



Πανεπιστήμιο Πειραιώς
Τμήμα Ψηφιακών Συστημάτων
Π.Μ.Σ. «Πληροφορικά Συστήματα & Υπηρεσίες»

Κατεύθυνση «Μεγάλα Δεδομένα &
Αναλυτική»

Διπλωματική Εργασία

Νέες Μετρικές σε Συστήματα Συστάσεων
Επιβλέπουσα Καθηγήτρια: Μαρία Χαλκίδα

Χρήστος Σωτήρης Καραίσκος
Μεταπτυχιακός Φοιτητής
Αριθμός Μητρώου: me2014

Σεπτέμβριος 2023

Περίληψη

Οι έρευνες πάνω στα συστήματα συστάσεων έχουν επικεντρωθεί σε παραδοσιακές μετρικές όπως η ακρίβεια και στο πόσο κοντά είναι η προβλεπόμενη αξιολόγηση του συστήματος σε σχέση με την πραγματική αξιολόγηση του χρήστη. Ωστόσο τα τελευταία χρόνια έχει γίνει ευρέως αποδεκτό ότι υπάρχουν και άλλες μετρικές που μπορούν να έχουν αντίκτυπο στην ποιότητα των προτάσεων ενός συστήματος συστάσεων. Αυτές οι μετρικές έχουν να κάνουν με το πόσο ποικίλες ή καινούργιες είναι οι προτάσεις για το χρήστη και πόσο μεγάλο εύρος καλύπτουν από τα συνολικά διαθέσιμα αντικείμενα που μπορούν να προταθούν.

Στην παρούσα εργασία γίνεται ανάλυση τριών μετρικών εξ αυτών και πιο συγκεκριμένα των diversity, novelty και coverage. Παρουσιάζονται οι ορισμοί τους και διάφορες τεχνικές οπτικοποίησης που συναντάμε στη βιβλιογραφία όπως επίσης και σχετικές εργασίες που επικεντρώνονται στη βελτίωση της απόδοσης των μετρικών αυτών. Μέσα από τη βιβλιογραφία αντιληφθήκαμε ότι δεν είναι ξεκάθαρο αν και πως η βελτίωση κάθε μετρικής επηρεάζει τις υπόλοιπες.

Για αυτό το λόγο χρησιμοποιώντας μια τεχνική αναδιάταξης με βάση το diversity κάναμε διάφορα offline πειράματα στο movielens σύνολο δεδομένων και προσπαθήσαμε να συγκρίνουμε τις συσχετίσεις που έχει η οπτικοποιημένη μετρική diversity σε σχέση με τις υπόλοιπες μετρικές. Η τεχνική αναδιάταξης εφαρμόστηκε πάνω σε state of the art αλγόριθμους των δύο βασικών κατηγοριών του συνεργατικού φιλτραρίσματος. Από τα αποτελέσματα μας αν και σε περιορισμένο εύρος βρήκαμε θετική συσχέτιση μεταξύ του diversity και του novelty και αρνητική συσχέτιση μεταξύ του diversity και της ακρίβειας.

Ευχαριστίες

Θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες στην επιβλέπουσα καθηγήτρια κυρία Μαρία Χαλκίδη, για την καθοδήγηση και την άμεση ανταπόκριση σε τυχόν απορίες κατά τη διάρκεια της εκπόνησης αυτής της διπλωματικής εργασίας. Επίσης, θέλω να ευχαριστήσω την οικογένειά μου και τους κοντινούς μου ανθρώπους για την υποστήριξη που μου παρείχαν καθ' όλη τη διάρκεια αυτής της πορείας. Τέλος δεν μπορώ να παραλείψω να ευχαριστήσω το πανεπιστήμιο Πειραιώς και την ευρύτερη κοινότητα των data scientists για τις πολύτιμες γνώσεις που μου προσέφεραν από την πρώτη έως και την τελευταία μέρα της συμμετοχής μου στο μεταπτυχιακό πρόγραμμα.

Περιεχόμενα

1	Εισαγωγή	1
1.1	Αναγκαιότητα Χρήσης Νέων Μετρικών	3
1.2	Κίνητρο Έρευνας	4
1.3	Ορισμός Προβλήματος	5
1.4	Στόχοι Εργασίας	5
1.5	Δομή Εργασίας	6
2	Συνεργατικό Φιλτράρισμα	7
2.1	Ιστορική Αναδρομή	7
2.2	Κύριες Κατηγορίες Συνεργατικού Φιλτραρίσματος	9
2.2.1	Memory-Based	9
2.2.2	Model-Based	12
2.3	Memory-Based Αλγόριθμοι	13
2.3.1	K-Nearest Neighbors	13
2.3.1.1	Προσέγγιση κοντινών γειτόνων	14
2.3.1.2	kNN Basic	15
2.3.1.3	kNN with Means	15
2.3.1.4	kNN with Z-Score	16
2.3.2	Μετρικές Απόστασης	16
2.4	Model-Based Αλγόριθμοι	19
2.4.1	Slope One	19
2.4.2	SVD	21
2.4.3	SVD++	23
2.5	Βαθμολογίες Χρηστών	24
2.6	Κανονικοποίηση Βαθμολογιών	25
2.7	Προβλήματα Συνεργατικού Φιλτραρίσματος	27
2.7.1	Αραιότητα Πίνακα	27
2.7.2	Εγγραφή Νέου Χρήστη/Αντικειμένου	28
3	Μετρικές Αξιολόγησης	29
3.1	Diversity	29
3.1.1	Reranking Τεχνικές Diversity	30
3.2	Novelty	32
3.2.1	Τεχνικές Βελτίωσης Novelty	33
3.3	Coverage	35
3.3.1	Συσχέτιση Coverage με τις Υπόλοιπες Μετρικές	36
3.4	Σχετικές Δουλειές πάνω στη Χρήση Νέων Μετρικών	37
3.5	Feedback Χρηστών στις Νέες Μετρικές	38
4	Μεθοδολογία	40
4.1	Αλγόριθμος Αναδιάταξης	40
4.2	Diversity Τεχνική Αναδιάταξης	41
4.3	Ακολουθούμενη Προσέγγιση	42

5	Πειραματική Διαδικασία	45
5.1	Αποτελέσματα	46
5.2	Σχολιασμός Αποτελεσμάτων	51
6	Επίλογος και Μελλοντική Δουλειά	52
	Αναφορές	53

1 Εισαγωγή

Το ίντερνετ παίζει πλέον πολύ σημαντικό ρόλο στις καθημερινές μας εμπορικές συναλλαγές. Καθημερινά εκατομμύρια άνθρωποι περιηγούνται στο διαδίκτυο για να πραγματοποιήσουν μια αγορά. Με το πέρασμα των χρόνων η παροχή προϊόντων και υπηρεσιών που είναι διαθέσιμα στο ίντερνετ άρχισε να αυξάνεται με εκθετικούς ρυθμούς, κάτι που έδωσε στους χρήστες πολλές επιλογές αλλά παράλληλα δημιούργησε μια μεγάλη πρόκληση. Έπρεπε να βρεθεί ένας τρόπος ώστε από όλες αυτές τις επιλογές ο εκάστοτε χρήστης να μπορεί εύκολα να βρίσκει προϊόντα και υπηρεσίες ανάλογα με τις προτιμήσεις του. Αυτό είχε σαν αποτέλεσμα την ανάπτυξη των συστημάτων συστάσεων.

Στις μέρες μας τα συστήματα συστάσεων είναι διαδεδομένα σε πάρα πολλά διαφορετικά πεδία όπως για παράδειγμα σε ιστοσελίδες ρούχων, ηλεκτρικών συσκευών, σε εφαρμογές μουσικής, ταινιών και άλλα. Τα συστήματα συστάσεων είναι μια τεχνολογία εξατομικευμένου φιλτράρισματος πληροφοριών, που στοχεύουν να παρέχουν στο χρήστη προσωποποιημένες προτάσεις έτσι ώστε να ικανοποιήσουν τις ανάγκες του [4].

Για πολλά χρόνια οι μελέτες στα συστήματα συστάσεων προσπαθούσαν να δώσουν απάντηση στο πως να αξιολογήσουν την αξία ενός συστήματος συστάσεων. Τι είναι αυτό που θα ορίσει αν ένα σύστημα είναι πετυχημένο ή όχι. Σε πολλές βιβλιογραφικές μελέτες οι περισσότερες προσεγγίσεις επικεντρώθηκαν στην ακρίβεια που παρείχαν οι προτάσεις προς ένα χρήστη. Προσπαθούσαν δηλαδή να υπολογίσουν πόσο κοντά είναι η τιμή που προβλέπει το σύστημα σε μία αξιολόγηση ενός χρήστη σε σχέση με την πραγματική αξιολόγηση του χρήστη [12].

Όμως, πολλές νέες μελέτες έχουν αρχίσει να μελετάνε και νέες μετρικές πέρα από την ακρίβεια. Για παράδειγμα στο [1] μέσα από μία σειρά επιχειρημάτων υποστηρίζεται ότι το να στηρίζεται ένα σύστημα συστάσεων αποκλειστικά στην ακρίβεια καταλήγει να κάνει κακό στο ίδιο το σύστημα.

Στην εισαγωγή του αναφέρει ως παράδειγμα ένα σύστημα συστάσεων ταξιδιωτικών προορισμών επιχειρηματολογώντας ότι αν το σύστημα δουλεύει αποκλειστικά με γνώμονα την ακρίβεια είναι σα να εκπαιδεύεται να προτείνει στο χρήστη προορισμούς που έχει ήδη επισκεφτεί και καταλήγει ότι είναι επιτακτική ανάγκη να ληφθούν υπόψιν και άλλες μετρικές [1]. Μερικές από αυτές τις μετρικές με τις οποίες θα ασχοληθούμε στην παρούσα εργασία είναι το diversity, το novelty και το coverage.

Το diversity είναι μία έννοια που προέρχεται από το πεδίο ανάκτησης πληροφορίας (Information Retrieval) [56]. Βασικό αίτιο για την ανάπτυξη της ιδέας αυτής ήταν η πεποίθηση ότι το να παρουσιάζεις στο χρήστη προτάσεις με αποκλειστικό κριτήριο την ακρίβεια μπορεί να μην ικανοποιήσει πλήρως το χρήστη λόγω της ομοιότητας των αντικειμένων.

Έτσι ξεκίνησαν να αναπτύσσονται τρόποι ώστε να μελετηθεί η σχέση μεταξύ ακρίβειας και ποικιλίας προϊόντων και ήταν πλέον κοινή πεποίθηση ότι μία λίστα προϊόντων που μπορεί να χάνει σε ακρίβεια και να δίνει μεγαλύτερη έμφαση στην ποικιλία μπορεί να αφήσει τους χρήστες περισσότερο ευχαριστημένους [2].

Το novelty θεωρείται μία ποιοτική μετρική που στόχο έχει να προτείνει κάτι νέο στο χρήστη, δηλαδή ένα αντικείμενο που δεν έχει ξαναδεί και το θεωρεί άγνωστο. Βασιζόμενο σε αυτό τον ορισμό μπορούμε να διακρίνουμε δύο σημαντικές πτυχές για το novelty. Πρώτον ένα νέο αντικείμενο δεν πρέπει να έχει ανακαλυφθεί από το χρήστη και δεύτερον θα πρέπει να είναι ενδιαφέρον, χρήσιμο και σχετικό με τα ενδιαφέροντα του χρήστη. Ο τρόπος προσδιορισμού για το αν ένα αντικείμενο θεωρείται νέο συνδέεται με την δημοτικότητα του. Για παράδειγμα σε ένα σύστημα με αξιολογήσεις χρηστών μπορούμε να καθορίσουμε πόσο δημοφιλές είναι ένα αντικείμενο μετρώντας τον αριθμό των βαθμολογιών. Όσο λιγότερες έχει τόσο πιθανότερο είναι να προταθεί σαν νέο στο χρήστη [2].

Σε κάποιες μελέτες της βιβλιογραφίας η έννοια του novelty συγχέεται με την έννοια του serendipity. Όμως, υπάρχει διαφορά μεταξύ τους καθώς το serendipity είναι μία έννοια που προϋποθέτει τόσο το να είναι νέο ένα αντικείμενο για το χρήστη και ταυτόχρονα να του προκαλεί έκπληξη. Για περισσότερες πληροφορίες σχετικά με το serendipity ο αναγνώστης μπορεί να ανατρέξει στο [3].

Το coverage είναι μία μετρική που υποδηλώνει το ποσοστό κάλυψης που έχουν οι προτάσεις που παράγει το σύστημα συστάσεων σε σχέση με τα συνολικά διαθέσιμα αντικείμενα. Σε αυτή την περίπτωση αφενός μεν θέλουμε το σύστημα να είναι ικανό να παράγει μεγάλο εύρος αντικειμένων από τον κατάλογο αφετέρου αυτές οι προτάσεις να είναι αποδοτικές όταν προταθούν στο χρήστη [3]. Η συγκεκριμένη μετρική συνδέεται συχνά στη βιβλιογραφία με το diversity και το novelty και όπως είναι εύκολα αντιληπτό όσο μεγαλύτερη είναι η κάλυψη σε ένα σύστημα τόσο καλύτερα για τον χρήστη αλλά και για την υπηρεσία που παρέχει το σύστημα συστάσεων στο χρήστη.

Οι αλγόριθμοι των συστημάτων συστάσεων χωρίζονται σε φιλτράρισμα βάσει περιεχομένου, συνεργατικό φιλτράρισμα και υβριδικό φιλτράρισμα [5]. Παρακάτω θα γίνει μία συνοπτική περιγραφή για τον τρόπο που λειτουργεί το συνεργατικό φιλτράρισμα καθώς στη μελέτη μας θα στηριχθούμε σε κάποιους από τους αλγόριθμους του συνεργατικού φιλτραρίσματος.

Το συνεργατικό φιλτράρισμα προτείνει αντικείμενα βασιζόμενο σε χρήστες που έχουν παρόμοιες προτιμήσεις χωρίς να απαιτούνται πληροφορίες για τα αντικείμενα. Προϋποθέτει πληροφορίες για τις ήδη υπάρχουσες αξιολογήσεις από τους χρήστες πάνω σε κάποια αντικείμενα [7]. Ο χρήστης αξιολογεί ένα αντικείμενο σε μία κλίμακα βαθμίδας ανάλογα με την προτίμησή του. Όσο περισσότερο του αρέσει κάτι τόσο υψηλότερη βαθμολογία αναμένεται να δώσει ή μπορεί να έχει την δυαδική επιλογή μου αρέσει/δε μου αρέσει ένα αντικείμενο.

Με βάση τις αξιολογήσεις στα αντικείμενα του χρήστη, θα του προταθούν αντικείμενα τα οποία έχουν υψηλή βαθμολογία ή αρέσουν σε όμοιες ομάδες χρηστών. Ας υποθέσουμε για παράδειγμα ότι ο χρήστης A έχει αξιολογήσει θετικά τις ταινίες A, B, Γ και όμοιες ομάδες χρηστών έχουν επίσης βαθμολογήσει θετικά τις ταινίες A,B,Γ και την ταινία Δ. Τότε το σύστημα θα προτείνει στον χρήστη A την ταινία Δ θεωρώντας ότι είναι πολύ πιθανό να βρίσκεται στις προτιμήσεις του χρήστη A.

1.1 Αναγκαιότητα Χρήσης Νέων Μετρικών

Στην εισαγωγή μιλήσαμε για την τάση που υπάρχει τα τελευταία χρόνια να μελετώνται και να αναπτύσσονται νέες μετρικές. Σε αυτή την ενότητα θα εξηγήσουμε γιατί είναι σημαντικό να λαμβάνονται υπόψη και άλλες μετρικές πέρα από την ακρίβεια. Η επικέντρωση αποκλειστικά και μόνο στις μετρικές ακρίβειας ενδέχεται να έχει αρνητικό αντίκτυπο στην αξιοπιστία και την απόδοση των συστημάτων συστάσεων [82]. Για παράδειγμα μπορεί να υπάρχει πολύ μεγάλη προσαρμογή σε δημοφιλή αντικείμενα. Το σύστημα συστάσεων μπορεί να προτείνει τα ίδια κοινότυπα και πολύ γνωστά αντικείμενα στους περισσότερους χρήστες κάτι που οδηγεί σε έλλειψη εξατομίκευσης και διαφορετικότητας στις προτάσεις [65].

Επιπλέον ένα σύστημα επικεντρωμένο στην ακρίβεια μπορεί να δώσει προτεραιότητα σε αντικείμενα που είναι κυρίως αγνοώντας τα λιγότερο γνωστά. Για παράδειγμα σε ένα σύστημα ταινιών μπορεί να δοθεί προτεραιότητα σε blockbuster ταινίες αγνοώντας έτσι ταινίες που βρίσκονται στην μακριά ουρά και είναι λιγότερο δημοφιλείς. Αυτό μπορεί να προκαλέσει δυσαρέσκεια σε χρήστες που έχουν πιο ιδιαίτερες προτιμήσεις καθώς το σύστημα δε μπορεί να ανταποκριθεί στις ανάγκες τους. Άμεση συνέπεια αυτού είναι και ο περιορισμός στην εξερεύνηση επιλογών νέων και λιγότερο γνωστών αντικειμένων. Στον χρήστη μπορεί να προτείνονται μόνιμα γνωστές προτάσεις αποτρέποντας τον έτσι να ανακαλύψει αντικείμενα που αποτελούν κρυφά διαμάντια [82].

Από τα παραπάνω γίνεται φανερό η σημαντικότητα που μπορεί να έχουν σε ένα σύστημα συστάσεων μετρικές πέραν της ακρίβειας όπως οι diversity, novelty και coverage. Με την βοήθεια του diversity μπορούμε να προτείνουμε στο χρήστη ένα μεγαλύτερο εύρος αντικειμένων που θα είναι λιγότερο γνωστά αλλά θα παραμένουν σχετικά με τα ενδιαφέροντα του χρήστη. Για παράδειγμα ένα σύστημα που προτείνει ταινίες θα μπορούσε να ενσωματώσει τη διαφορετικότητα προτείνοντας κάποιες επιλογές διεθνών ταινιών από διαφορετικές χώρες και κουλτούρες. Έτσι ο χρήστης μπορεί να ανακαλύψει μία ποικιλία κινηματογραφικών εμπειριών πέραν των γνωστών ταινιών που παράγει το Χόλιγουντ. Αυτό θα ικανοποιήσει το χρήστη και θα αυξήσει την εμπιστοσύνη του στο σύστημα συστάσεων [81].

Αντίστοιχα με τη βοήθεια του novelty ο χρήστης έχει την ευκαιρία να έρθει σε επαφή με αντικείμενα που είναι λιγότερο δημοφιλή ανάμεσα στους χρήστες αλλά μπορεί να είναι σχετικές με τα ενδιαφέροντα του. Δίνει θα λέγαμε μεγαλύτερη έμφαση στα αντικείμενα της μακριάς ουράς που δεν είναι τόσα γνωστά στο ευρύ κοινό. Παραδείγματος χάριν σε ένα σύστημα που προτείνει ταινίες, το novelty θα μπορούσε να προτείνει ταινίες που έχουν λάβει την αποδοχή των κριτικών σε διάφορα φεστιβάλ κινηματογράφου αλλά ενδεχομένως δεν έχουν γίνει ευρέως γνωστές. Αυτό μπορεί να παρακινήσει το ενδιαφέρον των χρηστών που ψάχνουν πιο ιδιαίτερες επιλογές και δεν είναι εύκολο να τις βρουν αλλιώς [83].

Η μετρική coverage έχει στόχο την μεγαλύτερη κάλυψη αντικειμένων του συνόλου δεδομένων κάτι που ενισχύεται από την αύξηση του diversity ή του novelty. Συμπερασματικά μπορούμε να πούμε ότι το βέλτιστο αποτέλεσμα σε ένα σύστημα συστάσεων μπορεί να επιτευχθεί συνδυάζοντας την ακρίβεια και μετρικές όπως το diversity, το novelty και το coverage. Με μία καλή ισορροπία ανάμεσα σε αυτές τις μετρικές τα συστήματα προτάσεων μπορούν να ενισχύσουν τη συμμετοχή των χρηστών, να ενθαρρύνουν την τυχαία ανακάλυψη και να καλύψουν μια ευρύτερη γκάμα προτιμήσεων των χρηστών [82].

1.2 Κίνητρο Έρευνας

Ερευνώντας τη βιβλιογραφία παρατηρήσαμε ότι το μεγαλύτερο κομμάτι της έρευνας πάνω στα συστήματα συστάσεων ήταν επικεντρωμένο σε παραδοσιακές μετρικές αξιολόγησης όπως η ακρίβεια. Με τον όρο ακρίβεια εννοούμε την προσπάθεια των μοντέλων να προβλέψουν πόσο κοντά είναι μία προβλεπόμενη αξιολόγηση στην πραγματική αξιολόγηση.

Όμως, τα τελευταία χρόνια έχει ξεκινήσει μία προσπάθεια μελέτης νέων μετρικών που δεν έχουν σαν κυρίαρχο στόχο να επιτυχούν την ακρίβεια αλλά προσπαθούν να λάβουν υπόψιν την ποικιλομορφία των προτεινόμενων προϊόντων και πόσο καινούργια είναι αυτά για το χρήστη.

Σκοπός αυτής της διπλωματικής είναι να παρουσιάσει τις σημαντικότερες εξ αυτών των νέων μετρικών όπως το coverage, το diversity και το novelty. Θα κάνουμε μία βιβλιογραφική αναφορά στους ορισμούς αυτών των μετρικών όπως και σε κάποιες τεχνικές οπτικοποίησης που έχουν υλοποιηθεί και αναπτυχθεί στη βιβλιογραφία για την καλύτερη απόδοση τους. Παράλληλα θα αναπτύξουμε μία από τις τεχνικές οπτικοποίησης του diversity και θα δούμε πως αλληλεπιδρά στο σύνολο δεδομένων σε σχέση με τις υπόλοιπες μετρικές.

Μέσα από αυτή την εργασία ευελπιστούμε ο αναγώστης να αποκτήσει μία καλή εικόνα για τις μετρικές πέραν της ακρίβειας, να κατανοήσει τους λόγους που ένα σύστημα συστάσεων δεν είναι απαραίτητα καλό όταν βασίζεται αποκλειστικά στην ακρίβεια αλλά πρέπει να λάβει υπόψιν και πόσο ενδιαφέρουσα ή καινούργια θα φαίνεται στο χρήστη μία πρόταση..

1.3 Ορισμός Προβλήματος

Το πρόβλημα μας χωρίζεται σε δύο άξονες. Το πρώτο κομμάτι αφορά την επιλογή των αλγορίθμων συνεργατικού φιλτραρίσματος ώστε να παράγουν προτάσεις σχετικές με τις προτιμήσεις των χρηστών. Το δεύτερο κομμάτι αφορά την υλοποίηση ενός αλγορίθμου αναδιάταξης με έμφαση στην αύξηση του diversity και τις επιπτώσεις που έχει στις 5 μετρικές που εξετάζουμε rmse, f-measure, diversity, novelty και coverage.

Έστω ότι έχουμε ένα χρήστη $u \in U$ και ένα σύνολο αντικειμένων $i \in I$. Αρχικά θέλουμε να βρούμε μία λίστα αντικειμένων που θα είναι σχετικές με τις προτιμήσεις του χρήστη και θα του αρέσουν. Οι προβλέψεις αυτές θα προκύψουν από το προφίλ του χρήστη, βλέποντας τα αντικείμενα με τα οποία έχει αλληλεπιδράσει στο ιστορικό του κάτι το οποίο θα μας δώσει ενδείξεις για τα ενδιαφέροντα του.

Η σχετικότητα μεταξύ ενός χρήστη και ενός αντικειμένου καθορίζεται από την βαθμολογία του χρήστη με τη σχέση $r(u,i)$, όπου μπορεί να είναι δυαδική - του αρέσει ή δεν του αρέσει ή κλιμακωτή, για παράδειγμα από 1 έως 10 αστέρια. Αρχικά θα ορίσουμε τη συνάρτηση $\hat{r}: U \times I \rightarrow [1, r_{max}]$ η οποία θα προβλέπει τις αξιολογήσεις του χρήστη u για το αντικείμενο i μέσω της σχέσης $\hat{r}(u,i)$.

