιμοποιούν τα προφίλ των χρηστών για να παράξουν υποδείξεις. Δηλαδή αξιοποιούν τις βαθμολογίες που έχουν δώσει για τα αντικείμενα που έχουν δει, συνεπώς προτρέπουν τους χρήστες να βαθμολογούν ώστε να έχουν περισσότερα δεδομένα στη διάθεσή τους. Τα εν λόγω συστήματα έχουν συνδεθεί θετικά με την αύξηση της πιστότητας των χρηστών και με την αύξηση των πωλήσεων οπότε συχνά γίνονται στόχοι επίθεσης από κακόβουλες οντότητες. Ως επίθεση ορίζεται ένας μετασχηματισμός που αντιστοιχίζει τη βάση δεδομένων του συστήματος σε μία νέα βάση δεδομένων. Ο πιο δημοφιλής και κοινός τρόπος επίθεσης αποτελείται από τη δημιουργία μίας ομάδας μεροληπτικών προφίλ που εισάγονται στο σύστημα που αποτελούν θόρυβο. Κάθε ένας από τους λογαριασμούς του κακόβουλου χρήστη που λαμβάνει μέρος στην επίθεση ονομάζεται προφίλ επίθεσης και σκοπό έχουν τη μείωση της ορθότητας των υποδείξεων (δηλαδή το ίδιο το σύστημα) ή το θετική ή αρνητικό επηρεασμό της βαθμολογίας συγκεκριμένου αντικειμένου (αντικείμενο στόχος) . Η ευρωστία (robustness) *(Groot et al., 2000)* αποτελεί το βαθμό κατά τον οποίο ένα σύστημα ή ένα υποσύστημα λειτουργεί ορθά παρουσία θορύβου στα δεδομένα ή κάτω από πιεστικές περιβαλλοντολογικές συνθήκες. Η πιο διαδεδομένη μετρική για την ευρωστία είναι η μεταβολή πρόβλεψης (Prediction Shift). Η μεταβολή πρόβλεψης για το αντικείμενο στόχος

είναι η διαφορά της μέσης προβλεπόμενης βαθμολογίας του αντικειμένου πριν και μετά την επίθεση για όλους τους χρήστες. Ως μέση μεταβολή πρόβλεψης ονομάζεται η μέση αλλαγή πρόβλεψης πριν και μετά την επίθεση για όλα τα αντικείμενα στόχους.

Ένα εύρωστο σύστημα δεν είναι απόλυτο ότι παρέχει υψηλή ορθότητα υποδείξεων αλλά ούτε και το αντίθετο. Αυτό συμβαίνει κυρίως γιατί η ορθότητα μιας πρότασης είναι υποκειμενικό θέμα και εναπόκειται στα γούστα του κάθε χρήστη. Δηλαδή μπορεί το αντικείμενο στόχος του οποίου η βαθμολογία μετά την επίθεση αυξήθηκε να είναι ποιοτικό και καλό προκαλώντας αύξηση στην ορθότητα των προβλέψεων ενώ η ευρωστία έχει μειωθεί. Στην βιβλιογραφία υπάρχουν αρκετές έρευνες που υποδεικνύουν τη θετική σχέση εμπιστοσύνης και ορθότητας, υπάρχει μόνο μία που να συνδέει την εμπιστοσύνη με την ευρωστία *O' Donovan end Smyth (2005).* Στην εν λόγω έρευνα αποδείχθηκε ότι όταν η εμπιστοσύνη χρησιμοποιηθεί στον υπολογισμό του βάρους που πολλαπλασιάζεται με κάθε βαθμολογία κάθε γείτονα για την παραγωγή της βαθμολογίας υπόδειξης προς τον τρέχοντα χρήστη τα αποτελέσματα είναι βελτιωμένα όσον αφορά την ορθότητα αλλά και συντελεί στην αύξηση της ευρωστίας του συστήματος.