Στη συνέχεια έχοντας τις λίστες προτάσεων που μας έδωσαν οι αλγόριθμοι με βάση τις αξιολογήσεις του χρήστη θα χρησιμοποιήσουμε μία τεχνική οπτικοποίησης του diversity και θα τρέξουμε ένα greedy αλγόριθμο αναδιάταξης στα αντικείμενα της λίστας αυτής με σκοπό να δώσουμε έμφαση όχι μόνο στην ακρίβεια αλλά και στο diversity. Σκοπός μας σε αυτό το κομμάτι είναι να συγκρίνουμε πως επηρεάζεται η τελική λίστα προτάσεων σε σχέση με την αρχική και αν καταφέρνουμε να αυξήσουμε κάποια από τις υπόλοιπες μετρικές διατηρώντας παράλληλα μια ισορροπία στην ακρίβεια των προτάσεων μας.

1.4 Στόχοι Εργασίας

Η παρακάτω εργασία έχει στόχο να καλύψει τα παρακάτω:

- Περιγραφή του συνεργατικού φιλταρίσματος και κάποιων state of the art αλγορίθμων του (memory-based και model-based αλγόριθμοι).
- Ανάλυση και ορισμός των μετρικών αξιολόγησης diversity, novelty, coverage μέσα από τη βιβλιογραφία.
- Περιγραφή του reranking αλγορίθμου και της αντικειμενικής συνάρτησης που θα χρησιμοποιήσουμε για να τρέξουμε το πείραμα μας.
- Διεξαγωγή πειράματος στο πραγματικό σύνολο δεδομένων MovieLens με τη βοήθεια της βιβλιοθήκης surprise της γλώσσας προγραμματισμού Python και σύγκριση των μετρικών αξιολόγησης.
- Παρουσίαση και αξιολόγηση των αποτελεσμάτων του πειράματος. Πως τελικά επηρεάζει η αύξηση μίας μετρικής την απόδοση των υπολοίπων;

1.5 Δομή Εργασίας

Αυτή η εργασία αποτελείται συνολικά από 5 κεφάλαια:

- Στο κεφάλαιο 1 γίνεται μια συνοπτική εισαγωγή στα συστήματα συστάσεων, αναφερόμαστε στους στόχους της εργασίας και ορίζουμε το πρόβλημα πάνω στο οποίο θα δουλέψουμε.
- Στο κεφάλαιο 2 γίνεται ανασκόπηση της εξέλιξης του συνεργατικού φιλτραρίσματος με το πέρασμα των χρόνων, περιγράφονται οι δύο κύριες κατηγορίες του και οι κυριότεροι αλγόριθμοι ανά κατηγορία.
- Στο κεφάλαιο 3 περιγράφονται οι μετρικές diversity, novelty και coverage που θα χρησιμοποιήσουμε για να αξιολογήσουμε τους παραπάνω αλγορίθμους. Επιπλέον παρουσιάζονται τεχνικές και σχετικές εργασίες από τη βιβλιογραφία για τη βελτίωση της απόδοσης των μετρικών αυτών.
- Στο κεφάλαιο 4 περιγράφεται ο αλγόριθμος αναδιάταξης που χρησιμοποιήσαμε για το πείραμα μας και η προσέγγιση που ακολουθήθηκε για την αξιολόγηση των μετρικών.
- Στο κεφάλαιο 5 ακολουθεί η διεξαγωγή πειράματος στο Movielens dataset με τη βοήθεια της βιβλιοθήκης Surprise και η παρουσίαση των αποτελεσμάτων για την απόδοση κάθε μετρικής.
- Στο κεφάλαιο 6 γίνεται μία σύνοψη της δουλειάς μας και μελλοντικές προτάσεις για περαιτέρω έρευνα.

2 Συνεργατικό Φιλτράρισμα

Σε αυτό το κεφάλαιο θα γίνει επισκόπηση του συνεργατικού φιλτραρίσματος. Ξεκινώντας θα γίνει μία μικρή ιστορική αναδρομή για την αρχή και την εξέλιξη του συνεργατικού φιλτραρίσματος. Στη συνέχεια θα αναλύσουμε τις δύο μεγάλες κατηγορίες του συνεργατικού φιλτραρίσματος και τέλος θα μιλήσουμε για τον τρόπο αξιολόγησης των χρηστών αλλά και τις προκλήσεις που καλείται να ξεπεράσει ένα σύστημα συστάσεων ώστε να είναι αποδοτικό.

2.1 Ιστορική Αναδρομή

Ο όρος συνεργατικό φιλτράρισμα χρησιμοποιήθηκε για πρώτη φορά από την Tapestry με σκοπό να χειριστεί ένα μεγάλο όγκο εγγράφων (για παράδειγμα μαιλ και μηνύματα) [12]. Βασικό ρόλο στη σχεδίαση και υλοποίηση του έπαιξε η πεποίθηση ότι οι άνθρωποι παρέχουν την πιο αξιόπιστη αξιολόγηση των εγγράφων αυτών. Εφάρμοσαν το συνεργατικό φιλτράρισμα με τον παρακάτω τρόπο: Αρχικά έβαλαν τους χρήστες να σχολιάζουν τα έγγραφα αυτά και στη συνέχεια με βάση αυτά τα σχόλια μπορούσαν να τα φιλτράρουν με διάφορα queries. Αν και το σύστημα που χρησιμοποίησαν δεν ήταν αυτοματοποιημένο ήταν ένα πολύ σημαντικό βήμα για την καθιέρωση αυτής της μεθόδου.

Το πρώτο αυτοματοποιημένο σύστημα συνεργατικού φιλτραρίσματος αναπτύχθηκε για το usenet από μία ερευνητική ομάδα της GroupLens [13]. Το Usenet είναι ένα παγκόσμιο κατακευματισμένο σύστημα συζήτησης, που δημοσιεύονται μηνύματα (ονομάζονται είτε άρθρα είτε αναρτήσεις) σε μία ή περισσότερες κατηγορίες γνωστές ως ομάδες συζήτησης και ο κάθε χρήστης μπορεί να αξιολογήσει αν του άρεσε ή όχι το μήνυμα αυτό. Θα λέγαμε ότι είναι ο πρόδρομος των σημερινών διαδικτυακών φόρουμ.

Με την ανάπτυξη μίας στρατηγικής κοντινών γειτόνων, η οποία χρησιμοποιούσε σαν μετρική απόστασης την συσχέτιση Pearson, η πλατφόρμα της GroupLens παρείχε εξατομικευμένες προβλέψεις αξιολόγησης για κάθε νέο μήνυμα [14]. Έτσι ανάμεσα σε χιλιάδες καθημερινά μηνύματα σε κάποιον χρήστη που του άρεσαν για παράδειγμα τα αθλητικά δρώμενα θα του πρότεινε κατά προτεραιότητα άρθρα ή συζητήσεις αθλητικού περιεχομένου.

Ακολούθησαν και άλλες τεχνολογίες που χρησιμοποιούσαν το συνεργατικό φιλτράρισμα βασιζόμενο στις ομοιότητες μεταξύ των χρηστών όπως το Mosaic, ο πρώτος γραφικός περιηγητής που μπορούσε να συμπεριλάβει εικόνες μαζί με κείμενο και επέτρεπε στους χρήστες να δημοσιεύουν στις ιστοσελίδες σχόλια και σημειώσεις όπως και οι μουσικές εφαρμογές HOMR και Ringo.

Το HOMR (Helpful Online Music Recommendations) αυτοματοποίησε μια διαδικασία που επέτρεπε στους χρήστες να εκφράζουν τη γνώμη τους για διάφορα τραγούδια ή συγκροτήματα και στη συνέχεια χρησιμοποιούσε αυτές τις πληροφορίες για τις ανάγκες των χρηστών. Το Ringo ήταν μια πλατφόρμα μουσικής που επέτρεπε στους χρήστες να κάνουν διάφορες προτάσεις για μουσική και μέσα από το φιλτράρισμα της πληροφορίας πρότεινε στο χρήστη μία λίστα τραγουδιών που μπορεί να του αρέσει [40].

Στις αρχές της δεκαετίας του 2000 η εταιρεία Amazon υιοθέτησε μία τεχνική συνεργατικού φιλτραρίσματος, η οποία αποτελεί σταθμό στην ιστορία των συστημάτων συστάσεως (2003). Ενώ όλοι είχαν επικεντρωθεί στην ανάπτυξη των user-based τεχνικών, η Amazon χρησιμοποίησε ένα αλγόριθμο συνεργατικού φιλτραρίσματος που βασιζόταν στην ομοιότητα μεταξύ των αντικειμένων παρά των χρηστών [15]. Με την προσέγγιση αυτή, ο αλγόριθμος έβλεπε τις πρόσφατες αγορές του επισκέτη και για κάθε αγορά έκανε μία λίστα με σχετικά αντικείμενα. Έτσι αντικείμενα που εμφανίζονταν συνεχώς σε όλες τις λίστες ήταν υποψήφια για προτάσεις στον επισκέπτη. Οι υποψήφιες προτάσεις είχαν μεγαλύτερο ή μικρότερο βάρος ανάλογα με το πόσο σχετικά ήταν με τις πρόσφατες αγορές του χρήστη [16].

Το 2006 η πλατφόρμα streaming Netflix ξεκίνησε ανοιχτό διαγωνισμό με έπαθλο 1 εκατομμύριο δολάρια καλώντας όλη την κοινότητα της πληροφορικής να υλοποιήσει ένα αλγόριθμο που θα ξεπερνούσε τον υπάρχοντα αλγόριθμο 'Cinematch', τον βασικό αλγόριθμο του Netflix, κατά 10%, χρησιμοποιώντας ως μετρική το μέσο τετραγωνικό σφάλμα (RMSE) [17]. Το σύνολο δεδομένων όπου θα γινόταν η εκπαίδευση των αλγορίθμων αποτελούνταν από 100 εκατομμύρια σημεία και η εξέταση του αλγορίθμου έγινε αρχικά σε δύο test set με τη μορφή quiz test κατά το οποίο υποβάλλανε τις προβλέψεις τους και στη συνέχεια ενημερωθήκαν για το RMSE που είχαν επιτύχει. Το βραβείο απονεμήθηκε 3 χρόνια μετά σε μία συνεργαζόμενη ομάδα των κορυφαίων υποψηφίων. Η μεγαλύτερη συνεισφορά του διαγωνισμού αυτή ήταν η υιοθέτηση model based αλγορίθμων στο συνεργατικό φιλτράρισμα και η απαρχή της μετέπειτα εξέλιξης τους.

2.2 Κύριες Κατηγορίες Συνεργατικού Φιλτραρίσματος

Το συνεργατικό φιλτράρισμα είναι μία αποδοτική τεχνική στα συστήματα συστάσεων. Θα μπορούσαμε να χαρακτηρίσουμε το συνεργατικό φιλτράρισμα σαν μία διαδικασία επεξεργασίας μεγάλου όγκου παραγόμενων δεδομένων που τελικό σκοπό έχει να φιλτράρει τα άσχετα για τον χρήστη δεδομένα και να κρατήσει μόνο εκείνα που είναι πιθανό να αποτελούν καλό ταίριασμα στις προτιμήσεις του χρήστη. Μέχρι σήμερα αποτελεί την περισσότερο χρησιμοποιημένη και πετυχημένη μέθοδο στα συστήματα συστάσεων. Χωρίζεται σε δύο κύριες μεθόδους την memory-based που στηρίζεται κυρίως στην λογική των κοντινών γειτόνων και την model-based που βασίζεται στις κρυμμένες σχέσεις ανάμεσα στα αντικείμενα (latent factor models). Σε αυτή την υποενότητα θα γίνει λεπτομερή περιγραφή των δύο αυτών κατηγοριών.

2.2.1 Memory-Based

Οι memory-based αλγόριθμοι χρησιμοποιούν όλο τον πίνακα R χρήστη - προϊόντος για να κάνουν μία πρόβλεψη. Σκοπός τους είναι να βρουν κοντινούς γείτονες με τον χρήστη στον οποίο θέλουμε να προτείνουμε ένα προϊόν. Κοντινοί γείτονες θεωρούνται οι χρήστες που έχουν αγοράσει όμοια προϊόντα ή έχουν παρόμοιες προτιμήσεις με βάση τις αξιολογήσεις που έχουν κάνει στα εκάστοτε προϊόντα. Χωρίζονται σε δύο μεγάλες υποκατηγορίες, την user-based και την item-based προσέγγιση.

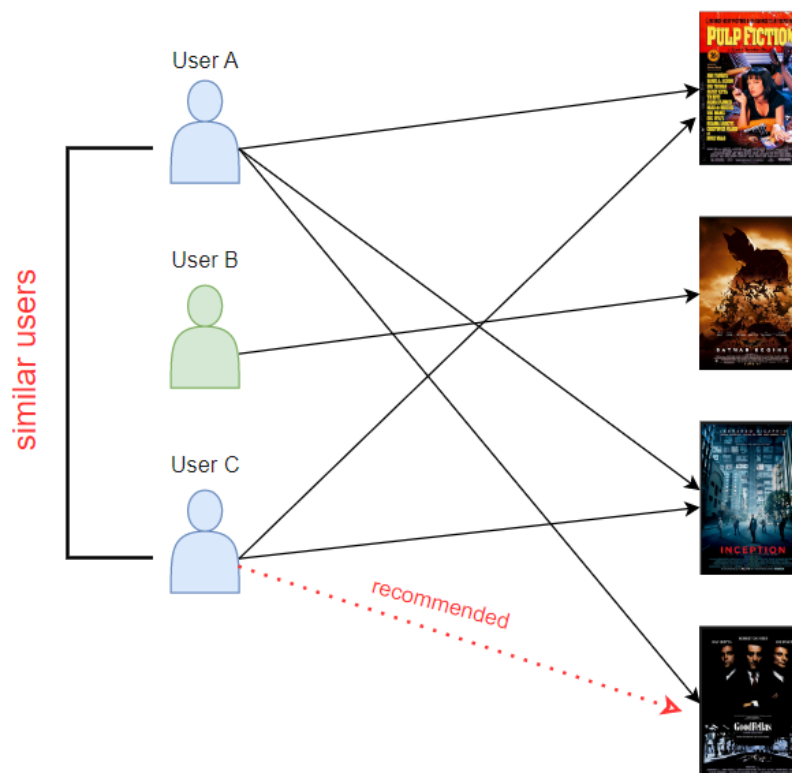


Figure 1: Προσέγγιση user-based μεθόδου

Η προσέγγιση user-based (Figure 1) ψάχνει ομοιότητα χρηστών που έχουν ίδιες αξιολογήσεις σε ίδια αντικείμενα, η οποία προκύπτει από διάφορες μετρικές απόστασης για τις οποίες

θα μιλήσουμε παρακάτω. Η λογική της προσέγγισης αυτής είναι η εξής: Έστω ότι έχουμε μία ομάδα χρηστών που τους αρέσουν τα ίδια αντικείμενα και ένα χρήστη u που του αρέσουν επίσης τα ίδια αντικείμενα. Όλοι αυτοί οι χρήστες ονομάζονται γείτονες του u . Στη συνέχεια ψάχνουμε στην ομάδα χρηστών τα αντικείμενα τα οποία τους αρέσουν αλλά δεν τα έχει δει ο χρήστης u και προτείνουν κάποιο j αντικείμενο από αυτά στο χρήστη u . Με δεδομένο ότι ο u έχει κοντινές προτιμήσεις με τους γείτονες του είναι αρκετά πιθανό να επιλέξει το αντικείμενο j και να μείνει ικανοποιημένος.

Η user-based προσέγγιση αντιμετωπίζει ένα πολύ σημαντικό θέμα αυτό του scalability. Ο αριθμός των χρηστών στις περισσότερες κατηγορίες είναι δυσανάλογα μεγάλος με τον αριθμό των αντικειμένων. Για παράδειγμα το IMDB έχει στη βάση του εκατομμύρια χρήστες εγγεγραμμένους και οι διαθέσιμες ταινίες είναι μερικές χιλιάδες. Λύση σε αυτό το πρόβλημα προσπαθήσε να δώσει η προσέγγιση item-based.

Η item-based προσέγγιση [18] βασίζεται σε έναν υπολογισμό της ομοιότητας μεταξύ των αντικειμένων σαν να ήταν user-based. Για παράδειγμα όπως βλέπουμε και στην εικόνα η ταινία A και η ταινία C είναι παρόμοιες έτσι θεωρούνται γειτονικές. Από τη στιγμή που ο χρήστης A έχει δηλώσει ότι του αρέσει η ταινία A, το item-based σύστημα θα προτείνει την ταινία C στο χρήστη αφού είναι γειτονική με την ταινία A. Η προσέγγιση item-based αναλύει δηλαδή ένα σύνολο αντικειμένων που ο χρήστης έχει ήδη αξιολογήσει και υπολογίζει την ομοιότητα μεταξύ αυτών των αντικειμένων και όλων των αντικειμένων. Τότε το σύστημα μπορεί να κάνει προβλέψεις και να προτείνει στον χρήστη παρόμοια αντικείμενα με βάση το ιστορικό του. Για τον υπολογισμό της απόστασης χρησιμοποιούνται προσεγγίσεις όμοιες με τη συσχέτιση Pearson αλλά και μετρικές συνημιτόνου τις οποίες θα δούμε αναλυτικότερα στο κεφάλαιο 3.

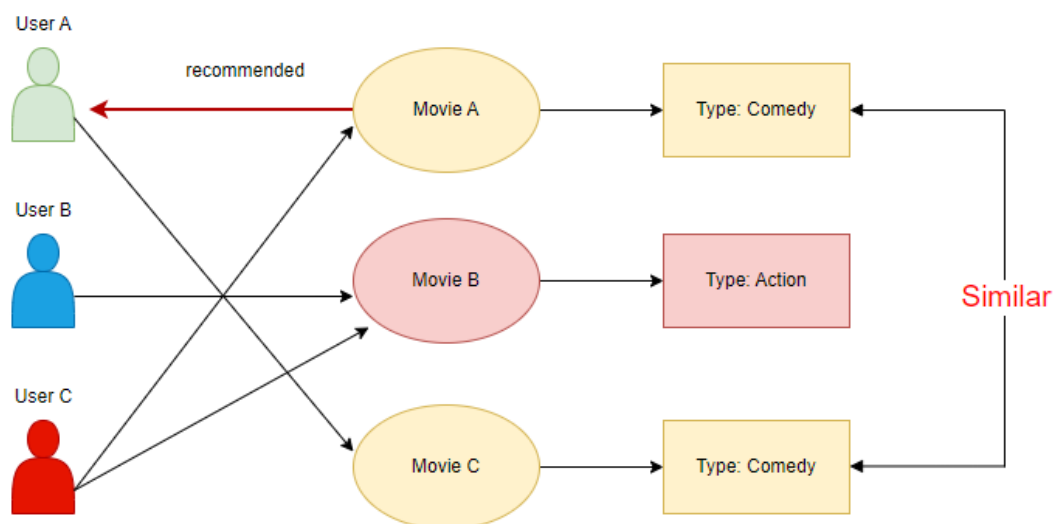


Figure 2: Προσέγγιση item-based μεθόδου

Κάθε μία από τις δύο προσεγγίσεις έχει τα δικά της πλεονεκτήματα και μειονεκτήματα και όταν γίνεται επιλογή ανάμεσα τους θα πρέπει να ληφθούν υπόψιν τα παρακάτω κριτήρια [19]:

- **Ακρίβεια:** Η ακρίβεια αυτών των μεθόδων εξαρτάται από την αναλογία μεταξύ χρηστών και αντικειμένων στο σύστημα. Σε περιπτώσεις όπου ο αριθμός των χρηστών είναι πολύ μεγαλύτερος από των αντικειμένων, όπως στην περίπτωση της Amazon, οι item-based μέθοδοι μπορούν να δώσουν μεγαλύτερη ακρίβεια. Αντιθέτως, σε περιπτώσεις όπου οι χρήστες είναι λιγότεροι από τα αντικείμενα όπως για παράδειγμα στην περίπτωση ενός ερευνητικού άρθρου όπου υπάρχουν χιλιάδες χρήστες αλλά εκατοντάδες χιλιάδες άρθρα για να προταθούν, οι user-based μέθοδοι μπορούν να είναι πιο αποδοτικοί.
- **Σταθερότητα:** Η σταθερότητα ενός συστήματος εξαρτάται από το πόσο συχνά και τι ποσοστό αλλαγών γίνονται μεταξύ χρηστών και αντικειμένων. Αν η λίστα διαθέσιμων αντικειμένων παραμένει σταθερή σε σύγκριση με τους χρήστες, η προσέγγιση item-based μπορεί να αποδειχθεί λιγότερο κοστοβόρα υπολογιστικά καθότι οι υπολογισμοί ομοιότητας μεταξύ αντικειμένων μπορούν να γίνονται σε αραιά διαστήματα χωρίς να επηρεάζονται οι προτάσεις προς τους χρήστες. Ένα τέτοιο παράδειγμα είναι μία online πλατφόρμα ταινιών. Από την άλλη σε ένα σύστημα όπου τα αντικείμενα συνεχώς αλλάζουν όπως σε ένα ειδησιογραφικό σαιτ βολεύουν περισσότερο οι user-based μέθοδοι.
- **Δικαιολογησιμότητα:** Σε αυτή την περίπτωση μία item-based προσέγγιση έχει πλεονέκτημα. Η λίστα των γειτονικών αντικειμένων που έχουν μεγάλο βαθμό ομοιότητας με αυτά που αρέσουν στο χρήστη, μπορούν να προταθούν στο χρήστη με αυτή την εξήγηση. Αυτό είναι πιο δύσκολο όμως να συμβεί στις user-based μεθόδους όπου ο χρήστης δεν ξέρω για τον τρόπο που λειτουργούν και αξιολογούν οι υπόλοιποι χρήστες.
- **Πρωτοτυπία Προτάσεων:** Σε item-based μεθόδους τα αντικείμενα που προτείνονται τείνουν να είναι όμοια με αυτά που αρέσουν ήδη στο χρήστη. Για παράδειγμα ένας χρήστης A που έχει αξιολογήσει θετικά πολλές κωμωδίες είναι πάρα πολύ πιθανό να του προτείνονται μόνο κωμωδίες με την item-based μέθοδο. Αυτό είναι κάτι που μπορεί να αντιμετωπίσει η user-based μέθοδο προτείνοντας και ταινίες πέραν των κωμωδιών. Για παράδειγμα αν ένας χρήστης B βαθμολογεί και αυτός θετικά τις κωμωδίες αλλά και τις περιπέτειες, τότε είναι πολύ πιθανό να προταθεί και μια ταινία περιπέτειας στον χρήστη A.
- **Υπολογιστικό Κόστος:** Το υπολογιστικό κόστος στα μοντέλα εκπαίδευσης εξαρτάται επίσης από την αναλογία χρηστών και αντικειμένων. Αν ο αριθμός των χρηστών υπερβαίνει κατά πολύ τον αριθμό των αντικειμένων τότε μία item-based προσέγγιση απαιτεί λιγότερο χρόνο και μνήμη για να υπολογίσει τα βάρη ομοιότητας μεταξύ των αντικειμένων.

2.2.2 Model-Based

Στις model-based μεθόδους χρησιμοποιούνται τεχνικές από τη γραμμική άλγεβρα, τη μηχανική μάθηση και την εξόρυξη δεδομένων στο πλαίσιο των προγνωστικών μοντέλων. Μερικά παραδείγματα τέτοιων μεθόδων είναι τα decision trees, μοντέλα βασιζόμενα στον κανόνα, Bayesian μοντέλα, μοντέλα με λανθάνοντες παράγοντες (latent factor models) και νευρωνικά δίκτυα [20]. Η model based μέθοδος προτάθηκε για να αντιμετωπίσει τους περιορισμούς των memory-based μεθόδων

Μία σημαντική διαφορά των model-based μεθόδων είναι ότι δεν χρησιμοποιούν όλο το σύνολο δεδομένων για να κάνουν προβλέψεις. Αυτό δίνει κάποια πολύ σημαντικά πλεονεκτήματα όπως ό,τι οι αλγόριθμοι model-based έχουν συνήθως γρηγορότερη απόδοση από τους memory-based. Επίσης αφού τα μοντέλα που προκύπτουν από τα model-based είναι αρκετά μικρότερα από τα πραγματικά dataset, έτσι καταλήγουν να είναι αρκετά μικρά για να χρησιμοποιηθούν αποδοτικά, δίνοντας έτσι την ικανότητα να έχουν scalability. Το μεγάλο μειονέκτημα της model-based μεθόδου είναι ότι το μοντέλο πρέπει να υπολογίζεται από την αρχή όταν ο πίνακας αλλάζει, δηλαδή όταν μια καινούργια αξιολόγηση προστεθεί.

Μία από τις καλύτερες σε απόδοση τεχνικές που χρησιμοποιείται ολοένα και περισσότερο τα τελευταία χρόνια είναι η παραγοντοποίηση πίνακα [21]. Στη μέθοδο αυτή ο πίνακας R χρηστών και αντικειμένων μπορεί να παραγοντοποιηθεί σε δύο μικρότερους πίνακες U και V, όπου ο πίνακας U αντικατοπτρίζει τη σχέση χρηστών - χαρακτηριστικών και ο πίνακας V τη σχέση αντικειμένων - χαρακτηριστικών.

Η κεντρική ιδέα των τεχνικών αυτών είναι να προσπαθούν να ανακαλύψουν κρυμμένα μοτίβα βασιζόμενα στις συσχετίσεις μεταξύ χρηστών και αξιολογήσεων. Αυτά τα κρυμμένα μοτίβα μπορεί να έχουν να κάνουν με χαρακτηριστικά του χρήστη ή του αντικειμένου και περιγράφουν ουσιαστικά τη διακύμανση των δεδομένων αξιολόγησης κάτι που αρκετές φορές οδηγεί στην εύρεση μοτίβων που δεν ερμηνεύονται από την ανθρώπινη σκοπιά, δηλαδή μπορεί να βρεθεί ένα μοτίβο που δεν αντιστοιχεί σε κάποιο χαρακτηριστικό του χρήστη ή του αντικειμένου.

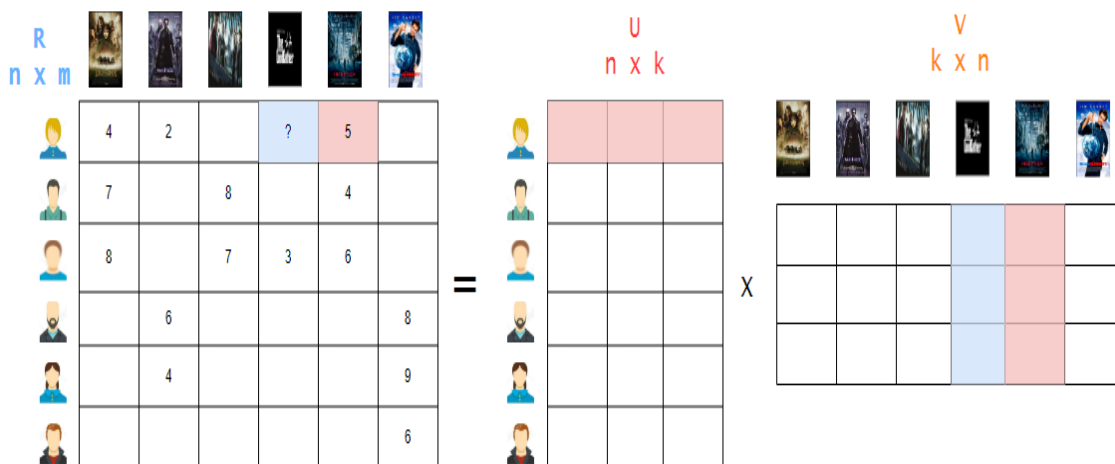


Figure 3: Προσέγγιση model-based μεθόδου

2.3 Memory-Based Αλγόριθμοι

Οι αλγόριθμοι memory-based χρησιμοποιούν τις αξιολογήσεις χρηστών για να υπολογίσουν ομοιότητες μεταξύ χρηστών ή αντικειμένων. Αρχικά θα επικεντρωθούμε στις προσεγγίσεις που βασίζονται στους κοντινούς γείτονες. Στη συνέχεια θα γίνει περιγραφή των μετρικών απόστασης που χρησιμοποιούνται για τον υπολογισμό της ομοιότητας των κοντινών γειτόνων.

2.3.1 K-Nearest Neighbors

Οι k -κοντινότεροι γείτονες είναι μία μέθοδο κατηγοριοποίησης βασιζόμενη σε στιγμιότυπα, η οποία είναι απλή αλλά αποδοτική σε πολλές περιπτώσεις [32]. Για να κατηγοριοποιήσουμε έναν χρήστη u , βρίσκουμε τους k πιο κοντινούς γείτονες του από το training σύνολο δεδομένων. Στη συνέχεια στο χρήστη u εκχωρείται μία ετικέτα κλάσης σύμφωνα με τις ετικέτες κλάσεις των πλησιέστερων γειτόνων. Συνήθως η ομοιότητα δύο σημείων σε ένα σύνολο δεδομένων υπολογίζεται από διάφορες μετρικές απόστασης.

Η μεγαλύτερη πρόκληση στον αλγόριθμο kNN είναι η επιλογή της τιμής του k . Σε περίπτωση που το k είναι πολύ μικρό θα είναι ευαίσθητο σε σημεία θορύβου, ενώ αν είναι πολύ μεγάλο ελλοχεύει ο κίνδυνος μια γειτονιά να συμπεριλάβει πολλά σημεία από άλλες ετικέτες κλάσης. Στη βιβλιογραφία έχουν προταθεί διάφορες λύσεις για την όσο το δυνατόν λιγότερη εξάρτηση από την επιλογή του k .

Για παράδειγμα στο [33] προτείνονται διάφορα σεντ με κοντινούς γείτονες παρά ένα σεντ με k -κοντινούς γείτονες. Ένας από τους πιο συνηθισμένους τρόπους είναι να τρέξουμε τον αλγόριθμο αρκετές φορές αλλάζοντας την τιμή του k και στο τέλος να διαλέξουμε το k με την υψηλότερη απόδοση.

Η τεχνική των κοντινών γειτόνων είναι μία από τις πιο κοινές προσεγγίσεις στο συνεργατικό φιλτράρισμα καθώς όπως γίνεται αντιληπτό και από τα παραπάνω η λογική του στηρίζεται στη λογική που είναι χτισμένο το συνεργατικό φιλτράρισμα. Όπως δηλαδή στο συνεργατικό φιλτράρισμα ψάχνουμε να βρούμε χρήστες ή αντικείμενα με όμοια χαρακτηριστικά έτσι και στους κοντινούς γείτονες κατηγοριοποιούμε με βάση την ομοιότητα σε ετικέτες κλάσης.

Ένα άλλο πλεονέκτημα του είναι ότι δεν απαιτεί την εκμάθηση και διατήρηση ενός δεδομένου μοντέλου. Είναι δηλαδή μία τεμπέλικη μέθοδο εκμάθησης και μπορεί να εφαρμοστεί με καλά αποτελέσματα σε τομείς όπου απαιτείται δυναμική κατηγοριοποίηση. Το μειονέκτημα στη συγκεκριμένη περίπτωση είναι ότι απαιτείται μεγάλο υπολογιστικό κόστος για τον επανυπολογισμό του πίνακα χρήστη αντικειμένου όταν προστίθενται νέες αλλαγές. Έχουν αναπτυχθεί διάφορες τεχνικές για την μείωση του υπολογιστικού κόστους, οι οποίες δε θα μελετηθούν στην εργασία μας αλλά ο αναγνώστης μπορεί να βρει περισσότερες πληροφορίες στο [34].

2.3.1.1 Προσέγγιση κοντινών γειτόνων

Η προσέγγιση που θα ακολουθήσουμε είναι των κορυφαίων N κοντινών γειτόνων. Σε αυτή την προσέγγιση διαλέγονται οι N πιο κοντινοί γείτονες με βάση τις μετρικές ομοιότητας και οι υπόλοιποι μένουν εκτός των προτάσεων που θα δώσει το σύστημα. Όσο μεγαλύτερο είναι το σύνολο δεδομένων μας, τόσο θα εξισορροπείται η διακύμανση των αξιολογήσεων. Αυτό, όμως, ενέχει τον κίνδυνο να μειωθεί η ικανότητα καλής πρόβλεψης.

Σε αυτή την προσέγγιση τίθεται το ερώτημα για το ποιος είναι ο καλύτερος δείκτης N χρηστών ώστε να έχουμε αξιόπιστα αποτελέσματα, κάτι που φυσικά συσχετίζεται με το σύνολο δεδομένων. Ένας συνηθισμένος τρόπος για να καθοριστεί το N είναι μέσα από την μέθοδο cross validation.

Η διαδικασία επαναλαμβάνεται με αρκετά διαφορετικά N και τελικά καταλήγουμε στο N που μας δίνει τα καλύτερα αποτελέσματα. Η συγκεκριμένη προσέγγιση μας βοηθάει στο scalability του συνόλου δεδομένων από τη στιγμή που τελικά θα χρησιμοποιηθεί ένας συγκεκριμένος αριθμός χρηστών θα έχουμε καλύτερη απόδοση όταν τρέξουμε τον αλγόριθμο μας [31].

Ένα επίσης σημαντικό πλεονέκτημα είναι ότι το σύστημα συστάσεων είναι σε θέση να παρέχει τη λίστα με τους κοντινούς γείτονες για έναν χρήστη και με αυτό τον τρόπο να αιτιολογήσει για ποιο λόγο του έκανε τις συγκεκριμένες προτάσεις. Από επιχειρηματικής σκοπιάς έχει αποδειχθεί ιδιαίτερα ευεργετικό ότι προτείνονται στους χρήστες αντικείμενα εξηγώντας τους λόγους για την εκάστοτε πρόταση. Είναι σημαντικό για έναν χρήστη να βλέπει ότι ένα αντικείμενο του έχει προταθεί με ένα συγκεκριμένο σκεπτικό και όχι τυχαία [42].

Στον αντίποδα η συγκεκριμένη προσέγγιση κατά πάσα πιθανότητα δεν θα συμπεριλάβει κάθε χρήστη που βρίσκεται στο σύστημα. Για κάποιους χρήστες μπορεί να μην είναι εύκολο να βρούμε κοντινούς γείτονες και να χρειάζονται περισσότεροι γείτονες για να μπορέσουμε να παράγουμε καλύτερης ποιότητας προτάσεις. Επιπλέον κάποιοι χρήστες που δεν είναι τόσο καλά συσχετισμένοι με τους υπόλοιπους N χρήστες της ομάδας που συμπεριλήφθηκαν μπορεί να μειώσουν την απόδοση στο πόσο καλή θα είναι τελικά η πρόταση.

2.3.1.2 kNN Basic

Η βασική μορφή του αλγορίθμου kNN όπως ορίζεται στο [9]. Η πρόβλεψη για το $\hat{r}(u,i)$ ορίζεται ως εξής;

Για την προσέγγιση user-based έχουμε τον παρακάτω τύπο:

$$\hat{r}_{ui} = \frac{\sum_{v \in N_i^k(u)} \text{sim}(u, v) \cdot r_{vi}}{\sum_{v \in N_i^k(u)} \text{sim}(u, v)} \quad (1)$$

Για την προσέγγιση item-based έχουμε τον παρακάτω τύπο [9]:

$$\hat{r}_{ui} = \frac{\sum_{j \in N_u^k(i)} \text{sim}(i, j) \cdot r_{uj}}{\sum_{j \in N_u^k(i)} \text{sim}(i, j)} \quad (2)$$

2.3.1.3 kNN with Means

Στην υποενότητα 2.5 Κανονικοποίηση Βαθμολογιών είχαμε αναφερθεί στην έννοια της κανονικοποίησης, σαν τρόπου αντιμετώπισης της διαφορετικής κλίμακας βαθμολόγησης μεταξύ των χρηστών. Υπενθυμίζουμε ότι μέσω της κανονικοποίησης όλες οι βαθμολογίες των χρηστών προσαρμόζονται σε μία κοινή κλίμακα για όλους τους χρήστες. Μιλήσαμε για την τεχνική είναι η mean-centering. Σε αυτή την υποενότητα θα δούμε μια βελτίωση του kNN αλγορίθμου στον οποίο ενσωματώνεται και η mean-centering κανονικοποίηση.

Έτσι, ο αλγόριθμος kNN with Means στην user-based προσέγγιση για τον υπολογισμό του \hat{r}_{ui} δίνεται από τον παρακάτω τύπο [9]:

$$\hat{r}_{ui} = \mu_u + \frac{\sum_{v \in N_i^k(u)} \text{sim}(u, v) \cdot (r_{vi} - \mu_v)}{\sum_{v \in N_i^k(u)} \text{sim}(u, v)} \quad (3)$$

Ενώ στην item-based προσέγγιση για τον υπολογισμό του \hat{r}_{ui} έχουμε τον παρακάτω τύπο [9]:

$$\hat{r}_{ui} = \mu_i + \frac{\sum_{j \in N_u^k(i)} \text{sim}(i, j) \cdot (r_{uj} - \mu_j)}{\sum_{j \in N_u^k(i)} \text{sim}(i, j)} \quad (4)$$

2.3.1.4 kNN with Z-Score

Στην υποενότητα 2.5 Κανονικοποίηση Βαθμολογιών είχε επίσης γίνει αναφορά για την τεχνική Z-Score. Θυμίζουμε συνοπτικά ότι έχει παρόμοια προσέγγιση με την mean-centering τεχνική αλλά λαμβάνεται υπόψιν και η τυπική απόκλιση των αξιολογήσεων. Όμοια σε αυτή την ενότητα παρουσιάζεται μια βελτίωση της εκδοχής του kNN αλγορίθμου με την ενσωμάτωση του Z-Score.

Στον αλγόριθμο kNN with Z-Score στην user-based προσέγγιση για τον υπολογισμό του \hat{r}_{ui} έχουμε τον παρακάτω τύπο [9]:

$$\hat{r}_{ui} = \mu_u + \sigma_u \frac{\sum_{v \in N_i^k(u)} \text{sim}(u, v) \cdot \frac{(r_{vi} - \mu_v)}{\sigma_v}}{\sum_{v \in N_i^k(u)} \text{sim}(u, v)} \quad (5)$$

Ενώ στην item-based προσέγγιση για τον υπολογισμό του \hat{r}_{ui} έχουμε τον παρακάτω τύπο [9]:

$$\hat{r}_{ui} = \mu_i + \sigma_i \frac{\sum_{j \in N_u^k(i)} \text{sim}(i, j) \cdot \frac{(r_{uj} - \mu_j)}{\sigma_j}}{\sum_{j \in N_u^k(i)} \text{sim}(i, j)} \quad (6)$$

2.3.2 Μετρικές Απόστασης

Οι μετρικές απόστασης παίζουν πολύ σημαντικό ρόλο στα συστήματα συστάσεων τα οποία μετράνε την ομοιότητα μεταξύ χρηστών ή αντικειμένων. Είναι ένας τρόπος για να προσδώσουμε στους γείτονες μεγαλύτερη ή μικρότερη βαρύτητα στις προβλέψεις. Σε αυτή την υποενότητα θα περιγράψουμε δύο από τις πιο δημοφιλείς μεθόδους, την ομοιότητα συνημιτόνου και την συσχέτιση Pearson (Pearson Correlation) [36] τις οποίες και θα χρησιμοποιήσουμε για το πείραμα μας στο επόμενο κεφάλαιο.

Συσχέτιση Pearson: Μία από τις πρώτες προσεγγίσεις Pearson εμφανίστηκε σαν κατάλληλο μέτρο ομοιότητας στο σύστημα Group-Lens. Είναι ίσως η πιο χρησιμοποιημένη προσέγγιση για τις user-based προσεγγίσεις. Στη συσχέτιση Pearson η κλίμακα ομοιότητας παίρνει τιμές στο διάστημα από -1 έως και +1. Στην τιμή -1 θεωρείται ότι έχουν πολύ χαμηλή συσχέτιση (αρνητική) και στην +1 πολύ υψηλή συσχέτιση (θετική). Όταν παίρνουν την τιμή μηδέν θεωρείται ότι δεν έχουν καμία συσχέτιση.

Έστω ότι έχουμε τον πίνακα I_{uv} όπου εκφράζει το σύνολο των αντικειμένων που έχουν βαθμολογηθεί από τους χρήστες u και v , r_{ui} και r_{vi} είναι οι βαθμολογίες των χρηστών u και v για το αντικείμενο i αντίστοιχα και μ_u , μ_v ο μέσος όρος των αξιολογήσεων των χρηστών u και v .

Η συσχέτιση pearson $p(u,v)$ μεταξύ των χρηστών u και v δίνεται από την εξίσωση [9]:

$$p(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \mu_u)(r_{vi} - \mu_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \mu_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{vi} - \mu_v)^2}} \quad (7)$$

Η συσχέτιση pearson μεταξύ δύο αντικειμένων i και j ορίζεται ως εξής [9]:

$$p(i, j) = \frac{\sum_{u \in U_{ij}} (r_{ui} - \mu_i)(r_{uj} - \mu_j)}{\sqrt{\sum_{u \in U_{ij}} (r_{ui} - \mu_i)^2} \sqrt{\sum_{u \in U_{ij}} (r_{uj} - \mu_j)^2}} \quad (8)$$

με U_{ij} να εκπροσωπεί το σύνολο των χρηστών που έχουν βαθμολογήσει τα αντικείμενα i και j αντίστοιχα, μ_i και μ_j τους μέσους όρους αξιολόγησης των προϊόντων i και j μεταξύ όλων των χρηστών που βαθμολόγησαν το αντικείμενο.

Ομοιότητα Συνημιτόνου: Η ομοιότητα συνημιτόνου είναι μία μετρική απόστασης που είναι περισσότερο ταιριαστή σε item-based προσεγγίσεις του συνεργατικού φιλτραρίσματος. Χρησιμοποιεί την ίδια κλίμακα με την μετρική Pearson όπου το +1 εκπροσωπεί υψηλή (θετική) συσχέτιση το -1 υψηλή (αρνητική) συσχέτιση και το 0 δείχνει ότι δεν υπάρχει διακύμανση.

Σε μορφή διανύσματος μετράει τη γωνία μεταξύ των διανυσμάτων \vec{a} , \vec{b} και ορίζεται ως εξής [9]:

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}|^2 \cdot |\vec{b}|^2}$$

Για την user based προσέγγιση δηλαδή τον υπολογισμό ομοιότητας μεταξύ των χρηστών ο τύπος διαμορφώνεται ως εξής [9]:

$$\cos(u, v) = \frac{\sum_{i \in I_{uv}} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in I_{uv}} r_{ui}^2} \sqrt{\sum_{i \in I_{uv}} r_{vi}^2}}$$

Ενώ για την item based προσέγγιση δηλαδή τον υπολογισμό ομοιότητας μεταξύ των αντικειμένων ο τύπος διαμορφώνεται ως εξής [9]:

$$\cos(i, j) = \frac{\sum_{u \in U_{ij}} r_{ui} \cdot r_{uj}}{\sqrt{\sum_{u \in U_{ij}} r_{ui}^2} \sqrt{\sum_{u \in U_{ij}} r_{uj}^2}}$$

Μέση Τετραγωνική Απόσταση: Η Μέση Τετραγωνική Απόσταση (Mean Squared Distance) για την user-based προσέγγιση ορίζεται ως εξής [40]:

$$s(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - r_{vi})^2}{|I_{uv}|}$$

όπου $s(u, v)$ η ομοιότητα μεταξύ των χρηστών u και v ανάμεσα σε όλα τα προϊόντα που έχουν βαθμολογήσει από κοινού.

Στην περίπτωση της item-based προσέγγισης ο τύπος υπολογισμού ορίζεται ως εξής [40]:

$$s(i, j) = \frac{\sum_{u \in U_{ij}} (r_{iu} - r_{ju})^2}{|U_{ij}|}$$

όπου $s(i, j)$ η ομοιότητα μεταξύ των χρηστών i και j ανάμεσα σε όλα τα αντικείμενα i και j που έχουν βαθμολογηθεί από όλους τους χρήστες που έχουν βαθμολογήσει και τα δύο αντικείμενα.

Αξίζει να σημειωθεί ότι σε κάποιες περιπτώσεις στη βιβλιογραφία συνίσταται η χρήση του αντίστροφου μέσης τετραγωνικής απόστασης που ορίζεται ως εξής για την user-based προσέγγιση [40]:

$$\hat{s}(u, v) = \frac{1}{s(u, v) + 1}$$

και όπως παρακάτω για την item-based προσέγγιση [40]:

$$\hat{s}(i, j) = \frac{1}{s(i, j) + 1}$$

με $\hat{s}(u, v)/\hat{s}(i, j)$ να συμβολίζουν την ομοιότητα του μέσου τετραγώνου και $s(u, v)/s(i, j)$ την μέση τετραγωνική απόσταση. Ο αριθμός 1 στον παρονομαστή έχει μπει έτσι ώστε ο παρονομαστής να μην πάρει ποτέ την τιμή 0.

2.4 Model-Based Αλγόριθμοι

Σε αυτήν την υποενότητα θα γίνει περιγραφή των model-based μοντέλων μας. Αρχικά θα μιλήσουμε για τον απλό αλλά αρκετά αποδοτικό αλγόριθμο Slope One που βασίζεται σε τεχνική εξόρυξης δεδομένων και στη συνέχεια θα επικεντρωθούμε σε μεθόδους παραγοντοποίησης πίνακα αναλύοντας την μέθοδο SVD (Ανάλυση Ιδιαζουσών Τιμών) και μία βελτιωμένη εκδοχή της όπως παρουσιάστηκε στο [41] την SVD++.

2.4.1 Slope One

Ο αλγόριθμος Slope One αναπτύχθηκε για πρώτη φορά στο [37] το 2005 και είναι σχεδιασμένος ώστε να προβλέπει πως ο χρήστης θα βαθμολογήσει ένα προϊόν στο online βασισμένο στις αξιολογήσεις συνεργατικό φιλτράρισμα. Είναι ένας γραμμικός αλγόριθμος, της μορφής $f(x) = x + b$, που υπολογίζει την μέση διαφορά ανάμεσα σε ένα αντικείμενο και ένα άλλο για χρήστες που έχουν βαθμολογήσει και τα δύο.

Τα πλεονεκτήματα του αλγορίθμου αυτού είναι τα εξής:

- Είναι εύκολος στην υλοποίηση και στη διατήρηση του.
- Είναι εύκολα ανανεώσιμος καθώς η προσθήκη μίας καινούργιας αξιολόγησης μπορεί να αλλάξει άμεσα όλες τις προβλέψεις.
- Είναι αποδοτικός με σύντομο χρόνο απόκρισης στα queries.
- Ένας σχετικά νέος χρήστης μπορεί να λάβει εύστοχες προτάσεις.
- Έχει ανταγωνιστικά ποσοστά ακρίβειας σε σχέση με τα υπόλοιπα memory-based μοντέλα.

Ο αλγόριθμος slope one είναι σχεδιασμένος στην αρχή της "διαφοράς δημοτικότητας" ανάμεσα στα αντικείμενα για τους χρήστες. Αντικείμενο και χρήστες χωρίζονται κατά μία έννοια σε ζευγάρια και προσπαθούμε να αποφασίσουμε πόσο περισσότερο ένα αντικείμενο αρέσει από ένα άλλο. Ο τρόπος μέτρησης είναι απλά να αφαιρούμε το μέσο όρο αξιολόγησης δύο αντικειμένων. Στη συνέχεια αυτή τη διαφορά τη χρησιμοποιήσουμε για να προβλέψουμε την αξιολόγηση που έχει κάνει ο χρήστης σε ένα από αυτά τα αντικείμενα, δεδομένου ότι έχει βαθμολογήσει το άλλο.

Ας δούμε πρώτα ένα απλοϊκό παράδειγμα για την κατανόηση του τρόπου λειτουργίας του αλγορίθμου. Έστω ότι έχουμε δύο χρήστες τον χρήστη A και τον χρήστη B και δύο αντικείμενα J και I. Έστω ότι ο χρήστης A έχει βαθμολογήσει με 3 το αντικείμενο J και με 5 το αντικείμενο I και ο χρήστης B έχει βαθμολογήσει με 2 το αντικείμενο J.

Τότε παρατηρούμε ότι ο χρήστης έχει βαθμολογήσει με $5 - 3 = 2$ πόντους περισσότερο το αντικείμενο I και έτσι μπορούμε να προβλέψουμε ότι ο χρήστης B θα βαθμολογήσει το αντικείμενο I με $2+2 = 4$ πόντους.

Ο χρήστης B ονομάζεται προβλεπόμενος χρήστης και το αντικείμενο I προβλεπόμενο αντικείμενο. Σε ένα σύνολο δεδομένων που υπάρχουν πολλές τέτοιες διαφορές για κάθε αντικείμενο I που θέλουμε να προβλέψουμε παίρνουμε τους μέσους όρους αυτών των διαφορών μεταξύ των αντικειμένων i και j για όσους χρήστες τα έχουν βαθμολογήσει και τα δύο.

	Item j	Item i	
User A	3	5	item i - item j = 2
User B	2	?	