Όμως η χρήση της εμπιστοσύνης είναι περιορισμένη καθώς χρησιμοποιείται μόνο ως συνιστώσα του βάρους ενώ η μετρική της ομοιότητας χρησιμοποιείται τόσο για την επιλογή των γειτόνων αλλά και στον υπολογισμό του βάρους. Επίσης οι *O' Donovan et al* υπολογίζουν την εμπιστοσύνη έμμεσα στηριζόμενοι στην ομοιότητα. Οι παραπάνω λόγοι οδηγούν στο ερώτημα αν η εμπιστοσύνη μπορεί να θεωρηθεί εύρωστη μετρική και αν μπορεί να σταθεί αυτόνομα σε ένα μοντέλο Συνεργατική Μεθόδου ή αν πρέπει να χρησιμοποιείται συνδυαζόμενη με άλλες μετρικές και κυρίως την ομοιότητα. Οι *O'Donovan & Smyth (2006)* ισχυρίζονται ότι η εμπιστοσύνη είναι εύρωστη στο σχετικό άρθρο αλλά πιστεύω ότι χρήζει περεταίρω διερεύνησης λόγω της περιορισμένης χρήσης της στο μοντέλο υπόδειξης και γιατί δεν υπάρχει άλλη ερευνητική εργασία που να συσχετίζει εμπιστοσύνη και ευρωστία.

Βασιζόμενη στις παραπάνω παρατηρήσεις και με σκοπό να ελεγχθούν οι δυνατότητες εμπιστοσύνης στα συστήματα υπόδειξης εξετάστηκε η συγκεκριμένη μετρική ως προς δύο σημαντικούς παράγοντες αξιολόγησης των Συστημάτων Υπόδειξης. Την ορθότητα , δηλαδή κατά πόσο η εμπιστοσύνη βοηθά την υπόδειξη αντικειμένων που να αρέσουν στους χρήστες και την ευρωστία που μετρά το βαθμό αντιμετώπισης του θορύβου στο σύστημα

και κυρίως μετά την εισαγωγή τεχνητού θορύβου (επίθεση με την εισαγωγή κακόβουλων προφίλ).

Για να μπορέσει να απαντηθεί το παραπάνω ερώτημα αλλά για να έχουμε μια αίσθηση του τι συνέπειες έχει μία επίθεση σε ένα σύστημα που από τη φύση του οφείλει να είναι ανοικτό στις διαφορετικές απόψεις πραγματοποιήθηκε πείραμα χρησιμοποιώντας τη βάση δεδομένων του *Epinions.com*

Από το σύνολο του dataset του *Epinions.com* χρησιμοποιήθηκε ο πίνακας στον οποίο είναι καταχωρημένες οι βαθμολογίες των χρηστών για αντικείμενα ώστε να παραχθούν οι υποδείξεις. Οι χρήστες του *Epinions.com* κατά σύμβαση θεωρούνται στα πλαίσια της συγκεκριμένης εργασίας ως γνήσιοι, δηλαδή ότι είναι πραγματικοί και δεν έχουν εισαχθεί στο σύστημα από κάποια κακόβουλη οντότητα που σκοπό έχει τον επηρεασμό κατά το δοκούν των υποδείξεων.

Στη δεύτερη φάση προσθέσαμε στο αρχικό dataset –που όπως αναφέρθηκε θεωρούμε ότι περιέχει μόνο γνήσια προφίλ- κακόβουλα προφίλ ακολουθώντας δύο διαφορετικού τύπου επιθέσεις, την τυχαία (Random) και αυτή του μέσου όρου (average). Στην τυχαία επίθεση η επιλογή των αντικειμένων που θα βαθμολογήσει κάθε προφίλ και οι βαθμολογίες που καταχωρούνται παράγονται τυχαία από μία γεννήτρια αριθμών, ενώ στην επίθεση μέσου όρου η επιλογή των αντικειμένων γίνεται τυχαία και η βαθμολόγησή τους είναι κοντά στον μέσο όρο της βαθμολογίας του αντικειμένου στο σύστημα. Το μέγεθος των προφίλ και στις δύο περιπτώσεις επιλέχθηκε να είναι ίσο με το μέσο προφίλ των γνησίων χρηστών, δηλαδή να αποτελείται από 20 βαθμολογημένα αντικείμενα. Η σύγκριση που ακολούθησε αφορά τα αποτελέσματα των διαφορετικών επιθέσεων μεταξύ τους (τυχαίας και μέσου όρου) αλλά των αποτελεσμάτων πριν και μετά τις επιθέσεις. Η σύγκριση πραγματοποιήθηκε χρησιμοποιώντας μετρικές που αφορούν την ορθότητα και την ευρωστία του συστήματος πριν και μετά τις επιθέσεις.

Το κύριο συμπέρασμα της εργασίας είναι ότι η χρήση της εμπιστοσύνης βελτιώνει σημαντικά τις αποδόσεις της Συνεργατικής Μεθόδου κυρίως όταν χρησιμοποιείται ως κριτήριο επιλογής των γειτόνων κάθε χρήστη. Συγκεκριμένα οδηγεί σε βελτίωση της ορθότητας των υποδείξεων είτε υπάρχει θόρυβος στα δεδομένα είτε όχι αλλά και σε μείωση των αρνητικών επιπτώσεων των επιθέσεων. Στην τυχαία επίθεση η αρνητική επίπτωση αφορά τη μείωση της ορθότητας ενώ στην μέση επίθεση η κύρια αρνητική

επίπτωση αφορά τη μεταβολή της βαθμολογίας του αντικειμένου στόχου προς την τιμή που επιθυμεί ο επιτιθέμενος.

6.1 Συμπεράσματα

Από την εκτέλεση του πειράματος και την αξιολόγηση των ευρημάτων προκύπτουν τα ακόλουθα συμπεράσματα:

- Τα διαφορετικά μοντέλα επίθεσης έχουν και διαφορετικό στόχο. Συγκεκριμένα, οι τυχαίες επιθέσεις οι οποίες επιλέγουν για τη δόμηση των κακόβουλων προφίλ τυχαία αντικείμενα και τυχαίες βαθμολογίες μειώνουν τη συνολική ορθότητα των υποδείξεων δημιουργώντας πρόβλημα στην αξιοπιστία του συστήματος. Όταν οι υποδείξεις που προτάσσονται στους χρήστες δεν ικανοποιούν τις ανάγκες τους και δεν συμβαδίζουν με τις προτιμήσεις τους τότε οι χρήστες χάνουν την πιστότητα τους προς το σύστημα και είναι εύκολο να οδηγηθούν σε ένα ανταγωνιστικό σύστημα.

  Αντίθετα οι επιθέσεις μέσου όρου (average) αυξάνουν τη συνολική ορθότητα των προβλέψεων αφού όλα τα αντικείμενα πλην του αντικειμένου στόχου λαμβάνουν βαθμολογία ίση με το μέσο όρο που λαμβάνουν στο σύστημα. Αυτό έχει μεν θετικό αντίκτυπο στη λειτουργία του συστήματος αλλά υπάρχουν σημαντικές επιπτώσεις γιατί πλέον οι υποδείξεις (κυρίως για το αντικείμενο στόχο της επίθεσης) δεν είναι γνήσιες αλλά έχουν τεχνηέντως ωθηθεί στην επιθυμητή βαθμολογία από τον επιτιθέμενο (αύξησή της ή μείωσή της ανάλογα με το συμφέρον του). Συμπεραίνουμε λοιπόν ότι οι average επιθέσεις δεν έχουν στόχο την εύρυθμη λειτουργία του συστήματος -όπως οι Random- αλλά συγκεκριμένα αντικείμενα (αντικείμενα στόχους) ή προμηθευτές συγκεκριμένων προϊόντων επηρεάζοντας τη φήμη τους (αυξάνοντας ή μειώνοντας την προτασσόμενη βαθμολογία τους) και κατά συνέπεια τα επίπεδα πώλησής τους.

- Το κριτήριο επιλογής των χρηστών που θα δομήσουν τη κάθε γειτονιά είναι σημαντικότερη των συντελεστών βαρών που θα έχουν στον υπολογισμό. Όπως φάνηκε από τα αποτελέσματα ένας έμπιστος χρήστης είναι πολύ πιο αξιόπιστος από έναν όμοιο χρήστη. Αυτό το συμπέρασμα μπορεί να δικαιολογηθεί λόγω του ότι η ομοιότητα

αφορά μόνο παρελθούσες βαθμολογίες και δεν συμπεριλαμβάνει καμία είδους αλληλεπίδραση μεταξύ των χρηστών του δικτύου. Αντίθετα η εμπιστοσύνη εκτός της ομοιότητας που συμπεριλαμβάνει στο μοντέλο της επηρεάζεται και από το πόσο καλές βρίσκει ο ενεργός χρήστης τις προτάσεις του τρέχοντα.