$? = 2 + (5 - 3) = 4$

Figure 4: Παράδειγμα slope one αλγορίθμου

Διατυπώνοντας το παραπάνω παράδειγμα σε μία γενικευμένη μαθηματική εξίσωση έχουμε το εξής: Έστω χ ένα σύνολο δεδομένων εκπαίδευσης και δύο αντικείμενα i, j με αξιολογήσεις u_j και u_i από ένα χρήστη $u \in U_{i,j}$. Η μέση διακύμανση ενός αντικειμένου i σε σχέση με το αντικείμενο j ορίζεται ως εξής:

$$dev(i, j) = \sum_{u \in U_{i,j}} \frac{r_{ui} - r_{uj}}{|U_{i,j}|}$$

Τότε σύμφωνα με τον παραπάνω τύπο, η πρόβλεψη \hat{r}_{ui} δίνεται παρακάτω:

$$\hat{r}_{ui} = \sum_{j \in R_i(u)} \frac{dev(i, j) + \mu_j}{|R_i(u)|}$$

με $R_i(u) = \{j \mid j \in U, j \neq i, U_{ij} > 0\}$ να είναι το σύνολο όλων των σχετικών αντικειμένων.

Σε ορισμένες περιπτώσεις της βιβλιογραφίας [38] διατυπώνεται μία παραλλαγή του Slope One αλγορίθμου weighted Slope One, στην οποία θέτουμε βάρη για την πρόβλεψη του \hat{r}_{ui} και ο τύπος τροποποιείται ως εξής:

$$\hat{r}_{ui} = \sum_{j \in R_i(u)} \frac{(dev(i, j) + \mu_j) \cdot c_{ij}}{|R_i(u)|}$$

όπου c_{ij} εκπροσωπεί τον αριθμό των χρηστών που αξιολογήσανε και το αντικείμενο i και το αντικείμενο j .

2.4.2 SVD

Η μέθοδος SVD χρησιμοποιείται για την ανακάλυψη κρυμμένων παραγόντων που προκύπτουν από τα μοτίβα αξιολόγησης. Ανήκει στα μοντέλα παραγοντοποίησης πίνακα, μία τεχνική στην οποία οι χρήστες και τα αντικείμενα αντιστοιχίζονται σε ένα κοινό χώρο διάστασης f έτσι ώστε οι αλληλεπιδράσεις που προκύπτουν μεταξύ τους μέσα από τις αξιολογήσεις να μοντελοποιούνται στο χώρο αυτό.

Με πιο απλά λόγια μπορούμε να αναφέρουμε το παράδειγμα χρηστών και ταινιών, στο οποίο οι κρυμμένοι παράγοντες που προσπαθεί να εντοπιστεί είναι για παράδειγμα το είδος της ταινίας αν είναι ντοκιμαντέρ ή επιστημοντικής φαντασίας ή αν απευθύνεται σε ενήλικο κοινό.

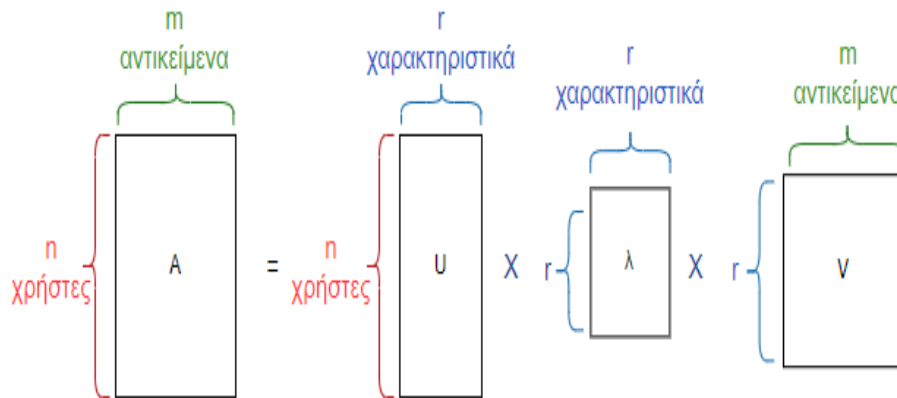


Figure 5: Περιγραφή SVD μεθόδου

Η Ανάλυση Ιδιάζουσων Τιμών αρχικά παραγοντοποιεί τον πίνακα A σε $A = UV^t$, με τον πίνακα A να έχει διαστάσεις $n \times m$, τον πίνακα U που είναι ο μοναδιαίος πίνακας χρηστών διαστάσεις $n \times r$ και τον πίνακα V που είναι ο μοναδιαίος πίνακας ταινιών διαστάσεις $r \times m$. Το r είναι ο αριθμός των κρυμμένων παραγόντων που προκύπτει από την παραγοντοποίηση πίνακα ενώ ο λ πίνακας είναι διαγώνιος πίνακας και δίνει την βαρύτητα των κρυμμένων παραγόντων [40].

Βασιζόμενοι στην παραπάνω αρχή παίρνουμε τον αλγόριθμο που χρησιμοποιείται στο συνεργατικό φιλτράρισμα. Κάθε αντικείμενο i εκπροσωπείται με ένα διάνυσμα $q_i \in R^f$ και κάθε χρήστης u συνδέεται με ένα διάνυσμα $p_u \in R^f$. Για το αντικείμενο i τα στοιχεία του q_i μετράνε το πόσο ένα αντικείμενο διαθέτει αυτούς τους παράγοντες, υψηλά ή αρνητικά. Για έναν χρήστη u αντίστοιχα τα στοιχεία του p_u μετράνε το πόσο ενδιαφέρον έχουν για τους χρήστες, υψηλό ή αρνητικό, για αντικείμενα τα οποία είναι ψηλά στους αντίστοιχους παράγοντες [11].

Το γινόμενο $q_i^T p_u$ υπολογίζει την αλληλεπίδραση μεταξύ του χρήστη u και του αντικειμένου i δείχνοντας την προτίμηση του χρήστη u για τα χαρακτηριστικά του κάθε αντικειμένου. Στον υπολογισμό της προβλεπόμενης αξιολόγησης θα προστεθούν και κάποιες υποκειμενικές εκτιμήσεις που εξαρτώνται μόνο από τον χρήστη ή το αντικείμενο, σαν ένας είδος προκατάληψης στις προτιμήσεις του χρήστη.

Τελικά το \hat{r}_{ui} υπολογίζεται ως εξής [9]:

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u$$

με b_u, b_i να υποδηλώνουν την προκατάληψη προς τον χρήστη u και το αντικείμενο i αντίστοιχα.

Για να υπολογίσουμε τις παραμέτρους (b_u, b_i, p_u και q_i) ελαχιστοποιούμε το κανονικοποιημένο τετραγωνικό σφάλμα [9]:

$$\sum_{r_{ui} \in R_{\text{train}}} (r_{ui} - \hat{r}_{ui})^2 + \lambda(b_u^2 + b_i^2 + \|q_i\|^2 + \|p_u\|^2)$$

Η σταθερά λ ελέγχει το βαθμό της κανονικοποίησης και συνήθως καθορίζεται από τη μέθοδο cross validation. Η ελαχιστοποίηση στην περίπτωση μας μπορεί να υλοποιηθεί εφαρμόζοντας τη μέθοδο στοχαστικής κλίσης καθόδου όπως ορίζεται στο [39]. Ο αλγόριθμος τρέχει σε λούπα μέσω από όλες τις αξιολογήσεις του συνόλου εκπαίδευσης των δεδομένων. Για κάθε αξιολόγηση r_{ui} γίνεται μία πρόβλεψη \hat{r}_{ui} και από τη διαφορά τους προκύπτει το αντίστοιχο λάθος πρόβλεψης e_{ui} . Τελικά έχουμε [9]:

$$b_u \leftarrow b_u + \gamma(e_{ui} - \lambda b_u)$$

$$b_i \leftarrow b_i + \gamma(e_{ui} - \lambda b_i)$$

$$q_i \leftarrow q_i + \gamma(e_{ui} p_u - \lambda q_i)$$

$$p_u \leftarrow p_u + \gamma(e_{ui} q_i - \lambda p_u)$$

με γ να είναι ο ρυθμός εκμάθησης που καθορίζεται από το χρήστη.

2.4.3 SVD++

Στην ενότητα 2 έγινε αναφορά για το άμεσο και έμμεσο feedback και πως ο συνδυασμός τους είναι η καλύτερη δυνατή επιλογή για την προβλεπτικότητα ενός μοντέλου. Πάνω σε αυτή τη βάση αναπτύχθηκε το μοντέλο SVD++ [41] θέλοντας να πετύχει καλύτερη ακρίβεια πρόβλεψης από το SVD. Η επιπλέον βελτίωση δηλαδή της μεθόδου SVD++ είναι ότι λαμβάνει υπόψιν και το έμμεσο feedback του χρήστη. Ακόμα και σε περιπτώσεις όπου το έμμεσο feedback δεν υπάρχει αυτό που θεωρεί σαν σημαντικό στοιχείο ο αλγόριθμος είναι η καταμέτρηση των αξιολογήσεων για το κάθε προϊόν, χωρίς να δίνει έμφαση στην βαθμολογία αυτή καθ'αυτή.

Αυτό οδήγησε σε ένα δεύτερο σύνολο αντικειμένων που συσχετίζει το κάθε αντικείμενο i με ένα παράγοντα διανύσματος $\mathbf{y}_i \in \mathbb{R}^f$. Αυτός ο δεύτερος λανθάνοντας παράγοντας χρησιμοποιείται για να χαρακτηρίσει τους χρήστες βασιζόμενο στο σύνολο των προϊόντων που έχουν βαθμολογήσει. Ο τύπος υπολογισμού της πρόβλεψης του \hat{r}_{ui} ορίζεται ως εξής [41]:

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T \left(p_u + \frac{1}{\sqrt{|I_u|}} \sum_{j \in I_u} y_j \right)$$

με y_i να είναι ο παράγοντας που υπολογίζει το έμμεσο feedback και I_u να υπολογίζει όλα τα αντικείμενα που έχει βαθμολογήσει ο χρήστης u . Στην περίπτωση μας το έμμεσο feedback προσμετράται μόνο από το αν ο χρήστης u έχει ψηφίσει για ένα αντικείμενο και δε μας ενδιαφέρει με πόσο το έχει βαθμολογήσει. Τα y_i είναι επικεντρωμένα γύρω από το 0 και το άθροισμα τους κανονικοποιείται από το $|I_u|^{-\frac{1}{2}}$ έτσι ώστε να ομαλοποιηθεί η διασπορά ανάμεσα στο εύρος των τιμών του $|I_u|$.

Η ελαχιστοποίηση επιτυγχάνεται όπως και στη μέθοδο SVD μέσα από τη μέθοδο στοχαστικής κλίσης καθόδου [41]:

$$\begin{aligned} b_u &\leftarrow b_u + \gamma \cdot (e_{ui} - \lambda b_u) \\ b_i &\leftarrow b_i + \gamma \cdot (e_{ui} - \lambda b_i) \\ q_i &\leftarrow q_i + \gamma \cdot (e_{ui} \cdot (p_u + |I_u|^{-1/2} \sum_{j \in I_u} y_j) - \lambda_1 q_i) \\ p_u &\leftarrow p_u + \gamma \cdot (e_{ui} q_i - \lambda_1 p_u) \\ y_j &\leftarrow y_j + \gamma \cdot (e_{ui} \cdot |I_u|^{-1/2} \cdot q_i - \lambda_1 y_j) \end{aligned}$$

2.5 Βαθμολογίες Χρηστών

Οι προτιμήσεις του χρήστη μπορούν να εκφραστούν με δύο τρόπους [22]. Ο πρώτος τρόπος είναι και ο πιο άμεσος που γίνεται με την αξιολόγηση ενός προϊόντος σε μια συγκεκριμένη κλίμακα βαθμολογίας (explicit feedback). Για παράδειγμα όταν ένας χρήστης βάζει σε μια ταινία δράσης στην εφαρμογή του IMDB 10 αστεράκια, υποδηλώνει ότι η ταινία του άρεσε πάρα πολύ.

Ο δεύτερος τρόπος είναι πιο έμμεσος και μπορεί να μας δώσει πληροφορίες για το χρήστη με βάση τις αναζητήσεις που έκανε ή τις αγορές θεωρώντας ότι του αρέσει κάτι που αγόρασε (implicit feedback). Ένα παράδειγμα για αυτή την περίπτωση είναι η σελίδα του Amazon, όπου ο χρήστης μπορεί να κάνει μία αγορά προϊόντος κάτι που υποδηλώνει ότι αυτό το προϊόν του αρέσει.

Οι κλίμακες αξιολόγησης αυτών των τρόπων μπορούν να χωριστούν στις παρακάτω κατηγορίες: Στη δυαδική κατηγορία όπου ο τρόπος αξιολόγησης είναι αρέσει ή δεν αρέσει κάτι στο χρήστη, στην ακέραια κατηγορία όπου ο χρήστης βαθμολογεί σε μία κλίμακα από 1 έως N και στην μοναδιαία κατηγορία όπου ο χρήστης μπορεί να κάνει κλικ πολλές φορές σε ένα προϊόν ή να το αγοράσει [23].

Όπως γίνεται αντιληπτό οι δύο πρώτες μορφές αφορούν τον άμεσο τρόπο αξιολόγησης και η τελευταία τον έμμεσο τρόπο. Αξίζει να αναφερθεί ότι μία σημαντική διαφορά μεταξύ των δύο πρώτων και του τελευταίας μορφής είναι το γεγονός ότι στην πρώτη περίπτωση μπορούμε να αντλήσουμε πληροφορίες τόσο για το τι αρέσει όσο και για το τι δεν αρέσει στο χρήστη ενώ στην τελευταία μαθαίνουμε μόνο πληροφορίες για το τι αρέσει στο χρήστη.

Ο άμεσος τρόπος αξιολόγησης όταν συμβαίνει έχει πολύ μεγάλη σημασία, καθώς το σύστημα ενημερώνεται για το πως ακριβώς βλέπει ο χρήστης το συγκεκριμένο αντικείμενο. Με μία πρώτη σκέψη αυτό φαίνεται ιδανικό, όμως, τα πράγματα στην πράξη είναι αρκετές φορές πιο δύσκολα από τη θεωρία.

Υπάρχει μία τάση στους ανθρώπους να μη βαθμολογούν ένα προϊόν όταν είναι πολύ ευχαριστημένοι από εκείνο αλλά να το βαθμολογούν συχνότερα όταν είναι δυσαρεστημένοι. Επίσης αρκετές φορές άτομα που τους αρέσει το ίδιο ένα προϊόν άλλος θα το βαθμολογήσει με αυστηρά κριτήρια και άλλος με πιο γενναιόδωρο τρόπο. Όπως θα δούμε στην επόμενη υποενότητα λύση στο θέμα αυτό μπορεί να δώσει η μέθοδος της κανονικοποίησης.

Ένα άλλο σημαντικό ζήτημα του explicit feedback προκύπτει από το δυναμικό χαρακτήρα που έχει η έννοια της αξιολόγησης. Με το πέρασμα του χρόνου ένας χρήστης μπορεί να αξιολογήσει με διαφορετικό τρόπο ένα αντικείμενο που έχει πρόσφατο από ένα αντικείμενο που είχε αξιολογήσει πριν 2 χρόνια. Κατά αυτό τον τρόπο μεταβάλλεται ουσιαστικά η κλίμακα βαθμολόγησης με το πέρασμα του χρόνου.

Επίσης είναι πιθανό να αλλάξουν οι προτιμήσεις του όσο μεγαλώνει, την ίδια στιγμή που οι παλαιότερες αξιολογήσεις παραμένουν αμετάβλητες. Αυτό πολλές φορές έχει σαν αποτέλεσμα ο χρήστης να δέχεται προτάσεις που πλέον δεν του είναι τόσο ελκυστικές, μειώνοντας έτσι την αποδοτικότητα του συστήματος συστάσεων.

Ο έμμεσος τρόπος αξιολόγησης από την άλλη μεριά δε λαμβάνει υπόψιν τις αξιολογήσεις του χρήστη αλλά μετράει πόσο χρόνο πέρασε ο χρήστης σε μία συγκεκριμένη σελίδα, πόσα θεάσεις έκανε σε ένα προϊόν, αν το έβαλε στο καλάθι και αν εντέλει προχώρησε σε κάποια αγορά. Ως προς αυτή την οπτική είναι λιγότερο επηρεασμένος από τη ψυχοσύνθεση του χρήστη που αξιολογεί. Επίσης είναι σημαντικό το γεγονός ότι δεν χρειάζεται να κάνει ο ίδιος κάποια ενέργεια για να δώσει feedback για το αν του άρεσε ένα προϊόν, καθώς ο αλγόριθμος βασίζεται στις ενέργειες που ούτως ή άλλως θα έκανε για να εκπαιδευτεί.

Παρόλα αυτά σε αυτή την περίπτωση προκύπτουν άλλα ζητήματα όπως για παράδειγμα ό,τι δε μπορούμε να γνωρίζουμε αν ο χρήστης που είναι πολλή ώρα σε ένα προϊόν το κοιτάει όντως ή απλά έχει αφήσει ανοιχτή την οθόνη και στην περίπτωση που όντως το κοιτάει πολλή ώρα δεν μπορούμε να γνωρίζουμε αν είναι ένα προϊόν που τελικά του αφήνει θετική ή αρνητική εντύπωση.

Άλλες φορές ο χρήστης κάνει κατάλαθος κλικ σε ένα προϊόν το οποίο δεν σκοπεύει να δει ή μπορεί να κάνει μία έρευνα αγοράς για να αγοράσει κάτι σε κάποιον άλλον. Ένα ακόμα παράδειγμα που περιπλέκει την κατάσταση είναι η παράλληλη χρήση ενός υπολογιστή από διάφορα άτομα αν πρόκειται για οικογένεια. Και οι δύο αυτές προσεγγίσεις έχουν τα δικά τους πλεονεκτήματα και μειονεκτήματα και ο βέλτιστος τρόπος προσέγγισης είναι συνδυασμός τους για την καλύτερη κατανόηση των προτιμήσεων του χρήστη [31].

2.6 Κανονικοποίηση Βαθμολογιών

Μία λύση στο πρόβλημα του διαφορετικού τρόπου αξιολόγησης από χρήστη σε χρήστη μας δίνει η κανονικοποίηση των αξιολογήσεων. Υπάρχουν διάφορες τεχνικές που χρησιμοποιούνται για να εξισορροπήσουν τη διαφορά στη βαθμολογία από διαφορετικούς χρήστες, συμπεριλαμβανομένης της κλιμάκωσης χαρακτηριστικών, το συντελεστή διακύμανσης καθώς και μελετημένα υπολείμματα (studentized residual). Σε αυτή τη διπλωματική θα μιλήσουμε για δύο τεχνικές, την mean-centering και την Z-Score.

Στην τεχνική mean-centering αρχικά βγαίνει ο μέσος όρος των αξιολογήσεων του κάθε χρήστη και στη συνέχεια αφαιρείται κάθε αξιολόγηση από το μέσο. Αν η διαφορά είναι θετική σημαίνει ότι είναι πάνω από το μέσο όρο και άρα θεωρούμε θετική την αξιολόγηση του χρήστη, αν είναι κάτω από το μέσο όρο θεωρούμε αρνητική την αξιολόγηση του χρήστη και αν είναι ίση με μηδέν η αξιολόγηση θεωρείται κανονική [49].

Η φόρμουλα για την user-based προσέγγιση δίνεται από τον παρακάτω τύπο [35]:

$$h(r_{ui}) = r_{ui} - \bar{r}_u \quad (9)$$

με $h(r_{ui})$ την συνάρτηση που υπολογίζει την προσαρμοσμένη στο μέσο αξιολόγηση του χρήστη u για το αντικείμενο i , r_{ui} την πραγματική αξιολόγηση του χρήστη u για το αντικείμενο i και \bar{r}_u την μέση τιμή αξιολογήσεων του χρήστη u .

Η φόρμουλα για την item-based προσέγγιση αλλάζει σε [35]:

$$h(r_{ui}) = r_{ui} - \bar{r}_i \quad (10)$$

με \bar{r}_i τη μέση αξιολόγηση ενός αντικειμένου i από όλους τους χρήστες που το έχουν αξιολογήσει.

Μία όμοια τεχνική είναι η Z-Score κανονικοποίηση. Η διαφορά της σε σχέση με την mean-centering προσέγγιση είναι ότι λαμβάνει υπόψιν και το πόσο ευρεία είναι η κλίμακα αξιολόγησης ενός χρήστη. Έστω ότι έχουμε για παράδειγμα δύο χρήστες. Ο χρήστης A βαθμολογεί με τιμές από το 1 έως το 5 ενώ ο χρήστης B βάσει σταθερά 3. Αν και οι 2 χρήστες βαθμολογήσουν με 5 ένα αντικείμενο τότε για τον χρήστη B το αντικείμενο αυτό βρίσκεται σε μεγαλύτερη εκτίμηση από τον χρήστη A λόγω της κλίμακας βαθμολόγησης τους.

Στη Z-Score κανονικοποίηση η φόρμουλα για την user-based προσέγγιση δίνεται από τον παρακάτω τύπο [35]:

$$h(r_{ui}) = \frac{r_{ui} - \bar{r}_u}{\sigma_u} \quad (11)$$

με σ_u την τυπική απόκλιση των αξιολογήσεων ενός χρήστη u .

Η φόρμουλα για την item-based προσέγγιση αλλάζει σε [35]:

$$h(r_{ui}) = \frac{r_{ui} - \bar{r}_i}{\sigma_i} \quad (12)$$

με σ_i την τυπική απόκλιση των αξιολογήσεων ενός αντικειμένου i .

2.7 Προβλήματα Συνεργατικού Φιλτραρίσματος

Στην προηγούμενη υποενότητα περιγράψαμε κάποιες δυσκολίες που προκύπτουν από τις βαθμολογίες των χρηστών και είναι πιθανό να δυσκολέψουν το σύστημα συστάσεων να κάνει την καλύτερη δυνατή πρόταση για το χρήστη. Πέραν αυτών υπάρχουν και άλλες δυσκολίες - προκλήσεις που καλείται να αντιμετωπίσει ένα σύστημα συστάσεων τις οποίες θα περιγράψουμε σε αυτή την υποενότητα.

2.7.1 Αραιότητα Πίνακα

Ένα πολύ σημαντικό πρόβλημα είναι η αραιότητα του πίνακα. Υπάρχουν πολλοί χρήστες και αντικείμενα στον πίνακα R, όμως, οι περισσότεροι χρήστες δεν έχουν αξιολογήσει τα περισσότερα αντικείμενα με αποτέλεσμα τα περισσότερα κελιά του πίνακα να είναι κενά. Έτσι ενώ μπορεί να υπάρχουν πολλοί όμοιοι χρήστες με έναν χρήστη που θέλουμε να του κάνουμε μία πρόταση, το σύστημα δε μπορεί να τους εντοπίσει λόγω έλλειψης πληροφορίας. Αυτό έχει σαν συνέπεια τη μείωση της "κάλυψης" των αντικειμένων που μπορούν να προταθούν [26].

Τα πιθανά αντικείμενα μειώνονται λόγω των υπερβολικά λίγων αξιολογήσεων που λαμβάνουν το οποίο προκαλεί στο σύστημα δυσκολία στο να τα προτείνουν σε άλλους χρήστες. Για αυτό το λόγο γίνεται σαφές ότι πέραν του πόσο καλός είναι ένας αλγόριθμος ή μία τεχνική το πιο σημαντικό από όλα είναι να υπάρχει επαρκής πληροφορία, δηλαδή να έχουμε ένα καλό σύνολο δεδομένων. Όσο περισσότερη πληροφορία έχουμε τόσο καλύτερα αναμένονται τα αποτελέσματα μας.

Μία λύση στο πρόβλημα της αραιότητας του πίνακα μπορεί να δώσει η απόκτηση επαρκών αξιολογήσεων και για αυτό το λόγο πολλές εφαρμογές έχουν μηχανισμούς να ενθαρρύνουν τους χρήστες να κάνουν αξιολογήσεις. Γενικά θα λέγαμε ότι είναι δύσκολο να αποκτηθούν πολλές αξιολογήσεις από ένα χρήστη λόγω του υψηλού κόστους που απαιτεί αυτή η διαδικασία και λόγω ότι μία τέτοια επιλογή απαιτεί ιδιαίτερη προσοχή για το τι θα βάζαμε να αξιολογήσει ένας συγκεκριμένος χρήστης.