- Ένας έμπιστος χρήστης είναι λιγότερο πιθανό να είναι κακόβουλος σε σχέση με έναν όμοιο χρήστη. Κι αυτό γιατί ο έμπιστος εκτός του ότι έχει κοινές παρελθούσες βαθμολογίες με τον τρέχοντα έχει και θετικό ιστορικό αποδοχής των υποδείξεών του, δηλαδή τον έχει αξιολογήσει ως αξιόπιστο.

- Η ομοιότητα μπορεί να μην είχε καλές αποδόσεις ως κριτήριο επιλογής γειτονιάς αλλά ως συντελεστής βάρους στον υπολογισμό της προτεινόμενη τιμής είναι καλή μετρική δείχνοντας ότι αποτελεί καλό δείκτη για το πόσο «ταιριάζουν» οι χρήστες ενώ η εμπιστοσύνη αποτελεί καλύτερο μέτρο που «αν ταιριάζουν» για αυτό προτείνεται φιλτράρισμα των χρηστών με την εμπιστοσύνη.

- Στη μετρική αξιολόγησης MAE φάνηκε ότι οι καθολικές μεταβλητές είχαν καλύτερα αποτελέσματα ενώ στη μετρική αξιολόγησης MAUE καλύτερα αποτελέσματα είχαν οι τοπικές μεταβλητές. Πιθανόν αυτό να οφείλεται στο διαφορετικό ορισμό τους καθώς MAE αφορά το μέσο λάθος στο σύστημα ενώ το MAUE το μέσο λάθος ανά χρήστη. Δηλαδή στην καθολική μεταβλητή όπως και στο MAE υπάρχει συμφηφισμός των απόψεων ενώ στην τοπική μεταβλητή και στο MAUE κάθε χρήστης έχει το ίδιο βάρος. Συνεπώς φαίνεται εξάρτηση μετρικής αξιολόγησης και μετρικής εμπιστοσύνης.

- Η πρόταση που έγινε για ενίσχυση του ορισμού της εμπιστοσύνης των *O' Donovan & Smyth (2005)* με την ιδέα ότι ο κάθε χρήστης για να είναι έμπιστος πρέπει να έχει συνεισφέρει θετικά στις επιτυχείς προβλέψεις είχε θετικά αποτελέσματα, υποδεικνύοντας ότι η εμπιστοσύνη όπως έχει ορισθεί μέχρι σήμερα έχει κενά και δεν πρέπει να μπορεί ένας «κακός γείτονας» να κρυφτεί σε μια καλή γειτονιά και να ξεγελάσει το δίκτυο.

6.2 Παρατηρήσεις – μελλοντική έρευνα

Τα αποτελέσματα και τα συμπεράσματα που προκύπτουν στα πλαίσια της παρούσας διπλωματικής εργασίας προέρχονται από την επεξεργασία δευτερογενών δεδομένων [4] και όχι πρωτογενών. Κρίνεται σκόπιμο οι παρουσιασθείσες τεχνικές να χρησιμοποιηθούν σε πραγματικές συνθήκες για να αξιολογηθούν ως προς τη συνέπειά τους.

Ένα άλλο ζήτημα που αξίζει να συζητηθεί είναι ότι οι κακόβουλοι δημιουργήθηκαν και εισήχθηκαν στα δεδομένα των γνησίων (Epionions dataset). Αυτό είναι ένα αντικειμενικό πρόβλημα όλων των αντίστοιχων εργασιών καθώς η προσπάθεια αντιμετώπισης αφορά επιθέσεις που δεν είναι «φυσικές» αλλά έχουν δημιουργηθεί στο «εργαστήριο» βάσει συγκεκριμένων μοντέλων της βιβλιογραφίας. Σε περίπτωση που θα μπορούσαμε να συγκρίνουμε τις μετρικές για ευρωστία και ορθότητα σε πραγματικό dataset όπου θα είχε δεχθεί επίθεση και θα γνωρίζαμε ποια είναι τα κακόβουλα προφίλ θα προσέφερε περισσότερο αξιόπιστα αποτελέσματα αλλά κάτι τέτοιο είναι δύσκολο.
Βήμα επέκτασης της συγκεκριμένης έρευνας θα πρέπει να είναι η αξιολόγηση της ευρωστίας και ορθότητας της εμπιστοσύνης που δεν παράγεται πλέον έμμεσα αλλά αποτελεί τη ρητή κρίση κάθε χρήστη προς τον άλλο.