Αν για παράδειγμα ένας χρήστης έχει βαθμολογήσει πολλές ταινίες που ανήκουν στην κατηγορία του θρίλερ τότε να του ζητήσουμε να αξιολογήσει επιπλέον ταινίες της κατηγορίας θρίλερ δεν προσδίδει κάποια σημαντική βοήθεια για την πρόβλεψη αξιολόγησης παρόμοιων ταινιών. Θα είχε μεγαλύτερη χρησιμότητα από την άλλη να του ζητήσουμε να αξιολογήσει λιγότερο δημοφιλή είδη βοηθώντας έτσι στην αξιολόγηση τέτοιου είδους ταινιών.

Εδώ προκύπτει ένα άλλο ζήτημα καθώς δεν μπορούμε να γνωρίζουμε εκ των προτέρων αν ο χρήστης έχει παρακολουθήσει την ταινία που του ζητάμε να αξιολογήσει. Βλέπουμε δηλαδή ότι ανάλογα την στρατηγική που πάμε να ακολουθήσουμε υπάρχουν πλεονεκτήματα και μειονεκτήματα [24].

2.7.2 Εγγραφή Νέου Χρήστη/Αντικειμένου

Ένα από τα σημαντικότερα θέματα στα συστήματα συστάσεων είναι η έλλειψη πληροφορίας για ένα καινούργιο χρήστη που καταχωρείται στο σύστημα. Ένας νέος χρήστης δεν έχει καταχωρήσει αξιολογήσεις και συνεπώς το σύστημα δεν μπορεί να κάνει προτάσεις με βάση το ιστορικό του και τις προτιμήσεις του. Η ανικανότητα του συστήματος να προτείνει προϊόντα σε νέους χρήστες είναι γνωστό στην ορολογία ως πρόβλημα Cold Start [30].

Ένας τρόπος να αντιμετωπιστεί αυτό το πρόβλημα είναι να αναγκάζεις νέους χρήστες να αξιολογούν κάποια αντικείμενα πριν αρχίσουν να χρησιμοποιούν το σύστημα ή να απαντάνε σε μία σειρά δημογραφικών ερωτήσεων και με βάση αυτά χρησιμοποιώντας κάποια στερεότυπα να κάνουν κάποιες προτάσεις. Αυτού του είδους η στρατηγική απαιτεί προσπάθεια, καθώς δεν είναι εύκολο να καθορίσουμε ποια προϊόντα να παρέχουμε για αξιολόγηση και τα στερεότυπα κάποιες φορές μπορεί να είναι λανθασμένα ή και προσβλητικά. Για παράδειγμα μπορούμε να υποθέσουμε ότι ένας έφηβος 14-16 χρόνων είναι πιο πιθανό να ακούει ραπ από κλασική μουσική αλλά δε μπορεί κανένας να το επιβεβαιώσει με σιγουριά [19].

Μία ενδιαφέρουσα λύση με πολύ καλά αποτελέσματα παρουσιάζεται στο [27]. Σύμφωνα με αυτή μοντελοποιούμε το σύστημα ώστε να προτείνει προϊόντα σε ένα γκρουπ χρηστών ώστε να τους διατηρεί όλους χαρούμενους. Η μοντελοποίηση γίνεται μέσα από τη συγκέντρωση πληροφοριών για τις ατομικές προτιμήσεις του κάθε χρήστη και την εφαρμογή κατάλληλων τεχνικών ώστε να μετατραπούν σε ένα ομοιογενές σύνολο. Έτσι θεωρούμε ότι ο νέος χρήστης θα έχει όμοιες προτιμήσεις με κάποιον από το γκρουπ χρηστών (χωρίς να ξέρουμε ποιον) και αφού κάθε χρήστης του γκρουπ είναι χαρούμενος από τις προτάσεις του συστήματος, θεωρούμε ότι θα είναι και ο νέος χρήστης. Σταδιακά το σύστημα θα μπορεί να αντλήσει περισσότερες πληροφορίες για το νέο χρήστη μέχρι να προσαρμοστεί πλήρως στις προτιμήσεις του.

Παρόμοιο θέμα υπάρχει και με την εισαγωγή ενός νέου αντικειμένου στο σύστημα, με το πρόβλημα αυτό να ορίζεται ως Early Rater [29]. Τα νέα αντικείμενα δεν έχουν καμία αξιολόγηση και συνεπώς δεν θα προταθούν σε κανέναν. Ένας τρόπος σε αυτή την περίπτωση είναι να αναγκάσουμε τους χρήστες να αξιολογήσουν αυτά τα νέα αντικείμενα ή να τα προβάλλουμε με πρώτη προτεραιότητα στους χρήστες, με τον κίνδυνο πάντοτε να απευθυνόμαστε σε χρήστες που δεν γνωρίζουν το συγκεκριμένο αντικείμενο. Για την αντιμετώπιση αυτού του προβλήματος έχουν προταθεί στη βιβλιογραφία διάφορες τεχνικές εξόρυξης που δεν έχουν να κάνουν με το συνεργατικό φιλτράρισμα και ο αναγνώστης μπορεί να βρει περισσότερες πληροφορίες στο [28].

3 Μετρικές Αξιολόγησης

3.1 Diversity

Η έννοια του diversity (ποικιλομορφία) αναπτύχθηκε αρχικά στη βιβλιογραφία στο πεδίο της ανάκτησης πληροφορίας αρχείων. Η ανάγκη για μία τέτοια έννοια προέρχεται από το γεγονός ότι μία αναζήτηση του χρήστη μπορεί να επιδέχεται πολλών διαφορετικών ερμηνειών και είναι δύσκολο για το σύστημα να ανακτήσει το κατάλληλο αρχείο. Για παράδειγμα αν ένας χρήστης ψάξει σε μία μηχανή αναζήτησης τον όρο "τοξότης" ο όρος αυτός μπορεί να αναφέρεται στον αστερισμό, σε ένα από τα αστρολογικά ζώδια ή σε κάποιον πολεμιστή οπλισμένο με τόξο.

Στο παραπάνω παράδειγμα, που είναι τυπικό παράδειγμα αφηρημένης αναζήτησης του χρήστη, βλέπουμε ότι είναι αδύνατο για το σύστημα να καταλάβει για ποιον από τους τρεις τομείς ενδιαφέρεται πραγματικά ο χρήστης. Μία λύση στην περίπτωση μίας τέτοιας αναζήτησης είναι το σύστημα να επιστρέψει στο χρήστη αποτελέσματα από ένα μεγαλύτερο εύρος τομέων. Έτσι είναι αρκετά πιθανότερο για τον χρήστη να βρει πληροφορίες για αυτό που τελικά ψάχνει.

Ανάλογη ανάγκη υπάρχει και στο ερευνητικό πεδίο των συστημάτων συστάσεων όπου επίσης γίνονται προσπάθειες για βελτιστοποίηση της ποικιλομορφίας των αποτελεσμάτων μίας μηχανής αναζήτησης. Στη βιβλιογραφία η μέτρηση του diversity γίνεται παίρνοντας το μέσο όρο ή το σύνολο της ανομοιότητας μεταξύ των αντικειμένων της λίστας. Στο [50] δίνεται ο ορισμός μέτρησης του diversity σαν το μέση απόσταση ανά ζευγάρια μεταξύ των αντικειμένων στη λίστα:

$$Diversity(R) = \frac{\sum_{i \in R} \sum_{j \in R \setminus \{i\}} dist(i, j)}{|R|(|R| - 1)} \quad (13)$$

Στα συστήματα συστάσεων η συνάρτηση που μετράει απόσταση μεταξύ δύο αντικειμένων μπορεί να γίνει με διάφορους τρόπους που συνήθως καθορίζονται από το είδος της πληροφορίας των αντικειμένων. Για παράδειγμα αν έχουμε αξιολογήσεις χρηστών όπως στο [51] η συνάρτηση απόστασης βασίζεται στις προβλέψεις από τις βαθμολογίες των χρηστών, οι οποίες προκύπτουν χρησιμοποιώντας τη μέθοδο των κοντινών γειτόνων.

Σε άλλες μελέτες έχει χρησιμοποιηθεί η συσχέτιση Pearson [52], η απόσταση συνημιτόνου [58] και η απόσταση hamming [53]. Παρόμοιες μετρικές απόστασεις έχουν χρησιμοποιηθεί και σε περιπτώσεις που δεν το σύνολο δεδομένων δεν αποτελείται από αξιολογήσεις αλλά από γραπτό κείμενο. Τέλος υπάρχουν και μελέτες που χρησιμοποιούν model-based μεθόδους και η απόσταση υπολογίζεται με τη βοήθεια των λανθάνοντων παραγόντων [54].

3.1.1 Reranking Τεχνικές Diversity

Στα συστήματα συστάσεων οι τεχνικές που αφορούν το diversity χωρίζονται σε δύο κατηγορίες. Η πρώτη κατηγορία στηρίζεται στην ισορροπία σχετικότητας και ποικιλίας των αντικειμένων ενός συστήματος συστάσεων. Σε αυτή την περίπτωση αρχικά τρέχει ο υπάρχοντας αλγόριθμος (για παράδειγμα ο αλγόριθμος k-NN) και δημιουργεί μία λίστα από προτάσεις βασιζόμενος στην σχετικότητα των προϊόντων με αυτά που αρέσουν στο χρήστη. Στη συνέχεια σε αυτή την λίστα γίνεται αναδιάταξη των προτάσεων λαμβάνοντας υπόψιν αυτή τη φορά και την έννοια της ποικιλίας.

Η δεύτερη κατηγορία με την οποία δεν θα ασχοληθούμε στην παρούσα εργασία αναπτύχθηκε στη βιβλιογραφία τα τελευταία χρόνια στηρίζεται σε νέα μοντέλα που στοχεύουν στην ποικιλία των αντικειμένων που προτείνονται στο χρήστη.

Μία δημοφιλής τεχνική της πρώτης κατηγορίας είναι η Μέγιστη Οριακή Συνάφεια (Maximal Marginal Relevance) που χρησιμοποιεί μία greedy ανακατατάξη [56]. Στο συγκεκριμένο paper που ασχολείται με το πεδίο της ανάκτησης πληροφοριών παρουσιάζεται μία μέθοδος που προσπαθεί να μειώσει τον πλεονασμό και να αυξήσει τη διαφορετικότητα στα αποτελέσματα που γυρνάει στο χρήστη. Η αντικειμενική συνάρτηση είναι ένας γραμμικός συνδυασμός μεταξύ σχετικότητας και διαφορετικότητας και εκφράζεται από τον παρακάτω τύπο:

$$MMR = Arg \max_{D_i \in R \setminus \{S\}} [\lambda \cdot (Sim_1(D_i, Q) - (1 - \lambda) \max_{D_j \in S} Sim_2(D_i, D_j))]$$

όπου Q είναι το query του χρήστη, D το σύνολο των αρχείων που σχετίζονται με το ερώτημα Q, S το υποσύνολο των αρχείων στο R που έχουν επιλεγθεί ήδη, $R \setminus \{S\}$ τα μη επιλεγμένα αρχεία στο R και λ η σταθερά που παίρνει τιμές στο διάστημα [0,1] και ρυθμίζει την διαφορετικότητα των αποτελεσμάτων.

Στο [55] παρουσιάζεται η μέθοδος a-nDGG που μετράει το diversity σε σχέση με τη σχετικότητα εξετάζοντας τα αποτελέσματα που γυρνάει μία διαδικτυακή αναζήτηση του χρήστη. Στη μελέτη αυτή ορίζεται σαν n_i η πληροφορία που περιέχει το κάθε αρχείο, u η πληροφορία που χρειάζεται για έναν χρήστη για να διατυπώσει το query και d ένα έγγραφο που μπορεί να είναι σχετικό ή όχι με το u .

Στη συνέχεια με την εφαρμογή της μεθόδου DGG ερευνάται η πιθανότητα να υπάρχει ένα n_i έτσι ώστε να είναι η τομή του u και του d , δηλαδή την πιθανότητα η πληροφορία που χρειάζεται στο query του χρήστη να περιέχει το περιεχόμενο n_i και το έγγραφο να περιέχει επίσης αυτή την πληροφορία. Σε κάθε διάταξη τα αρχεία που βρίσκονται χαμηλότερα δέχονται μία μικρή ποινή αντανakλώντας την επιπρόσθετη προσπάθεια του χρήστη για να τα ανοίξει.

Στο paper [60] περιγράφεται μία μέθοδος που προσπαθεί να δώσει λύση στο πρόβλημα του filter bubble συνδυάζοντας την ποικιλία με την ακρίβεια. Το filter bubble είναι μία κατάσταση απομόνωσης που προέρχεται από προσωποποιημένη αναζήτηση όταν ένας αλγόριθμος προσπαθεί να μαντέψει τι πληροφορίες θα ήθελε ένας χρήστης να δει βασιζόμενος σε πληροφορίες που έχει ήδη για το χρήστη. Με άλλα λόγια το περιεχόμενο γίνεται αρκετά εξατομικευμένο και περιορίζει τα αποτελέσματα που γυρνάει στο χρήστη.

Λύση στο πρόβλημα αυτό προσπαθεί να δώσει η ανάπτυξη ενός αλγορίθμου που ισορροπεί την ποικιλία με την ακρίβεια βασιζόμενο σε κανόνες εξόρυξης διαδοχικών κανόνων και στους k κοντινούς γείτονες και στο τέλος εφαρμόζει μία greedy προσέγγιση για την βελτιστοποίηση των αποτελεσμάτων του.

Άλλη μία αξιοσημείωτη μελέτη που χρησιμοποιεί τον greedy αλγόριθμο παρουσιάζεται στο [59] όπου χρησιμοποιείται ένα σύστημα συστάσεων που βασίζεται στην αλληλεπίδραση με το χρήστη. Μέσω ενός μηχανισμού κάθε φορά που το σύστημα προτείνει μία λίστα με αντικείμενα ο χρήστης δίνει πληροφορίες για το πόσο του άρεσε το περιεχόμενο και βασιζόμενο σε αυτό το προτείνονται νέα αντικείμενα. Η διαδικασία ολοκληρώνεται όταν ο χρήστης μείνει ευχαριστημένος και η μετρική απόστασης που χρησιμοποιείται είναι η Hamming.

Ενώ οι παραπάνω μελέτες έχουν σαν κοινό παράγοντα ότι εξετάζουν την αλληλεπίδραση μεταξύ ακρίβειας και ποικιλίας στο [57] παρουσιάζεται μία διαφοροποιημένη μέθοδος η οποία λαμβάνει υπόψιν όχι μόνο την ακρίβεια και την ποικιλία αλλά και την αντιστάθμιση μεταξύ μεταξύ εκμετάλλευσης του προφίλ ενός χρήστη και της εξερεύνησης νέων προϊόντων. Για να το πετύχει αυτό βασίστηκε στην παρακάτω αντικειμενική συνάρτηση:

$$X_{PLODIV(i,U,R)} = a \cdot rel(i) + (1 - a) \cdot div(i, R) \cdot X_{PLO(i,U)}$$

$$X_{PLO(i,U)} = \beta \cdot xploit(i, U) + (1 - \beta)xplore(i, U)$$

όπου $rel(i)$, $div(i,R)$, $exploitation(i,U)$ και $exploration(i,U)$ είναι οι 4 βασικοί άξονες και a, β οι παράμετροι που καθορίζουν την ανταλλαγή μεταξύ σχετικότητας - ποικιλίας και εκμετάλλευσης-εξερεύνησης.

Μία διαφορετική οπτική παρουσιάζεται στο [58], στο οποίο υποστηρίζεται ότι η προσέγγιση αναδιάταξης ενός αλγορίθμου επικεντρώνεται στην οπτικοποίηση της ακρίβειας χωρίς να δίνεται μεγάλη βαρύτητα στο diversity και το novelty. Η προσέγγιση που προτείνουν οι συγγραφείς ώστε να λύσουν αυτό το πρόβλημα είναι υβριδική, συνδυάζοντας διαφορετικούς υπάρχοντες αλγορίθμους που διαφέρουν ως προς το επίπεδο ακρίβειας, diversity και novelty.

Το μεγάλο πρόβλημα που αντιμετωπίζουν είναι ότι αυτές οι 3 μετρήσεις μπορεί να οδηγήσουν σε ένα πρόβλημα αντικρουόμενων στόχων αφού η βελτίωση του ενός μπορεί να αποφέρει σημαντική μείωση στα άλλα δύο. Για να αντιμετωπίσουν αυτό το πρόβλημα χρησιμοποιούν την προσέγγιση Strength Pareto στην οποία διατηρούνται μόνο όσα υβρίδια δεν κυριαρχούν έναντι των άλλων. Έτσι καταφέρνουν να συνδυάσουν αλγορίθμους που βελτιώνουν μία μετρική χωρίς να πλήττονται σε μεγάλο βαθμό οι υπόλοιπες και μάλιστα με τέτοιο τρόπο ώστε το σύστημα να προσαρμόζεται δυναμικά στις ανάγκες του κάθε χρήστη.

3.2 Novelty

Το novelty είναι μία ακόμα μετρική που προέρχεται από το πεδίο ανάκτησης πληροφορίας. Στη βιβλιογραφία συναντάμε διαφορετικούς ορισμούς για το novelty και αρκετές φορές γίνεται άμεση συσχέτιση με το serendipity. Από τις πρώτες αναφορές που μιλούν για την αξία που μπορεί να προσδώσει το novelty είναι οι [61] και [62].

Το [62] ασχολείται με την ανάκτηση αρχείων σχετικές με τις πληροφορίες που δίνονται από το χρήστη. Το φιλτράρισμα αυτών των αρχείων θα μπορούσε να παρέχει καλύτερες υπηρεσίες στους χρήστες αν χωριζόταν σε τρεις κατηγορίες, μία για τα αρχεία που δεν είναι σχετικά, μία για αρχεία που είναι σχετικά αλλά δεν περιέχουν νέες πληροφορίες και τέλος τα αρχεία που είναι σχετικά και περιέχουν νέες πληροφορίες. Το μέτρο σύγκρισης για το αν οι πληροφορίες είναι σχετικές ή όχι καθορίζεται από τα αρχεία που έχει δει ο χρήστης στο ιστορικό του. Έτσι θα λέγαμε ότι το novelty παρουσιάζεται σαν μία λύση που θα βοηθούσε το χρήστη να διαλέξει αρχεία με βάση τη νέα πληροφορία, γλυτώνοντας τον έτσι από το να δει αρχεία που δεν του δίνουν καμία νέα πληροφορία.

Μέσα από τη βιβλιογραφία βλέπουμε αρκετές διαφορετικές οπτικές σχετικά με το πως ορίζεται το novelty. Σε κάποιες ορίζεται σαν ένα αντικείμενο που μπορεί να είναι άγνωστο για το χρήστη ή κάτι που δεν είχε ξαναδεί πριν [63]. Στο [64] δίνονται τρεις παράμετροι να είναι άγνωστο για το χρήστη, κοντά τις προτιμήσεις του και ταυτόχρονα να μην έχει μεγάλη ομοιότητα με αντικείμενα στο ιστορικό του. Βασιζόμενοι σε αυτές τις παραμέτρους μπορούμε εύκολα να κατανοήσουμε για ποιο λόγο συνδέεται τόσο στενά με το serendipity. Όταν σε ένα χρήστη προτείνεται ένα αντικείμενο που δεν είναι όμοιο με τα αντικείμενα στο ιστορικό του είναι πιθανό να αποτελέσει έκπληξη για εκείνον, κάτι που αποτελεί κυριότερο χαρακτηριστικό του serendipity.

Για να διαχωρίσουμε αυτές τις δύο έννοιες θα υιοθετήσουμε τον ορισμό που δίνεται στο [65] σύμφωνα με το οποίο ένα αντικείμενο μπορεί να είναι νέο στο χρήστη χωρίς απαραίτητα να είναι έκπληξη για εκείνον. Για να κάνει πιο κατανοητό το διαχωρισμό δίνει το εξής παράδειγμα: Έστω ότι έχουμε ένα σύστημα συστάσεων που προτείνει ταινίες που έχουν ως σκηνοθέτη τον αγαπημένο σκηνοθέτη του χρήστη. Αν ο χρήστης δεν έχει δει κάποια από αυτές τις ταινίες η ταινία θα είναι νέα για εκείνον αλλά πιθανότατα δε θα αποτελεί έκπληξη καθώς πρόκειται για ταινία του αγαπημένου του σκηνοθέτη. Ίσως μάλιστα είναι μία ταινία που θα μπορούσε να έχει ανακαλύψει και μόνος του ο χρήστης. Αντιθέτως μία ταινία νέα από άλλο σκηνοθέτη θα αποτελούσε μεγαλύτερη έκπληξη για το χρήστη.

Το επόμενο ερώτημα που καλούμαστε να απαντήσουμε είναι για το πως ορίζεται ένα άγνωστο αντικείμενο για το χρήστη. Σε συστήματα offline είναι πολύ δύσκολο να ξέρουμε αν ένα αντικείμενο είναι ή όχι νέο για τον χρήστη και ο πιο ασφαλής τρόπος θα ήταν να ρωτήσουμε τον ίδιο τον χρήστη να μας δώσει αυτή την πληροφορία. Κάτι τέτοιο θα μπορούσε να γίνει για παράδειγμα με ερώτηση προς το χρήστη μετά την πρόταση ενός αντικειμένου. Ακόμα και σε αυτή την περίπτωση, όμως, δεν είμαστε σίγουροι ότι ο χρήστης θα μας έδινε τη γνώμη του και πρόκειται για μία διαδικασία που μάλλον θα την έβρισκε βαρετή ή κουραστική.

Ένας συνηθισμένος τρόπος για να μετρήσουμε το πόσο νέο φαίνεται ένα αντικείμενο για το χρήστη είναι μετρώντας το συνολικό αριθμό αξιολογήσεων του. Βέβαια ένα αντικείμενο που δεν έχει αξιολογηθεί δε σημαίνει απαραίτητα ότι είναι άγνωστο στο χρήστη αλλά ένα

αντικείμενο που δεν έχει πολλές ψήφους και θεωρείται λιγότερο δημοφιλές τείνει να είναι πιο νέο για τον χρήστη από ένα δημοφιλές. Αυτή την προσέγγιση θα υιοθετήσουμε και εμείς στην παρούσα εργασία ορίζοντας το novelty βασιζόμενο στον αριθμό αξιολογήσεων ακολουθώντας και άλλες μελέτες της βιβλιογραφίας [66]. Έτσι έχουμε τον παρακάτω ορισμό για να μετρήσουμε το novelty μίας λίστας R:

$$Novelty(R) = \frac{\sum_{i \in R} -\log_2 p(i)}{|R|} \quad (14)$$

με $p(i)$ να είναι η αναλογία των χρηστών που έδωσαν αξιολόγηση σε ένα αντικείμενο i σε σχέση με τους συνολικούς χρήστες της λίστας R. Ο αρνητικός λογάριθμος $p(i)$ χρησιμοποιείται για να δοθεί μεγαλύτερη βαρύτητα στα σπάνια αντικείμενα [2].

3.2.1 Τεχνικές Βελτίωσης Novelty

Στη βιβλιογραφία υπάρχουν αρκετές μελέτες που προσπαθούν να βελτιώσουν το novelty στηριζόμενοι στον ορισμό που ακολουθήσαμε παραπάνω, δίνοντας δηλαδή μεγαλύτερη έμφαση στις προτάσεις αντικειμένων που είναι πιο σπάνια. Τα αντικείμενα αυτά μπορούν να συναντηθούν στην βιβλιογραφία και με τον όρο αντικείμενα της 'μακριάς ουράς' (long tail).

Η μακριά ουρά αποτελείται από ένα μικρό αριθμό δημοφιλών αντικειμένων και τα υπόλοιπα αντικείμενα στην ουρά είναι αυτά που δεν είναι δημοφιλή και δεν πουλάνε τόσο πολύ [66]. Στο [67] ο συγγραφέας θέτει δύο βασικές προϋποθέσεις για αυτού του είδους τα αντικείμενα: Η πρώτη είναι να είναι διαθέσιμα και η δεύτερη να μπορούν να βρεθούν από τους χρήστες. Η πρώτη προϋπόθεση θα λέγαμε ότι καλύπτεται στις μέρες μας αφού πλέον είναι εύκολη η πρόσβαση σε κάθε είδους αντικείμενο. Για να ικανοποιήσουμε την δεύτερη προϋπόθεση χρειαζόμαστε τη βοήθεια των συστημάτων συστάσεων τα οποία θα προτείνουν στο χρήστη τα κρυμμένα διαμάντια της μακριάς ουράς.

Στη βιβλιογραφία υπάρχουν αρκετές μελέτες που ασχολούνται με την μεθόδους προτάσεων αντικειμένων από αντικείμενα της μακριάς ουράς. Οι μέθοδοι προτάσεων τέτοιου είδους αντικειμένων χωρίζονται σε τέσσερις βασικές κατηγορίες: οι βασιζόμενες στην ομαδοποίηση, οι βασιζόμενες στο γράφημα, οι βασιζόμενες σε πολλούς στόχους βελτιστοποίησης και οι μέθοδοι που βασίζονται σε τεχνικές deep learning. Παρακάτω θα παρουσιάσουμε κάποιες αξιολογούμενες μελέτες από τις δύο πρώτες κατηγορίες. Για περισσότερες πληροφορίες σχετικά με τις υπόλοιπες κατηγορίες ο αναγνώστης μπορεί να ανατρέξει στο [67].

Οι Park και Tuzhilin πρότειναν πρώτοι μεθόδους βασισμένες στην ομαδοποίηση για τις προτάσεις αντικειμένων της μακριάς ουράς. Μία εξ αυτών ήταν η Clustered Tail μέθοδος. Στη μέθοδο αυτή χωρίζονται όλα τα αντικείμενα σε διαφορετικά γκρουπ και τότε χτίζεται ένα μοντέλο πρόβλεψης για κάθε γκρουπ. Η κατηγορία των αντικειμένων της μακριάς ουράς καθορίζεται από τον αριθμό των αξιολογήσεων όλων των αντικειμένων. Για παράδειγμα αν ο αριθμός αξιολογήσεων των αντικειμένων είναι 30 τότε τα αντικείμενα με περισσότερες από 30 ψήφους χωρίζονται σε short-head αντικείμενα και τα υπόλοιπα θεωρούνται αντικείμενα της μακριάς ουράς. Στα πειράματα που κάναμε με τη μέθοδο αυτοί παρατηρήθηκε βελτίωση στην ακρίβεια των προβλέψεων, ωστόσο στο κομμάτι του διαχωρισμού των αντικειμένων υπάρχουν πολλά περιθώρια βελτίωσης.

Η παραπάνω μελέτη αν και εισήγαγε την μέθοδο για προτάσεις αντικειμένων της μακριάς ουράς δεν επικεντρώθηκε στην αξιολόγηση της μετρικής novelty. Αντιθέτως οι Huang και Wu χρησιμοποίησαν το novelty και το diversity για να αξιολογήσουν τις μεθόδους που ανέπτυξαν [68]. Πιο συγκεκριμένα πρότειναν μία βελτιωμένη k-medoids μέθοδο ομαδοποίησης. Το μοντέλο τους βασιζόταν σε δύο μέρη, ένα για να εξάγουν τα προφίλ των προϊόντων και ένα να εξάγουν παρόμοια προϊόντα. Για κάθε ένα από τα μέρη αυτά ανέπτυξαν 4 μεθόδους και χρησιμοποίησαν μετρικές diversity και novelty για να τις συγκρίνουν [68].

Η δεύτερη κατηγορία αφορά μεθόδους βασιζόμενες στο γράφημα, δηλαδή την αναπαράσταση των διαδραστικών δεδομένων μεταξύ χρηστών και αντικειμένων με τη μορφή γραφημάτων, τα οποία ενισχύουν την διαθεσιμότητα των δεδομένων [67]. Στο [69] χρησιμοποιήθηκε ένα τριμερές γράφημα για να αντιπροσωπεύσουν τις σχέσεις μεταξύ χρηστών, αντικειμένων και τις κατηγορίες μεταξύ των αντικειμένων. Όταν ο χρήστης διαλέξει μια συγκεκριμένη κατηγορία αντικειμένου χρησιμοποιώντας το γράφημα, το αντικείμενο της μακριάς ουράς στη συλλογή των αντικειμένων αυτής της κατηγορίας μπορεί να προταθεί στο χρήστη. Η χρησιμοποιούμενη μέθοδος ήταν η αλυσίδα Markov και παρουσίασε αρκετά καλά αποτελέσματα.

Στο [71] προτάθηκε ένας αλγόριθμος που βασιζόταν στο κατανεμημένο βάρος σε ένα διμερές γράφημα. Το μοντέλο δούλευε αναθέτοντας βάρη σε αντικείμενα που είχαν βαθμολογηθεί από τον χρήστη που θέλαμε να προτείνουμε ένα αντικείμενο και μοίραζε ισόποσα τα βάρη κάθε αντικειμένου στους άλλους χρήστες που το είχαν βαθμολογήσει. Στη συνέχεια αυτά τα βάρη κατανέμονταν στα αντικείμενα των άλλων χρηστών. Με αυτό τον τρόπο ευνοούνταν η κατανομή βάρους σε κόμβους με λίγους συνδέσμους γραφημάτων και μπορούσε να προτείνει νέα αντικείμενα. Με απλά λόγια θα λέγαμε ότι χρησιμοποιώντας τη δημοφιλία των αντικειμένων κατάφερε να προτείνει λιγότερο δημοφιλή αντικείμενα.

Τέλος αξίζει να αναφέρουμε ότι τα τελευταία χρόνια παρατηρούνται αρκετές μελέτες που έχουν σα στοχο την εκπαίδευση ενός μοντέλου όχι μόνο για την βελτιστοποίηση όχι μόνο του novelty αλλά παράλληλα και άλλων μετρικών όπως η ακρίβεια, η κάλυψη και η ποικιλία. Στη παρούσα εργασία δεν έχουμε σκοπο την περαιτέρω ανάλυση αυτής της κατηγορίας, όμως, ο χρήστης μπορεί να δει στα [70], [72], [72] κάποιες από τις πιο ενδιαφέρουσες έρευνες.

3.3 Coverage

Το coverage σε ένα σύστημα συστάσεων είναι ένα μέτρο του τομέα των αντικειμένων στο οποίο μπορεί το σύστημα να κάνει προτάσεις. Στη βιβλιογραφία, ο όρος coverage συνδέεται κυρίως με δύο άξονες. Ο πρώτος έχει να κάνει με το ποσοστό των αντικειμένων για το οποίο το σύστημα είναι ικανό να παράγει μία πρόταση και ο δεύτερος αφορά το ποσοστό των διαθέσιμων αντικειμένων που προτείνονται αποδοτικά σε ένα χρήστη [3]. Προς διευκόλυνση μας θα ορίσουμε τον πρώτο άξονα σαν prediction coverage και τον δεύτερο σαν catalogue coverage.

Έστω ότι συμβολίζουμε με I το σύνολο των διαθέσιμων αντικειμένων και l_p το σύνολο των αντικειμένων για τα οποία μία πρόβλεψη μπορεί να γίνει. Τότε έχουμε [3]:

$$\text{Prediction Coverage} = \frac{l_p}{I} \quad (15)$$

Για να ετιμήσουμε το l_p υπάρχουν διάφοροι τρόποι. Θα αναφερθούμε στο συνεργατικό φιλτράρισμα, το οποίο χρησιμοποιούμε και στην παρούσα εργασία. Τα συστήματα συνεργατικού φιλταρίσματος μπορούν να κάνουν προβλέψεις για αντικείμενα που έχουν περισσότερο από ένα συγκεκριμένο αριθμό c αξιολογήσεων. Σε αυτή την περίπτωση μπορούμε να θεωρήσουμε το l_p σαν το σύνολο αντικειμένων που ξεπερνούν τον αριθμό αυτό. Παρόμοια λογική χρησιμοποιείται και στα υπόλοιπα συστήματα όπου ορίζεται κάποια συνθήκη και το σύνολο των διαθέσιμων αντικειμένων περιλαμβάνονται τα αντικείμενα που ικανοποιούν την συνθήκη.

Στον παραπάνω ορισμό μπορούμε να λάβουμε υπόψιν και την χρησιμότητα που έχει ένα αντικείμενο που προτείνουμε. Αυτό μπορούμε να το κάνουμε για παράδειγμα δίνοντας έμφαση σε αντικείμενα πέρα από τα δημοφιλή που λίγο πολύ ο χρήστης αναμένεται να γνωρίζει, αντικείμενα δηλαδή που συνδέονται και με το novelty για το οποίο μιλήσαμε στην προηγούμενη υποενότητα. Άλλη μία περίπτωση όπου το μεγαλύτερο εύρος προτάσεων θα ήταν χρήσιμο έχουμε όταν προτείνονται αντικείμενα σε ένα νέο χρήστη. Συχνά ο τρόπος μέτρησης για να συνδυάσουμε το coverage με τη σχετικότητα είναι η ακρίβεια, ιδιαίτερα σε συστήματα συνεργατικού φιλταρίσματος [3].

Ας δούμε τώρα τον ορισμό του Catalog Coverage. Έστω ότι R_u είναι το σύνολο όλων των προτάσεων για τον χρήστη u , U είναι το σύνολο των χρηστών και I το σύνολο των αντικειμένων. Τότε έχουμε:

$$\text{Catalog Coverage} = \frac{|\cup_{u \in U} R_u|}{|I|} \quad (16)$$

Το catalog coverage μπορεί να αποτελέσει ένα πολύτιμο τρόπο μέτρηση για τα συστήματα τα οποία προτείνουν μία λίστα αντικειμένων όπως για παράδειγμα τα δέκα πιο ταιριαστά αντικείμενα, μία παράμετρος που δε λαμβάνεται υπόψιν από το prediction coverage [3]. Στον παραπάνω τύπο παίζει σημαντικό ρόλο ο αριθμός των προτάσεων ανά χρήστη, ιδιαίτερα σε μικρά νούμερα όπου εκεί η αύξηση τους οδηγεί σε σημαντική αύξηση του catalog coverage. Επίσης μπορούν να γίνουν διάφορες δοκιμές για να διαπιστώσουμε σε ποιον αριθμό χρηστών η μετρική catalog coverage σταθεροποιείται και από ποιο αριθμό και πάνω σταματάει να έχει επιρροή στο catalog coverage.

3.3.1 Συσχέτιση Coverage με τις Υπόλοιπες Μετρικές

Η μετρική coverage σε αντίθεση με τις μετρικές novelty και diversity δεν ορίζεται σε ατομική επίπεδο χρήστη αλλά σε επίπεδο συστήματος. Παρόλα αυτά στη βιβλιογραφία οι 3 αυτές έννοιες είναι στενά συνδεδεμένες μεταξύ τους. Για παράδειγμα ένα υψηλό coverage ενός συνόλου δεδομένων συνεπάγεται ότι θα προταθούν και αντικείμενα που ανήκουν στην μακριά ουρά και είναι λιγότερη δημοφιλή και άρα μεγαλύτερο novelty.

Όπως είδαμε στην ενότητα του novelty, πολλές τεχνικές βελτίωσης του novelty σχετίζονται με το πόσο δημοφιλές είναι ένα αντικείμενο. Πάνω σε αυτό βασίζονται και οι περισσότερες τεχνικές βελτιστοποίησης του coverage, προσπαθούμε δηλαδή να μειώσουμε τα δημοφιλή αντικείμενα που προτείνουμε στο χρήστη προωθώντας τα πιο σπάνια. Μία τέτοια προσέγγιση ακολουθεί και η μελέτη [73]. Πιο συγκεκριμένα μέσα από μία αναδιάταξη της λίστας προτάσεων με σκοπό τη μεγιστοποίηση του coverage.

Στο [74] παρουσιάζεται μία πιο εξελιγμένη γραφική προσέγγιση που μοντελοποιεί το πρόβλημα της μεγιστοποίησης του coverage σαν πρόβλημα μεγιστοποίησης ροής δικτύου και έχει σχεδιαστεί ειδικά για τις κορυφαίες N διαφορετικές προτάσεις. Κάθε γράφος συνδέεται με ένα χρήστη μόνο αν το αντικείμενο φαίνεται να είναι σχετικό για το χρήστη. Τότε προσπαθούν να βρουν το μεγαλύτερο αριθμό ακμών με τα οποία συνδέονται χρήστες και αντικείμενα, με τέτοιο τρόπο ώστε κάθε χρήστης να έχει το πολύ N αντικείμενα. Η μέθοδος αυτή παρουσιάζει σημαντικά πλεονεκτήματα επί των μεθόδων αναδιάταξης που προσπαθούν να αντισταθμίσουν ακρίβεια και ποικιλομορφία.

Το coverage όπως προείπαμε έχει επίσης συσχετιστεί και με το diversity. Ενώ στη βιβλιογραφία είναι αποδεκτό ότι αυτές οι έννοιες αλληλοεπηρεάζονται δεν υπάρχουν επαρκή στοιχεία για το πως ακριβώς συμβαίνει αυτό. Για παράδειγμα στο [3] υποστηρίζεται ότι η αύξηση του coverage δεν είναι απαραίτητο ότι οδηγεί στην αύξηση του diversity αλλά η αύξηση του diversity συνεπάγεται μεγαλύτερο coverage.

Από την άλλη μεριά στο [73] υποστηρίζεται ότι η αύξηση στο diversity δε σημαίνει και απαραίτητα αύξηση του coverage. Σαν παράδειγμα παραθέτει μία περίπτωση όπου κάποια συγκεκριμένα diversity αντικείμενα προτείνονται σε πολλούς χρήστες. Έτσι το diversity παραμένει υψηλό αλλά το coverage περιορισμένο.

Από τα παραπάνω είναι φανερό ότι οι συσχετίσεις μεταξύ αυτών των μετρικών είναι ένα ανεξερεύνητο τοπίο στη βιβλιογραφία και χρήζει περαιτέρω μελέτης και διεξαγωγής πειραμάτων για να καταλήξουμε σε πιο ασφαλή συμπεράσματα.

3.4 Σχετικές Δουλειές πάνω στη Χρήση Νέων Μετρικών

Σε αυτή την ενότητα θα μιλήσουμε για διάφορες πειραματικές μετρήσεις που έχουν γίνει για τις μετρικές diversity, novelty και coverage σε διάφορα offline σύνολα δεδομένων.

Στο [58] παρουσιάζεται ένας συνδυασμός υπάρχοντων αλγορίθμων που κάθε ένας διαφέρει στην απόδοση ακρίβειας, diversity και novelty. Χρησιμοποιούνται διάφορων ειδών τεχνικές όπως η τεχνική των k-κοντινών γειτόνων, τεχνική βασιζόμενη στο περιεχόμενο, τεχνική βασιζόμενη στη δημοφιλία και τεχνική παραγοντοποίησης πίνακα. Η αξιολόγηση έγινε σε δύο σύνολα δεδομένων το Last.fm και το Movielens. Τα αποτελέσματα έδειξαν ότι κανένας αλγόριθμος δεν μπόρεσε να κυριαρχήσει και στις 3 κατηγορίες.

Στο movielens dataset τις καλύτερες μετρήσεις στην ακρίβεια έδωσε ο αλγόριθμος SVD με 50 παράγοντες, στο diversity η προσέγγιση με βάση τη δημοφιλία και για το novelty τα καλύτερα αποτελέσματα έδωσε η user-based προσέγγιση των κοντινών γειτόνων αντίστοιχα [58].

Στο lastfm σύνολο δεδομένων τα καλύτερα αποτελέσματα σε ακρίβεια έδωσε μία τεχνική παραγοντοποίησης πίνακα με σταθμισμένα βάρη, σε diversity μια τεχνική user-based κοντινών γειτόνων επίσης με σταθμισμένα βάρη και σε novelty ο αλγόριθμος SVD με 150 παράγοντες. Τέλος μέσα από μία προσέγγιση Pareto συνδύασε τους διάφορους αλγορίθμους και παρουσίασε υβριδικές λύσεις που έχουν μετρήσεις όμοιες με τις καλύτερες επιδόσεις σε κάθε ξεχωριστή κατηγορία έχοντας παράλληλα βελτίωση στις άλλες δύο κατηγορίες. Με άλλα λόγια μία υβριδική λύση μπορεί να έχει την καλύτερη μέτρηση σε ακρίβεια που είχε ο svd με τους 50 παράγοντες και παράλληλα να έχει καλύτερα αποτελέσματα σε novelty και diversity σε σχέση με τον svd [58].

Άλλη μία ενδιαφέρουσα μελέτη έχουμε στο [2] όπου παρουσιάζονται πολλοί διαφορετικοί αλγόριθμοι αναδιάταξης μελετώντας πως η κάθε προσέγγιση επηρεάζει τις μετρικές στα σύνολα δεδομένων. Τα σύνολα που εξετάζονται και σε αυτή την περίπτωση είναι το movielens και το last.fm. Στο movielens σύνολο τα καλύτερα αποτελέσματα στις μετρικές ακρίβειας και diversity τα παρουσιάζει ο αλγόριθμος learning to rank, ενώ στις μετρικές novelty και coverage τα καλύτερα αποτελέσματα δίνει η item-based προσέγγιση των κοντινών γειτόνων.

Στο σύνολο δεδομένων last.fm τα καλύτερα αποτελέσματα σε ακρίβεια μας τα δίνει η item-based προσέγγιση κοντινών γειτόνων ενώ ο learning to rank αλγόριθμος παρουσιάζει τα χειρότερα σκορ. Στη μετρική diversity τα καλύτερα σκορ δίνει ο αλγόριθμος svd ενώ σε novelty και coverage κυριαρχεί και εδώ η item-based προσέγγιση κοντινών γειτόνων [2]. Σε αυτή την έρευνα σε αντίθεση με την πρώτη έρευνα [58] βλέπουμε ότι η προσέγγιση user-based κοντινών γειτόνων έχει χειρότερα σκορ από την item-based προσέγγιση των κοντινών γειτόνων.

Στο [84] γίνεται αξιολόγηση σε μετρικές ακρίβειας, diversity, novelty και catalog coverage. Αναπτύσσονται τρεις διαφορετικές προσεγγίσεις οι οποίες είναι προσέγγιση με βάση τις αξιολογήσεις, προσέγγιση με βάση το περιεχόμενο και προσέγγιση με βάση το συνεργατικό φιλτράρισμα και τις κοινωνικές σχέσεις μεταξύ των χρηστών. Οι μετρήσεις γίνονται σε τρία ξεχωριστά σύνολα δεδομένων το movielens, το last.fm και το delicious. Από τα αποτελέσματα βλέπουμε ότι και σε αυτό το πείραμα δεν υπάρχει κάποιος αλγόριθμος που να δίνει τα καλύτερα σκορ σε όλες τις μετρικές ταυτόχρονα.

Όσον αφορά την ακρίβεια στα σύνολα δεδομένων last.fm και delicious τα καλύτερα αποτελέσματα δίνει η προσέγγιση με βάση τις κοινωνικές σχέσεις μεταξύ των χρηστών ενώ για το movielens σύνολο δεδομένων το καλύτερο σκορ το δίνει η προσέγγιση με βάση το περιεχόμενο. Οι αλγόριθμοι συνεργατικού φιλτραρίσματος δίνουν καλύτερα σκορ για το diversity και στα τρία σύνολα δεδομένων. Στο σύνολο movielens το υψηλότερο σκορ novelty μας το δίνει η προσέγγιση συνεργατικού φιλτραρίσματος και στα υπόλοιπα δύο η προσέγγιση με βάση το περιεχόμενο. Στις προσεγγίσεις με βάση το περιεχόμενο επιτυγχάνεται επίσης το υψηλότερο σκορ coverage και για τα τρία σύνολα δεδομένων [84].

Μέσα από τις παραπάνω έρευνες και συγκρίνοντας με τα αποτελέσματα μετρήσεων στη δική μας έρευνα γίνεται εύκολα αντιληπτό ότι δεν υπάρχει κάποιος αλγόριθμος που να μπορεί να παρουσιάσει ταυτόχρονα τις καλύτερες επιδόσεις σε όλες τις μετρικές. Επίσης είναι αξιοσημείωτο το γεγονός ότι στα ίδια σύνολα δεδομένων μεταξύ των ερευνών μπορεί στη μία έρευνα να δώσει το καλύτερο αποτέλεσμα σε μία μετρική ένας αλγόριθμος και στην άλλη έρευνα ένας διαφορετικός αλγόριθμος.

3.5 Feedback Χρηστών στις Νέες Μετρικές

Οι μετρικές που συζητήθηκαν στις προηγούμενες ενότητες επηρεάζουν τόσο τη χρησιμότητα ενός συστήματος όσο και το ποια θα είναι η οπτική του χρήστη για αυτό. Μπορεί στη θεωρία ένα μοντέλο που δημιουργήθηκε να αποδίδει καλά αποτελέσματα τόσο σε επίπεδο novelty όσο και σε επίπεδο diversity αλλά υπάρχει τρόπος να επιβεβαιώσουμε και στην πράξη ότι αυτές οι προτάσεις φαίνονται και στους χρήστες καινοτόμες ή ποικίλες; Με βάση αυτό το ερώτημα θα προσπαθήσουμε να κάνουμε μία σύνοψη παρουσιάζοντας κάποιες έρευνες που έχουν χρησιμοποιηθεί προς την κατεύθυνση απόκτησης feedback από τους χρήστες.

Μέχρι σήμερα δεν έχουν διεξαχθεί πολλές έρευνες ως προς αυτό το κομμάτι. Αυτό μπορεί να οφείλεται σε αρκετούς παράγοντες. Η διεξαγωγή μίας τέτοιας έρευνας χρειάζεται να είναι τόσο καλά σχεδιασμένη έτσι ώστε να μπορούν να είναι κατανοητές από τα άτομα που συμμετέχουν στην έρευνα οι έννοιες που εξετάζονται. Επιπλέον μία τέτοια έρευνα θα απαιτούσε αρκετά μεγάλο αριθμό ατόμων και θα έπρεπε να έχουμε εξασφαλίσει ότι η πλειοψηφία αυτών έχουν κατανοήσει με σωστό τρόπο την έννοια του novelty και του diversity.

Μία ακόμα δύσκολη παράμετρος είναι το αντικείμενο που πραγματεύεται η έρευνα. Για παράδειγμα αν ένα σύστημα προτείνει μία λίστα με ταινίες χρειάζεται αρκετά μεγάλο διάστημα ώστε ο χρήστης να δει κάποιες εξ αυτών και στη συνέχεια να κρίνει αν όντως ικανοποιούσαν το αίτημα του για novelty και diversity.

Μία ενδιαφέρουσα έρευνα για το πως οι χρήστες αξιολογούν τα συστήματα συστάσεων παρουσιάζεται στο [75]. Οι συμμετέχοντες μπορούσαν να χρησιμοποιήσουν ένα σύστημα συστάσεων όσες φορές θέλανε κατά τη διάρκεια της εικοσαήμερης έρευνας. Μετά από κάθε χρήση συμπληρώνανε ένα ερωτηματολόγιο 47 ερωτήσεων σχετικών με την εμπειρία τους. Σε κάθε νέο session άλλαζε ο αλγόριθμος που χρησιμοποιούσαν χωρίς ωστόσο να γνωρίζουν ποιον αλγόριθμο χρησιμοποιούν κάθε φορά. Ο άξονας της έρευνας περιστρέφεται γύρω από την επιλογή δυσκολίας, την αντιληπτή ακρίβεια του συστήματος, το αντιληπτό diversity, την απόδοση του συστήματος και την ικανοποίηση από τις επιλογές.

Από τα αποτελέσματα της έρευνας διαπιστώνεται ότι το αντιληπτό diversity από τη μεριά των χρηστών είναι διαφορετικό για τη χρήση κάθε αλγορίθμου. Μάλιστα οι αλγόριθμοι k-NN που χρησιμοποιήθηκαν χωρίς να επικεντρώνονται στο diversity θεωρήθηκαν από τους χρήστες σαν πιο ποικίλοι από τους αντίστοιχους αλγορίθμους αναδιάταξης που είχαν επικεντρωθεί στην αύξηση του diversity. Ένα ακόμα συμπέρασμα είχε να κάνει με το γεγονός πως όταν οι χρήστες αντιλαμβάνονταν diverse ένα αντικείμενο τους φαινόταν σαν πιο ακριβές και λιγότερο δύσκολο να το διαλέξουν κάτι που τους έκανε να νιώθουν περισσότερο χαρούμενοι με τις προτάσεις του συστήματος.

Μία παρόμοια έρευνα σε εμπορικά σαιτ προϊόντων με παρόμοιους άξονες μέτρησης της ποιότητας μπορούμε να βρούμε στο [76]. Στη συγκεκριμένη μελέτη αναπτύσσεται μία μέθοδος που ονομάζεται ResQue (Recommender systems Quality of user experience) και προσπαθεί να μετρήσει την ποιότητα που παρέχουν οι προτάσεις του συστήματος αξιολογώντας τη χρησιμότητα του συστήματος, πόσο χρήσιμο φαίνεται στο χρήστη και πως το novelty, το diversity και άλλες μετρικές επηρεάζουν την συμπεριφορά του χρήστη και το αν τελικά θα αγοράσουν ένα προϊόν.