Τέλος θα πρέπει να μελετηθεί η έννοια της εμπιστοσύνης με όρους άλλων επιστημονικών πεδίων πχ ψυχολογίας , οικονομικής θεωρίας κτλ ώστε να προταθεί ένας νέος ορισμός που δεν θα στηρίζεται στο μέτρο της ομοιότητας.

[4] Epionions.com

7 Βιβλιογραφία

1. Sinha, R.R., Swearingen, K.: Comparing recommendations made by online systems & friends. In: DELOS Workshop: Personalization & Recommender Systems in Digital Libraries, 2001.
2. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. & Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In: Proc. of the Conf. on Comp. Supp. Coop. Work, Chapel Hill, NC, 175-186. 1994.
3. Spiekermann, S. & C. Paraschiv . Motivating Human-Agent Interaction: Transferring Insights from Behavioral Marketing to Interface Design. Electronic Commerce Research, 2, pp. 255-285, 2002
4. Felfernig, K. Isak, K. Szabo, & P. Zachar. The VITA Financial Services Sales Support Environment, AAAI/IAAI 2007, pages 1692-1699, Vancouver, Canada, 2007
5. Massa P. & Avesani P.. Trust-aware collaborative filtering for recommender systems. Proceedings of International Conference on Cooperative Information Systems, Agia Napa, Cyprus, 25 Oct – 29 Oct 2004
6. Resnick P., & Varian H. 'Recommender Systems', Communications of the ACM, vol. 40, no. 3, March 1997, pp. 56-58
7. Burke R., Hybrid Recommender Systems: Survey & Experiments, User Modeling & User-Adapted Interaction, v.12 n.4, p.331-370, November 2002
8. Schafer J. Ben , Konstan J., Riedi J., Recommender systems in e-commerce, Proceedings of the 1st ACM conference on Electronic commerce, p.158-166, , Denver, Colorado, United States, November 03-05, 1999
9. Salton G., Automatic Text Processing: The Transformation, Analysis, & Retrieval of Information by Computer. Addison-Wesley, Reading, Massachusetts, 1989
10. Armstrong J.S., Principles of Forecasting A H&book for Researchers & Practitioners. Kluwer Academic, 2001.
11. Lilien, G.L., Kotler Ph., Moorthy, K.S., 1992. Marketing Models. Prentice-Hall, Englewood Cliffs, NJ.
12. Kantor P. B., Ricci F., Rokach L., & Shapira B.. Recommender Systems Handbook. Springer, 2010

13. Herlocker, J., Konstan, J., Riedl, J.: Explaining collaborative filtering recommendations. In: In proceedings of ACM 2000 Conference on Computer Supported Cooperative Work, pp. 241–250 2000.
14. Schafer, J. B., Konstan, J., & Riedl, J. Recommender Systems in E Commerce. In Proceedings of ACM E-Commerce 1999 conference.
15. Frederick F. Reichheld & W. Earl Sasser, Jr 1990. Zero Defections: Quality Comes to Services. Harvard Business School Review, 1990(5): pp. 105-111.
16. Frederick F. Reichheld 1993. Loyalty-Based Management. Harvard Business School Review, 1993(2): pp. 64-73.
17. B. Joseph Pine II, Peppers D. , & Rogers M. Do you want to keep your customers forever? Harvard Business School Review, 1995(2): pp. 103-114
18. Cremonesi P., Picozzi M., & Matera M., "A Comparison of Recommender Systems for Mashup Composition," in Proc. of RSSE 2012. IEEE Press, 2012, p. In print.
19. Montaner, M., L´opez, B., de la Rosa, J.L.: A taxonomy of recommender agents on the internet. Artificial Intelligence Review 19(4), 285–330 2003
20. Balabanović, M. & Shoham, Y.:, Fab: Content-based, collaborative recommendation. Communications of the ACM 40(3), 66-72 1997
21. Goldberg, D. Nichols, D. Oki, B. M. & Terry, D.: Using collaborative filtering to weave an information tapestry. Commun. ACM 35, 12, 1992.
22. Goldberg, K., Roeder, T., Gupta, D. & Chris Perkins. Eigentaste: A Constant Time Collaborative Filtering Algorithm. Information Retrieval, 4(2):133–151, 2001.
23. Breese, J., Heckerman, D., & Kadie, C. (May, 1998). An experimental comparison of collaborative filtering methods. Technical Report MSR-TR-98-12, Microsoft Research, Redmond, WA
24. Krulwich, B.: 'Lifestyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data'. *Artificial Intelligence Magazine* **18**(2), 37-45, 1997
25. Massa, P., Avesani, P.: Trust metrics on controversial users: balancing between tyranny of the majority & echo chambers. International Journal on Semantic Web & Information Systems 3, 39–64, 2007A.
26. Massa, P., Avesani, P.: Trust metrics in recommender systems. In: Golbeck, J. (ed.) Computing with Social Trust, pp. 259-285, 2009.
27. Pazzani, M., Billsus, D.: Content-based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web: Methods & Strategies of Web