Η έρευνα περιείχε 32 ερωτήσεις που χωρίζονταν σε 32 κατηγορίες. Στα ευρήματα της φαίνεται ότι το novelty έχει μεγαλύτερη αλληλεπίδραση και χρησιμότητα για το χρήστη από το diversity. Αυτό φυσικά μπορεί να οφείλεται σε πολλούς παράγοντες όπως η διαφορετικότητα των αντικειμένων που έγινε το πείραμα αλλά και του τρόπου διατύπωσης των ερωτήσεων.

Σε άλλες έρευνες όπως στο [77] μετρήθηκε ο συσχετισμός μεταξύ novelty και χρησιμότητας σε ένα σύστημα συστάσεων που πρότεινε τραγούδια στους χρήστες. Οι χρήστες μπορούσαν να αντιληφθούν τους αλγορίθμους που πρότειναν πιο novel τραγούδια, ωστόσο τα συγκεκριμένα τραγούδια πήρανε χαμηλότερες βαθμολογίες από τα λιγότερο novelty που προτάθηκαν στους χρήστες.

Σε παρόμοια ευρήματα κατέληξε η έρευνα στο [78] όπου αξιολογήθηκαν οι αλγόριθμοι των συστημάτων συστάσεων πάνω στο diversity που πρότειναν ταινίες. Και εδώ οι χρήστες αντιλήφθηκαν τους αλγορίθμους που πρότειναν περισσότερο diverse αντικείμενα. Για τους αλγορίθμους που διατήρησαν χαμηλό έως μέτριο επίπεδο diversity οι χρήστες φάνηκε να μένουν περισσότερο ικανοποιημένοι, ενώ σε αυτούς με υψηλότερο diversity δεν παρατηρήθηκε κάποια αύξηση της ικανοποίησης.

4 Μεθοδολογία

4.1 Αλγόριθμος Αναδιάταξης

Στο κεφάλαιο 4 περιγράψαμε διάφορες τεχνικές αναδιάταξης που έχουν χρησιμοποιηθεί, προκειμένου ένα σύστημα συστάσεων να κάνει προτάσεις όχι μόνο σχετικές με τις προτιμήσεις του χρήστη με βάση το accuracy αλλά να λαμβάνονται υπόψιν και άλλες μετρικές όπως το diversity ή το novelty.

Για το πείραμα μας θα χρησιμοποιήσουμε τον αλγόριθμο αναδιάταξης όπως ορίζεται στο [2] καθώς μας επιτρέπει αφενώς μεν να τρέξουμε τους υπάρχοντες state of the art αλγορίθμους παίρνοντας αρχικά κάποιες προτάσεις για κάθε χρήστη αφετέρου να ελέγξουμε μέσω της παραμέτρου α την συσχέτιση μεταξύ accuracy και diversity. Ο greedy αλγόριθμος είναι ο εξής [2]:

Algorithm 1 Greedy Reranking Algorithm

Data: N ; a set of candidate items C , $|C| > N$

Result: result list R , $|R| = N$

$R \leftarrow []$;

while $|R| < N$ **do**

$i \leftarrow \operatorname{argmax}_{i \in C} f_{\text{obj}}(i, R)$;

$R \leftarrow R ++ [i]$;

$C \leftarrow C \setminus \{i\}$;

end while

$R \leftarrow$ result list R ;

με $++$ να συμβολίζει την ένωση της λίστας και \setminus τη διαφορά συνόλου.

Ο reranking αλγόριθμος δουλεύει με τον ακόλουθο τρόπο: Αφού τρέξουμε τους αλγορίθμους συνεργατικού φιλτραρίσματος παίρνουμε μία λίστα αντικειμένων συνόλου C τα οποία είναι σχετικά με τις προτιμήσεις του χρήστη όπως προκύπτει από την ομοιότητα μεταξύ αντικειμένων ή χρηστών. Ύστερα μετράμε για κάθε ταινία στην προτεινόμενη λίστα το diversity (τύπος 19) και το relevance score (τύπος 18).

Στη συνέχεια συνδυάζουμε τα δύο score και μέσω της αντικειμενικής συνάρτησης f_{obj} ξεκινάμε την αναδιάταξη. Σε κάθε τρέξιμο του αλγορίθμου το αντικείμενο στο σύνολο C που μεγιστοποιεί την αντικειμενική συνάρτηση μεταφέρεται στη λίστα αποτελεσμάτων R και αφαιρείται από το σύνολο C . Ο αλγόριθμος επαναλαμβάνεται όσο η λίστα R είναι μικρότερη σε πλήθος από N , με N ένας αριθμός μικρότερος του συνόλου C .

4.2 Diversity Τεχνική Αναδιάταξης

Αντικειμενική Συνάρτηση

Η αντικειμενική συνάρτηση f_{obj} που χρησιμοποιούμε στην αναδιάταξη του αλγορίθμου βασίζεται στο diversity και είναι η παρακάτω όπως ορίζεται στο [2]:

$$f_{obj}(i, R) = a \cdot rel(i) + (1 - a) \cdot obj_{div}(i, R) \quad (17)$$

Με την παραπάνω συνάρτηση μπορούμε να συσχετίσουμε την ακρίβεια με το diversity. Με το $rel(i)$ (τύπος 18) εκφράζεται η προβλεπόμενη σχετικότητα του αντικειμένου i και με το $obj_{div}(i, R)$ (τύπος 19) θα μετρήσουμε το σκορ για το diversity ανά ταινία i στη λίστα αποτελεσμάτων R .

Relevance

Για να υπολογίσουμε το relevance θα χρησιμοποιήσουμε τον παρακάτω τύπο:

$$rel(i) = \mu_u - \hat{r}_{ui} \quad (18)$$

με μ_u να είναι η μέση τιμή βαθμολογίας του χρήστη u για τις ταινίες του train set και \hat{r}_{ui} η τιμή της προβλεπόμενης βαθμολογίας του χρήστη u για το αντικείμενο i .

Diversity Reranking

Για την αναδιάταξη με βάση το diversity θα χρησιμοποιήσουμε την μέση απόσταση (τύπος 20) μεταξύ των αντικειμένων στη λίστα R κάτι που έχει χρησιμοποιηθεί και στις μελέτες [2] [52]:

$$obj_{diversity}(i, R) = \frac{\sum_{j \in R} dist(i, j)}{|R|} \quad (19)$$

Η μέτρηση της απόστασης θα γίνει με βάση τις αξιολογήσεις των χρηστών που έχουν βαθμολογήσει και τις δύο ταινίες i και j , χρησιμοποιώντας την απόσταση συνημιτόνου.

Έτσι έχουμε [2]:

$$dist(i, j) = 1 - \frac{\sum_{u \in U} (r_{ui} - \mu_i) \cdot (r_{uj} - \mu_j)}{\sqrt{\sum_{u \in U} (r_{ui} - \mu_i)^2} \sqrt{\sum_{u \in U} (r_{uj} - \mu_j)^2}} \quad (20)$$

με μ_i και μ_j να είναι οι μέσες τιμές για τα αντικείμενα i και j , r_{ui} η βαθμολογία του αντικειμένου i από το χρήστη u και r_{uj} η βαθμολογία του αντικειμένου j από το χρήστη u για κάθε χρήστη u που έχει βαθμολογήσει και τις δύο ταινίες i και j .

4.3 Ακολουθούμενη Προσέγγιση

Για να μελετήσουμε τις συσχετίσεις ανάμεσα στις μετρικές που ερευνήσαμε στην εργασία μας χρησιμοποιήσαμε αλγόριθμους συνεργατικού φιλτραρίσματος τόσο memory-based όσο και model-based. Οι αλγόριθμοι αυτοί αναλύθηκαν με περισσότερες λεπτομέρειες στο κεφάλαιο 2. Η παρουσίαση των αποτελεσμάτων θα γίνει για τους αλγόριθμους kNN with Means με user - based προσέγγιση, kNN with Means με item-based προσέγγιση, SVD και Slope One. Οι υπόλοιποι αλγόριθμοι που περιγράφονται στο κεφάλαιο 3 έδωσαν πολύ κοντινά αποτελέσματα με τους υπόλοιπους τέσσερις και έτσι θεωρήσαμε ότι δεν θα προσδώσει καμία επιπλέον αξία στο πείραμα μας η παρουσίαση τους.

Για να τρέξουμε αυτούς τους αλγόριθμους χρησιμοποιήσαμε τη βιβλιοθήκη surprise της python [9]. Η surprise είναι σχεδιασμένη για χτίσιμο και ανάλυση συστημάτων συστάσεων που προέρχονται από σύνολα δεδομένων με explicit data κάτι το οποίο την καθιστά μία από τις ιδανικότερες βιβλιοθήκες που μπορούν να χρησιμοποιηθούν για το movielens σύνολο δεδομένων.

Ένα ακόμα πλεονέκτημα που προσφέρει είναι πως ο χρήστης μπορεί να πειραματιστεί αλλάζοντας διάφορες παραμέτρους για τα καλύτερα δυνατά αποτελέσματα. Για τον υπολογισμό των μετρικών μας χτίσαμε από το μηδέν κάποιους αλγόριθμους χρησιμοποιώντας τα πλεονεκτήματα που προσφέρουν κάποιες βιβλιοθήκες της python όπως η numpy και η pandas.

Για κάθε αλγόριθμο η παραγόμενη λίστα προτάσεων ήταν 40 ταινίες ($C=40$). Στη συνέχεια για κάθε λίστα βασιζόμενοι στην αναδιάταξη με βάση το diversity τρέξαμε το reranking αλγόριθμο και κρατήσαμε τις 30 ($N=30$) κορυφαίες προτάσεις ανά χρήστη με βάση το συνδυαστικό σκορ μεταξύ relevance και diversity και τις εκχωρήσαμε στην λίστα R.

Για την αντικειμενική συνάρτηση ορίσαμε ως παράμετρο $a = 0.7$ γιατί θέλαμε να δώσουμε ελαφρώς περισσότερη έμφαση στο diversity έναντι του accuracy. Ο λόγος που επιλέξαμε τα νούμερα $C=40$ και $N=30$ ήταν γιατί θεωρήσαμε ότι είναι ένα αρκετά καλό δείγμα για να έχουμε πληθώρα προτάσεων ανάμεσα στους υποψηφίους με δεδομένο ότι οι πόροι μας ήταν περιορισμένοι αλλά και παράλληλα θεωρήσαμε ότι το όριο των 30 προτάσεων είναι αρκετά καλό ώστε να διατηρήσουμε το ενδιαφέρον του χρήστη και να μη του προτείνουμε ταινίες που πιθανότατα δε θα του τραβήξουν την προσοχή.

Μετρήσεις Απόδοσης

Για να αξιολογήσουμε τα αποτελέσματά μας χρησιμοποιούμε 5 μετρικές αξιολογήσεις. Προτού προχωρήσουμε στην παρουσίαση των τύπων θα θέλαμε να επισημάνουμε ότι οι παρακάτω μετρικές δεν αποτελούν τεχνικές αναδιάταξης που χρησιμοποιήθηκαν για να μας δώσουν προτάσεις, αλλά μετρικές για να αξιολογήσουμε τις λίστες προτάσεων που παράγααν οι αρχικοί baseline αλγόριθμοι και οι αλγόριθμοι μετά την εφαρμογή της diversity τεχνικής αναδιάταξης.

RMSE Score

Μετρήσαμε το RMSE βασιζόμενοι στον ακόλουθο τύπο όπως ορίζεται στο [40]:

$$RMSE = \sqrt{\frac{\sum_{i \in R} (r_{ui} - \hat{r}_{ui})^2}{|R|}} \quad (21)$$

με r_{ui} να είναι η πραγματική βαθμολογία του αντικειμένου i από το χρήστη u , \hat{r}_{ui} η προβλεπόμενη βαθμολογία του αντικειμένου i από το χρήστη u και R το συνολικό πλήθος αντικειμένων στη λίστα προτάσεων.

F Measure Score

Για τον υπολογισμό του F-measure βασιστήκαμε στον παρακάτω τύπο [85]:

$$F_{measure} = 2 \cdot \frac{recall \cdot precision}{precision + recall} = \frac{2 \cdot tp}{2 \cdot tp + fp + fn} \quad (22)$$

με true positive να ορίζουμε όσες ταινίες βρίσκονται στην λίστα προτάσεων προς το χρήστη και έχουν πραγματική βαθμολογία περισσότερο από 3.5 αστέρια, false positive όσες ταινίες έχουν πραγματική βαθμολογία μικρότερη ή ίση από 3.5 αστέρια αλλά βρίσκονται στη λίστα προτάσεων και false negative όσες ταινίες βρίσκονται στο test set και έχουν πραγματική βαθμολογία μεγαλύτερη από 3.5 αστέρια αλλά δεν βρίσκονται στη λίστα προτάσεων.

Diversity Score

Για το diversity χρησιμοποιήσαμε τον ακόλουθο τύπο [50]:

$$Diversity(R) = \frac{\sum_{i \in R} \sum_{j \in R \setminus \{i\}} dist(i, j)}{|R|(|R| - 1)} \quad (23)$$

με την απόσταση να υπολογίζεται χρησιμοποιώντας τις αξιολογήσεις των χρηστών όπως στον τύπο απόστασης που θέσαμε στην υποενότητα 5.1 (τύπος 20).

Novelty

Για το novelty χρησιμοποιήσαμε τον ακόλουθο τύπο [66]:

$$Novelty(R) = \frac{\sum_{i \in R} -\log_2 p(i)}{nov_{max} \cdot |R|} \quad (24)$$

με $p(i)$ να είναι η αναλογία των χρηστών που έδωσαν αξιολόγηση σε ένα αντικείμενο i σε σχέση με τους συνολικούς χρήστες του συνόλου δεδομένων και $nov_{max} = -\log_2 \frac{1}{U}$ να είναι η μέγιστη δυνατή τιμή του novelty, την οποία χρησιμοποιούμε για να κανονικοποιήσουμε το σκορ κάθε αντικειμένου στο $[0,1]$ με U να είναι ο συνολικός αριθμός όλων των χρηστών του συνόλου δεδομένων μας.

Για όλα τα παραπάνω υπολογίσαμε το μέσο όρο σε όλους τους χρήστες του test συνόλου.

Catalog Coverage

Για το catalog coverage μετράμε όλες τις μοναδικές ταινίες που εμφανίζονται στις N - κορυφαίες προτάσεις κάθε λίστας και τις διαιρούμε με τον συνολικό αριθμό ταινιών του test set παίρνοντας έτσι μία τιμή ανάμεσα στο $(0,1]$ [2].

$$\text{Prediction Coverage} = \frac{l_p}{l} \quad (25)$$

με l_p να είναι το σύνολο των μοναδικών ταινιών στις λίστες προτάσεων και l ο συνολικός αριθμός ταινιών του test set.

Ποσοστό Κοινών Ταινιών στο Top10

Για το ποσοστό των κοινών ταινιών στο Top10 κάθε χρήστη συγκρίναμε τις 10 κορυφαίες προτάσεις που πρότεινε ο baseline αλγόριθμος με τις 10 κορυφαίες προτάσεις όπως προέκυψαν μετά την αναδιάταξη με την τεχνική του diversity και βγάλαμε το μέσο όρο των κοινών ταινιών για όλους τους χρήστες του test συνόλου δεδομένων.

$$\text{Common Movies Percentage} = \frac{|K|}{10} \quad (26)$$

με K να εκφράζει το κοινό πλήθος κοινών ταινιών ανάμεσα στο top10 της αρχικής και της λίστας με diversity τεχνική αναδιάταξης.

5 Πειραματική Διαδικασία

Σύνολο Δεδομένων: Το σύνολο δεδομένων μας είναι το Movielens dataset [80]. Το αρχικό αρχείο με την ονομασία ratings περιέχει 20 εκατομμύρια αξιολογήσεις χρηστών σε μία κλίμακα 5 αστεριών που ο χρήστης μπορεί να βαθμολογεί ανά μισό από το 0.5 έως και το 5] και 27278 ταινίες. Κάθε χρήστης εκπροσωπείται από ένα `userId` χωρίς να παρέχεται κάποια άλλη πληροφορία για το χρήστη.

Από το αρχικό dataset κρατήσαμε τις αξιολογήσεις που είχαν τιμή τουλάχιστον 0.5 και όσους χρήστες είχαν βαθμολογήσει τουλάχιστον 700 ταινίες. Έτσι το τελικό dataset πάνω στο οποίο δουλέψαμε είχε 4.535.739 εγγραφές με 3989 μοναδικούς χρήστες και 25652 μοναδικές ταινίες. Θεωρήσαμε ότι είναι ένα αρκετά μεγάλο και επαρκές δείγμα για να τρέξουμε τα πειράματά μας και να λάβουμε αξιόπιστα αποτελέσματα διατηρώντας παράλληλα το performance των αλγορίθμων. Το σύνολο δεδομένων μας χωρίστηκε σε train set και test set με αναλογία 85 - 15.

Παράμετροι Αλγορίθμων: Όπως προείπαμε η τεχνική αναδιάταξης με βάση το diversity αξιολογήθηκε πάνω σε τέσσερις αλγορίθμους συνεργατικού φιλτραρίσματος τον αλγόριθμο παραγοντοποίησης SVD, τον αλγόριθμο κοντινών γειτόνων user-based, τον αλγόριθμο κοντινών γειτόνων item-based και τον model-based αλγόριθμο Slope One. Γίνανε διάφορες δοκιμές για τις παραμέτρους του κάθε αλγορίθμου (εκτός από τον Slope One που δεν έχει παραμέτρους) και αποφασίστηκε να κρατηθούν όσες παράμετροι δίνουν τα καλύτερα αποτελέσματα στην ακρίβεια μια πρακτική που υιοθετείται και σε άλλες έρευνες της βιβλιογραφίας όπως στο [2]. Έτσι έχουμε τις ακόλουθες παραμέτρους:

- SVD: αριθμός παραγόντων = 100, αριθμός εκμάθησης = 0.005 .
- kNN user-based αριθμός κοντινών γειτόνων $k = 60$.
- kNN item -based αριθμός κοντινών γειτόνων $k = 60$.

5.1 Αποτελέσματα

Στο πείραμα μας συγκρίναμε πως μεταβάλλονται οι μετρικές rmse, fmeasure, diversity, novelty και catalog coverage μεταξύ των αρχικών προτάσεων των τεσσάρων αλγορίθμων και των τελικών προτάσεων που προέκυψαν μετά την τεχνική αναδιάταξης με βάση το diversity. Έχει γίνει κανονικοποίηση των μετρικών στην κλίμακα $[0,1]$ για να μπορούμε να συγκρίνουμε τα αποτελέσματα στην ίδια κλίμακα.

Σύγκριση Μετρικών Μεταξύ Αλγορίθμων

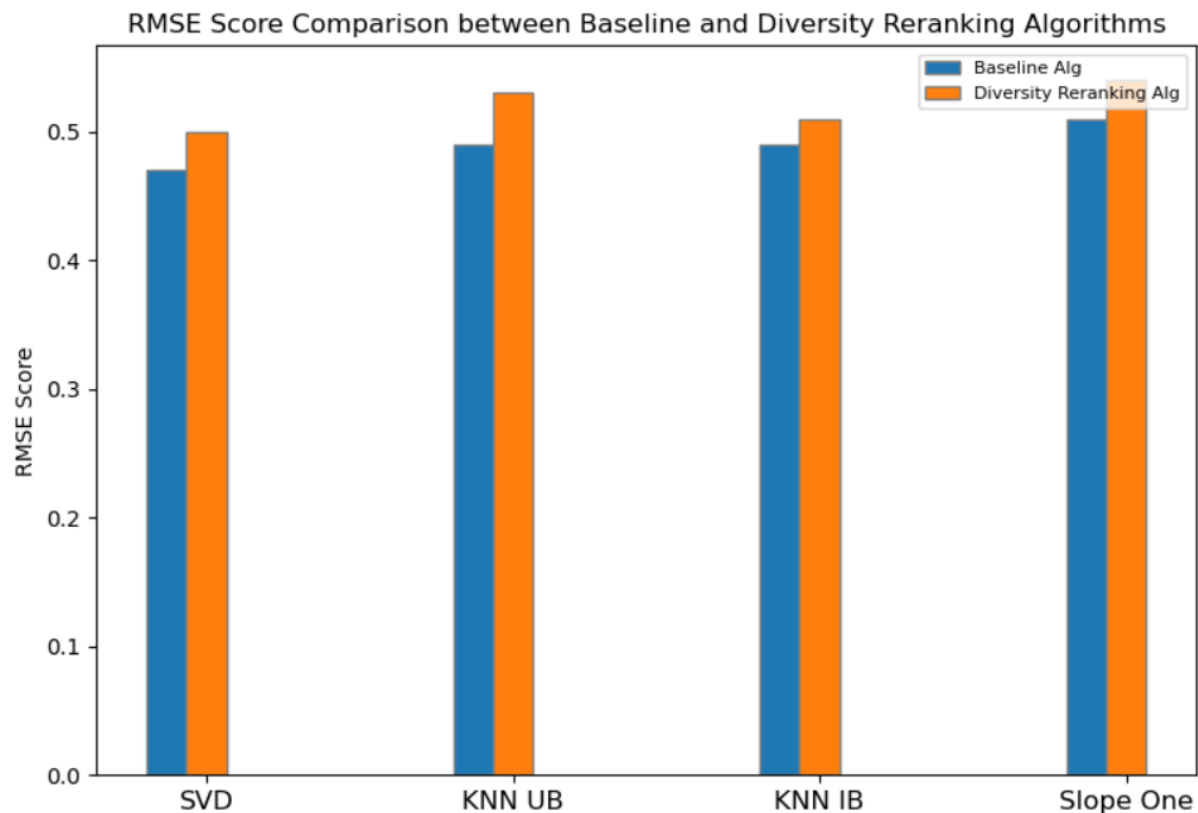


Figure 6: Σύγκριση RMSE σκορ πριν και μετά την αναδιάταξη μεταξύ των αλγορίθμων

Αρχικά βλέπουμε το γράφημα για το rmse (εικόνα 6). Για τις μετρήσεις των προτάσεων της αρχικής λίστας παρατηρούμε ότι το καλύτερο rmse σκορ (δηλαδή αυτό με τη μικρότερη τιμή) μας το δίνει ο αλγόριθμος SVD με 0.47. Οι knn-UB και knn-IB έρχονται δεύτεροι με παρόμοιο σκορ 0.49 ενώ ακολουθεί ο slope one με σκορ 0.51. Μετά την αναδιάταξη των προτάσεων με την τεχνική diversity παρατηρούμε ότι και για τους 4 αλγορίθμους η ακρίβεια μειώνεται όπως θα αναμέναμε. Ο αλγόριθμος SVD έχει το καλύτερο σκορ με 0.5 και ακολουθείται από τον knn-IB με σκορ 0.51 ενώ οι αλγόριθμοι knn-UB και Slope one έχουν 0.53 και 0.54 αντίστοιχα.