Personal Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York (2007) this volume

28. Pazzani M., & Billsus, D. (1997). Learning & revising user profiles: the identification of interesting web sites. Machine Learning 27, 313-331
29. Shardanand, U., Maes, P.: Social information filtering: algorithms for automating ”word of mouth”. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI’95), pp. 210–217 , 1995
30. Chirita, P.A., Nejdl, W., Zamfir, C.: Preventing shilling attacks in online recommender systems. In: WIDM ’05: Proceedings of the 7th annual ACM international workshop on Web information & data management, pp. 67–74. ACM, New York, NY, USA (2005)
31. O'Donovan J. , Smyth B., Trust in recommender systems, Proceedings of the 10th international conference on Intelligent user interfaces, San Diego, California, USA, January 10-13, 2005,
32. Mobasher B. et al., "Towards Trustworthy Recommender Systems: An Analysis of Attack Models & Algorithm Robustness," to be published in ACM Trans. Internet Technology, vol. 7, 2007.
33. Mobasher B., Burke R. & Sandvig J.J., "Model-Based Collaborative Filtering as a Defense against Profile Injection Attacks," Proc. 21st Nat',l Conf. Artificial Intelligence (AAAI 06), AAAI Press, 2006, pp. 1388 and ash,1393.
34. Burke R.. Knowledge-based recommender systems. In Encyclopedia of Library & Information Systems , A. Kent, Ed. Vol. 69. Marcel Dekker, New York, 2000
35. Groot P., Harmelen F. van, & Teije A. ten. Torture tests: a quantitative analysis for the robustness of knowledge-based systems. In *European Workshop on Knowledge Acquisition, Modelling & Management (EKAW’00)*. LNAI Springer-Verlag, October 2000.
36. Sztompka P.. Trust: A Sociological Theory. Cambridge University Press, 1999
37. Golbeck, J., Hendler, J.: FilmTrust: Movie Recommendations using Trust in Web-based Social Networks. In: IEEE Consumer Communication & Network Conference 2006, Las Vegas, USA ,2006
38. Guha, R, & Ravi Kumar ,"Propagation of Trust & Distrust" *Proceedings of the 13th Annual World Wide Web Conference*. 2004.

39. Gray, Elizabeth, Jean-Marc Seigneur, Yong Chen, & Christian Jensen"Trust Propagation in Small Worlds." *Proceedings of the First International Conference on Trust Management*. LNCS 2692, Springer-Verlag. 2003
40. Jøsang, A. The Right Type of Trust for Distributed Systems. In C. Meadows, editor, Proc. of the 1996 New Security Paradigms Workshop. ACM, 1996.
41. Jøsang, Audun, Elisabeth Gray, Michael Kinateder. (2003) " Analysing Topologies of Transitive Trust," *Proceedings of the First International Workshop on Formal Aspects in Security & Trust (FAST 2003)*.
42. Richardson, Matthew, Rakesh Agrawal, Pedro Domingos. (2003) “Trust Management for the Semantic Web,” Proceedings of the Second International Semantic Web Conference. Sanibel Isl&, Florida.
43. Hardin, R.Trust & trustworthiness Russell Sage Foundation, New York, 2002, pp. xxi,234
44. P. Massa & P. Avesani. Trust-aware recommender systems. In RecSys 2007, USA
45. Ziegler C., Lausen G., Spreading Activation Models for Trust Propagation, Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce & e-Service (EEE'04), p.83-97, March 28-31, 2004
46. Golbeck J. & Mannes A., "Using trust & provenance for content filtering on the semantic web," in Proceedings of the WWW'06 Workshop on Models of Trust for the Web (MTW'06), 2006.
47. Massa P. & Bhattacharjee B. Using trust in recommender systems: an experimental analysis. Proceedings of 2nd International Conference on Trust Managment, Oxford, Engl&, 2004.
48. Massa P. & Avesani P. Trust metrics on controversial users: balancing between tyranny of the majority & echo chambers. International Journal on Semantic Web & Information Systems (IJSWIS), 3(1), 2007.
49. Balabanović M. , Shoham Y, Fab: content-based, collaborative recommendation, Communications of the ACM, v.40 n.3, p.66-72, March 1997
50. Mobasher B. , Burke R. , Bhaumik R., Sandvig J. J., Attacks & Remedies in Collaborative Recommendation, IEEE Intelligent Systems, v.22 n.3, p.56-63, May 2007
51. Burke R. , Mobasher B. , Williams Ch. , Bhaumik R, Classification features for attack detection in collaborative recommender systems, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery & data mining, August 20-

23, 2006, Philadelphia, PA, USA

52. Dellarocas Chr., Immunizing online reputation reporting systems against unfair ratings & discriminatory behavior, Proceedings of the 2nd ACM conference on Electronic commerce, p.150-157, October 17-20, 2000, Minneapolis, Minnesota, United States
53. Hoffman K. , Zage D. , Rotaru Ch., A survey of attack & defense techniques for reputation systems, ACM Computing Surveys (CSUR), v.42 n.1, p.1-31, December 2009
54. O'Mahony M., Hurley N., Silvestre G., Promoting Recommendations: An Attack on Collaborative Filtering, Proceedings of the 13th International Conference on Database & Expert Systems Applications, p.494-503, September 02-06, 2002
55. Williams A., Mobasher B., Burke R.: Defending recommender systems: detection of profile injection attacks. Service Oriented Computing & Applications pp. 157–170 , 2007
56. Ray & A. Mahanti. Strategies for effective shilling attacks against recommender systems. In Proc. 2nd Int'l Workshop on Privacy, Security, & Trust in KDD, Las Vegas, NV, USA, 2008
57. O'Mahony M., Hurley N., Kushmerick N., Silvestre G., Collaborative recommendation: A robustness analysis, ACM Transactions on Internet Technology (TOIT), v.4 n.4, p.344-377, November 2004
58. Mehta B., Nejdl W., Unsupervised strategies for shilling detection & robust collaborative filtering, User Modeling & User-Adapted Interaction, v.19 n.1-2, p.65-97, February 2009
59. Mehta B., Hofmann T. , Fankhauser P., Lies & propaganda: detecting spam users in collaborative filtering, Proceedings of the 12th international conference on Intelligent user interfaces, January 28-31, 2007, Honolulu, Hawaii, USA
60. Chirita P. , Nejdl W. , Zamfir C., Preventing shilling attacks in online recommender systems, Proceedings of the 7th annual ACM international workshop on Web information & data management, November 04-04, 2005, Bremen, Germany
61. O'Donovan J., Smyth B.,, Is trust robust?: an analysis of trust-based recommendation, Proceedings of the 11th international conference on Intelligent user interfaces, January 29-February 01, 2006, Sydney, Australia
62. Lam S. , Riedl J., Shilling recommender systems for fun & profit, Proceedings of the 13th international conference on World Wide Web, May 17-20, 2004, New York, NY,

USA

63. O'Mahony M. , Hurley N., & Silvestre G.. Efficient & secure collaborative filtering through intelligent neighbour selection. In Proceedings of the 16th European Conference on Artificial Intelligence (ECAI'04), pages 383--387, August 23-27 2004
64. O'Mahony M. , Hurley N., Silvestre G., Utility-based neighbourhood formation for efficient & robust collaborative filtering, Proceedings of the 5th ACM conference on Electronic commerce, May 17-20, 2004, New York, NY, USA
65. Sarwar B, Karypis G, Konstan J, Reidl J, Item-based collaborative filtering recommendation algorithms, Proceedings of the 10th international conference on World Wide Web, p.285-295, May 01-05, 2001, Hong Kong, Hong Kong
66. O'Mahony P. , Hurley N , Silvestre G, Detecting noise in recommender system databases, Proceedings of the 11th international conference on Intelligent user interfaces, January 29-February 01, 2006, Sydney, Australia
67. Adomavicius G. , Tuzhilin A., Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art & Possible Extensions, IEEE Transactions on Knowledge & Data Engineering, v.17 n.6, p.734-749, June 2005
68. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: The Adaptive Web, pp. 291–324. Springer Berlin / Heidelberg , 2007
69. Polat, H., Du, W.: Privacy-preserving collaborative filtering using Randomized perturbation techniques. In: Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003), 19-22 December 2003, Melbourne, Florida, USA, pp. 625–628 , 2003
70. Takacs G., Pil´aszy I., N´emeth B., Tikk D.: Scalable collaborative filtering approaches for large recommender systems. J. Mach. Learn. Res. 10, 623–656 (2009)
71. Blaze M., Feigenbaum J., & Lacy J.. Decentralized trust management. In Proceedings of the 17th Symposium on Security & Privacy, pages 164-173, Oakl&, CA, USA, May 1996
72. Page L., Brin S., Motwani R., & Winograd T.. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.
73. Mui L., Mohtashemi M., & Halberstadt A.. A computational model of trust & reputation. In Proceedings of the 35th Hawaii International Conference on System

Sciences, pages 188-196, Big Isl&, HI, USA, January 2002.

74. Eschenauer L., Gligor V., & Baras J.. On trust establishment in mobile ad-hoc networks. Technical Report MS 2002-10, Institute for Systems Research, University of Maryl&, MD, USA, October 2002.
75. Asch,S.E.:Studies of independence & conformity: a minority of one against a unanimous majority. Pschol. Monogr. 70 ,1956
76. Cosley, D., Lam, S. K., Albert, I., Konstan, J., & Riedl, J.. Is seeing believing? How recommender systems influence users' opinions. In Proceedings of CHI 2003: Human Factorsin Computing Systems (pp. 585-592). New York: ACM Press ,2003
77. Avesani, P., Massa, P., Tiella, R.: Moleskiing.it: a trust-aware recommender system for ski mountaineering. International Journal for Infonomics 2005
78. Golbeck, J.: Generating predictive movie ratings from trust in social networks. In: Stølen, K, Winsborough, W.H., Martinelli, F., Massacci, F. (eds.) Lecture Notes in Computer Science 3986, pp. 93-104 2006
79. Golbeck, J., Mannes, A.: Using trust & provenance for content filtering on the semantic web. In: Proc. of the WWW06 Models of Trust for the Web Workshop (2006)
80. Mobasher, B., Burke, R., Bhaumik, R., Williams, C.: Effective attack models for shilling item-based collaborative filtering system. In Proceedings of the 2005 WebKDD Workshop 2005
81. Breese J. , Heckerman D., Kadie C., Empirical analysis of predictive algorithms for collaborative filtering, Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, p.43-52, July 24-26, 1998, Madison, Wisconsin
82. Zhang F., Bai L., Gao F., A user trust-based collaborative filtering recommendation algorithm, Proceedings of the 11th international conference on Information & Communications Security, December 14-17, 2009, Beijing, China
83. Montaner M. , López B. , Lluís de la Rosa J., Developing trust in recommender agents, Proceedings of the first international joint conference on Autonomous agents & multiagent systems: part 1, July 15-19, 2002, Bologna, Italy
84. Golbeck J. , Hendler J., Computing & applying trust in web-based social networks, University of Maryland at College Park, College Park, MD, 2005
85. Celma O., & Lamere, P. Music Recommendation Tutorial. Presented at the 8th International Conference on Music Information Retrieval, Vienna, Austria , 2007

.