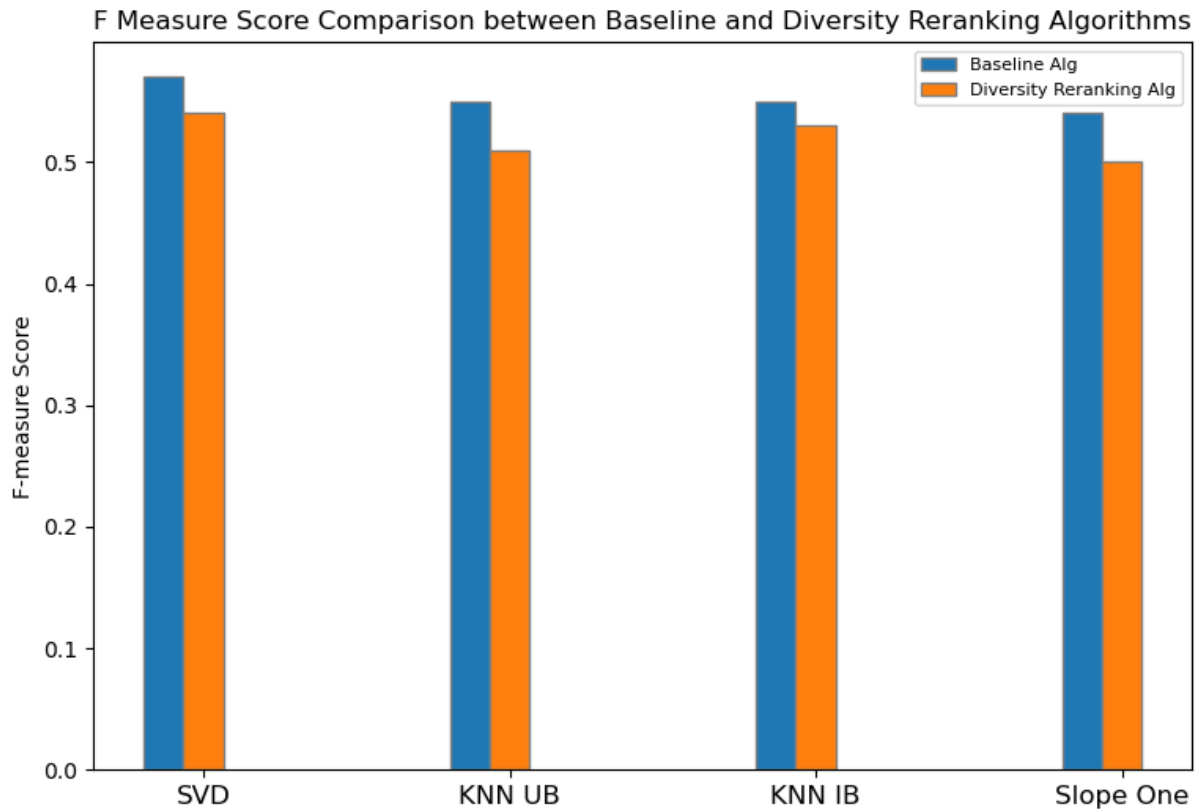


Figure 7: Σύγκριση F-measure σκορ πριν και μετά την αναδιάταξη μεταξύ των αλγορίθμων

Παρόμοια αποτελέσματα έχουμε για το F-measure ή F-score (εικόνα 7). Το μεγαλύτερο σκορ μας το δίνει ο SVD με 0.57 και ακολουθούν οι knn-IB, knn-UB με 0.55 ενώ ο Slope One ακολουθεί τελευταίος με 0.54. Και σε αυτή τη μετρική μετά την αναδιάταξη των προτάσεων έχουμε μείωση των σκορ για τους αλγορίθμους μας. Το μεγαλύτερο σκορ έχουν οι αλγόριθμοι svd με 0.54, knn-IB με 0.53 ακολουθούμενοι από τον knn-UB με 0.51 και Slope One με 0.49.

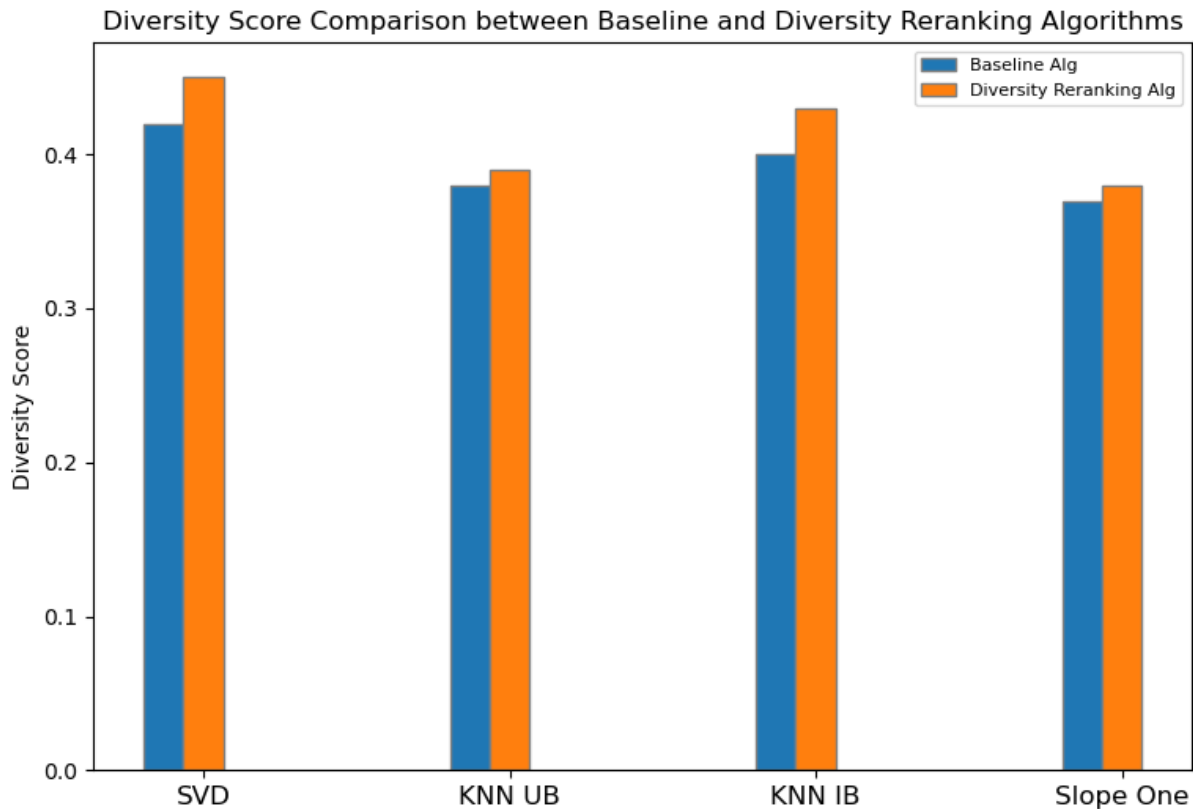


Figure 8: Σύγκριση Diversity σκορ πριν και μετά την αναδιάταξη μεταξύ των αλγορίθμων

Όσον αφορά το diversity (εικόνα 8) το μεγαλύτερο σκορ μας το έδωσε ο αλγόριθμος SVD με 0.42, ακολουθούμενος από τον knn IB με 0.4 και τρίτο τον knn UB με 0.38 ενώ ο Slope One αλγόριθμος έρχεται τελευταίος και σε αυτή την περίπτωση με 0.37. Μετά την αναδιάταξη αυξάνεται το diversity score και για τους τέσσερις αλγορίθμους. Το υψηλότερο σκορ μας το δίνει και πάλι ο SVD με 0.45, δεύτερος έρχεται ο knn-IB με 0.43 και οι αλγόριθμοι knn UB, Slope One με πολύ μικρή άνοδο σε σχέση με την αρχική τιμή με 0.39 και 0.38 αντίστοιχα.

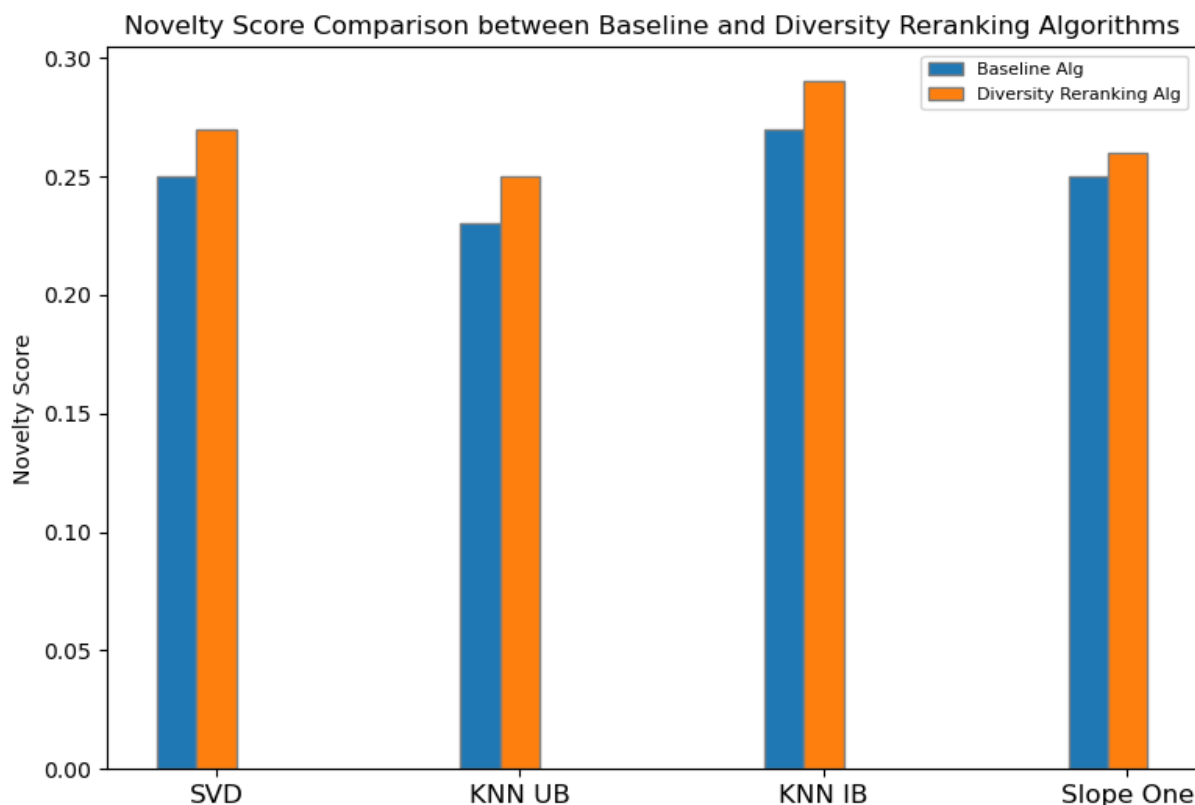


Figure 9: Σύγκριση Novelty σκορ πριν και μετά την αναδιάταξη μεταξύ των αλγορίθμων

Το μεγαλύτερο σκορ στο novelty (εικόνα 9) μας το δίνει ο αλγόριθμος knn IB με σκορ 0.27 και ακολουθούν οι αλγόριθμοι SVD και Slope One με 0.25 αντίστοιχα, ενώ στην τελευταία θέση αυτή τη φορά έρχεται ο knn UB με 0.23. Και σε αυτή την περίπτωση η μετρική novelty αυξάνεται μετά την αναδιάταξη των προτάσεων. Το υψηλότερο σκορ παρουσιάζει και εδώ ο αλγόριθμος knn-IB με 0.29 και ακολουθεί ο SVD με 0.27, ο Slope one με 0.26 και ο knn-UB με 0.25.

Όσον αφορά το catalog coverage (εικόνα 10) στην αρχική λίστα προτάσεων μεγαλύτερη κάλυψη έχουν οι αλγόριθμοι SVD και knn IB με 0.23 και 0.22 αντίστοιχα με τους αλγορίθμους knn UB και Slope να έχουν το μικρότερο ποσοστό κάλυψης με 0.16 και 0.13. Μετά την αναδιάταξη παρατηρήθηκε μείωση και στις 4 μετρικές. Το μεγαλύτερο ποσοστό κάλυψης έχουν οι αλγόριθμοι SVD με 0.20, knn IB με 0.22 και ακολουθούν οι αλγόριθμοι knn UB και Slope One με 0.16 και 0.12 αντίστοιχα.

Τέλος παρατηρούμε στο γράφημα της εικόνας 11 τα ποσοστά των κοινών ταινιών που βρέθηκαν στη λίστα των 10 κορυφαίων προτάσεων για κάθε χρήστη μεταξύ των baseline αλγορίθμων και των αλγορίθμων με την diversity τεχνική. Το μεγαλύτερο ποσοστό έχει ο αλγόριθμος Slope One με 75%, ακολουθούμενος από τον knn-UB με 74%. Τα μικρότερα ποσοστά δίνουν οι αλγόριθμοι knn-IB με 71% και ο SVD με 70%.

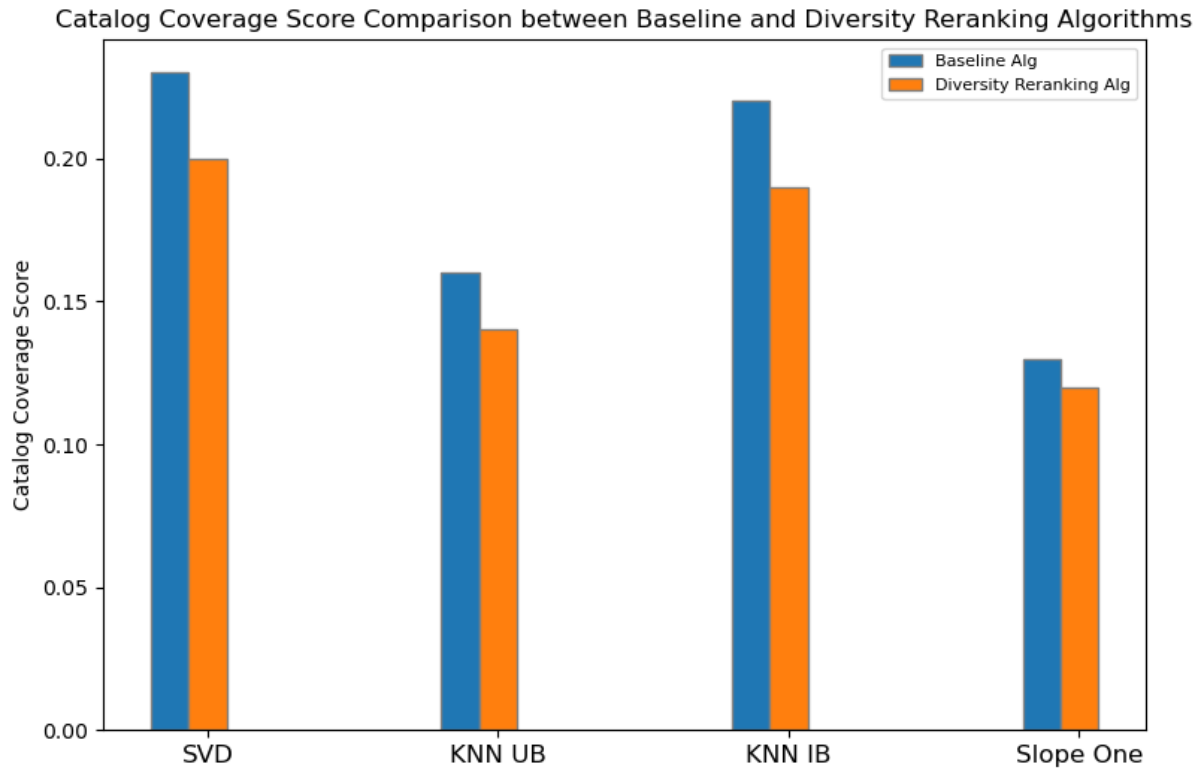


Figure 10: Σύγκριση Catalog Coverage σκορ πριν και μετά την αναδιάταξη των αλγορίθμων

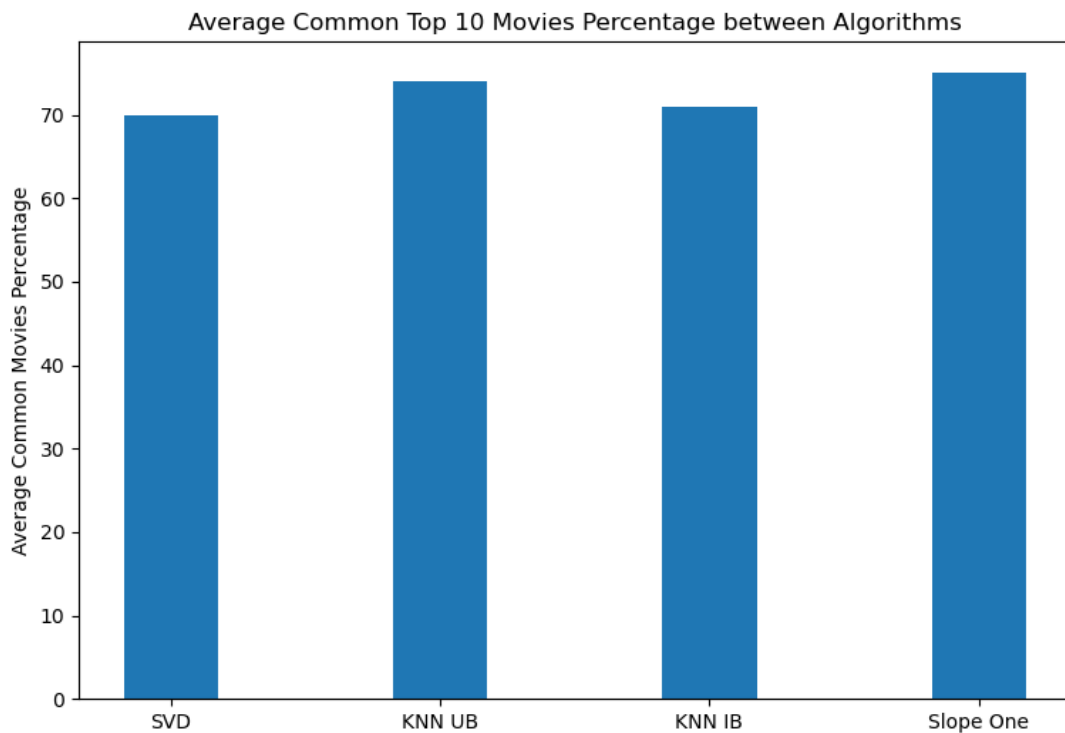


Figure 11: Σύγκριση ποσοστών top 10 κοινών ταινιών πριν και μετά την αναδιάταξη των αλγορίθμων

5.2 Σχολιασμός Αποτελεσμάτων

Από τα παραπάνω αποτελέσματα διαπιστώσαμε ότι τα δύο μέτρα που χρησιμοποιήσαμε για να μετρήσουμε την ακρίβεια μειώθηκαν μετά την αναδιάταξη των προτάσεων με βάση την τεχνική του diversity και ταυτόχρονα παρατηρήθηκε αύξηση στις μετρικές diversity και novelty. Υπάρχει δηλαδή μία τάση θετικής συσχέτισης ανάμεσα στο diversity και το novelty και αρνητικής συσχέτισης μεταξύ αυτών και της ακρίβειας. Αυτή η τάση θα μπορούσε να θεωρηθεί αναμενόμενη από τη στιγμή που στο πείραμα μας συμπεριλαμβάνουμε και προτάσεις αντικειμένων προσπαθώντας να δώσουμε και μία μικρή έμφαση στην ποικιλία των προτάσεων θυσιάζοντας ένα μικρό μέρος της ακρίβειας.

Μείωση παρατηρήθηκε και στη συνολική κάλυψη των προτεινόμενων ταινιών μετά την αναδιάταξη. Αυτό πιθανώς να οφείλεται στο ότι οι αλγόριθμοι επικεντρώνονται σε ένα συγκεκριμένο κομμάτι των long-tail αντικειμένων και τον μικρότερο αριθμό ταινιών που προτείνεται στη λίστα ανά χρήστη, όμως, χρειάζεται περαιτέρω ανάλυση προτού καταλήξουμε σε κάποιο συμπέρασμα.

Όσον αφορά την απόδοση των αλγορίθμων μας τα καλύτερα αποτελέσματα παρουσίασαν οι αλγόριθμοι SVD και knn IB που έδωσαν σε συνδυασμό τα υψηλότερα σκορ ανάμεσα σε ακρίβεια, diversity, novelty και catalog coverage. Πιο συγκεκριμένα μετά την αναδιάταξη του SVD με την τεχνική του diversity καταφέραμε να έχουμε αύξηση 7% στο diversity με παράλληλη μείωση 5,2% στο f-measure σκορ ενώ παράλληλα διατηρήθηκε το 70% των 10 κορυφαίων προτάσεων της αρχικής λίστας, στην τελική λίστα μετά την αναδιάταξη.

Μετά την αναδιάταξη του knn IB πετύχαμε αύξηση του diversity περίπου 7% και μείωση του f measure γύρω στο 3,5%, ενώ διατηρήθηκε το 71 % των 10 κορυφαίων προτάσεων της αρχικής λίστας, στην τελική λίστα μετά την αναδιάταξη. Μέτρια απόδοση είχαν οι αλγόριθμοι knn UB και Slope One που παρουσίασαν μια αύξηση της τάξης του 2,5% στο diversity με μείωση 7% στο f-measure σκορ.

6 Επίλογος και Μελλοντική Δουλειά

Σε αυτή την εργασία αρχικά αναφερθήκαμε στο συνεργατικό φιλτράρισμα και τη λογική πάνω στην οποία είναι σχεδιασμένο για να παράγει προτάσεις για τους χρήστες σε ένα σύστημα συστάσεων. Στη συνέχεια μιλήσαμε για τους state of the art αλγορίθμους του συνεργατικού φιλτραρίσματος τους οποίους χρησιμοποιήσαμε για να τρέξουμε το πείραμα μας. Περιγράψαμε τρεις από τις σπουδαιότερες μετρικές εκτός της ακρίβειας δηλαδή τις diversity, novelty, coverage που τα τελευταία χρόνια έχουν γίνει αντικείμενο μελέτης πολλών ερευνητών. Για κάθε μία από αυτές δώσαμε διάφορους ορισμούς από τη βιβλιογραφία και μιλήσαμε για διάφορες τεχνικές που έχουν χρησιμοποιηθεί για τη βελτιστοποίηση κάθε μίας εξ αυτών.

Επιπλέον τρέξαμε ένα πείραμα χρησιμοποιώντας μία τεχνική οπτικοποίησης του diversity και τον αλγόριθμο αναδιάταξης που περιγράφηκε στο κεφάλαιο 4. Αξιολογήσαμε τις αποδόσεις που είχαν οι μετρικές rmse, fmeasure, diversity, novelty και coverage στο σύνολο δεδομένων μας movielens. Ο κύριος σκοπός της εργασίας μας ήταν να συγκρίνουμε πως επηρεάζονται αυτές οι μετρικές μετά την αναδιάταξη των αρχικών προτάσεων. Αν και το πείραμα μας ήταν σε περιορισμένο εύρος παρατηρήσαμε μία θετική συσχέτιση ανάμεσα στο diversity και το novelty και αρνητική συσχέτιση μεταξύ αυτών και του accuracy. Δηλαδή όπως ίσως θα αναμέναμε η αύξηση του diversity στις προτάσεις προς τους χρήστες είχε ως αποτέλεσμα να χάσουν οι προτάσεις σε ακρίβεια.

Σαν μελλοντική δουλειά θα μπορούσαμε να επικεντρωθούμε σε πολλές διαφορετικές κατευθύνσεις. Για παράδειγμα θα μπορούσαμε να εφαρμόσουμε παρόμοιες τεχνικές οπτικοποίησης και για άλλες μετρικές όπως το novelty ή το serdendipity και να συγκρίνουμε πως αλληλεπιδρά η κάθε μία σε ένα σύνολο δεδομένων. Τέλος θα μπορούσε η μελέτη μας να επικεντρωθεί περισσότερο και στην οπτικοποίηση των διάφορων αλγορίθμων που τρέχουμε για να πάρουμε αποτελέσματα και να κρίνουμε ανά περίπτωση πως οι αλλαγές σε μία παράμετρο βελτιστοποιούν το αποτέλεσμα που θέλουμε.

Αναφορές

- [1] Mcnee S., Riedl J and Konstan J. 2006. Accurate is not always good: How Accuracy Metrics have hurt Recommender Systems, Conference on Human Factors in Computing Systems, Quebec.
- [2] Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems, MARIUS KAMINSKAS and DEREK BRIDGE, Insight Centre for Data Analytics, University College Cork, Ireland
- [3] Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity, Mouzhi Ge , Carla Delgado-Battenfeld , Dietmar Jannach
- [4] Deshpande, M., Karypis, G. (2004) Item-based top-n recommendation algorithms. ACM Trans. Inf. Syst., 22 (1), 143-177/
- [5] Content-based Recommender System for Movie Website, Master Thesis at VionLabs, Mihhail Matskin
- [6] Bobadilla, J., Ortega, F., Hernando, A., Gutierrez, A. (2013). Recommender systems survey. Knowledge-Based Systems, 46, 109-132
- [7] RECOMMENDATION SYSTEM USING COLLABORATIVE FILTERING, Yunkyoung Lee, San Jose State University, Fall 2015
- [8] Hybrid Recommender Systems: Survey and Experiments, Robin Burke
- [9] Surprise, A Python scikit for recommender systems, Jekyll and Hyde
- [10] Facebook user interests exploration and recommendation based on Facebook Social Graph data analysis, Author: Jakub Krzemien
- [11] Introduction to Recommender Systems Handbook, Francesco Ricci, Lior Rokach and Bracha Shapira
- [12] Using Collaborative Filtering to Weave an Information Tapestry, David Goldberg, David Nichols, Brian Oki, and Douglas Terry, September 1992
- [13] GroupLens is a research lab in the Department of Computer Science and Engineering at the University of Minnesota, Twin Cities specializing in recommender systems, online communities, mobile and ubiquitous technologies, digital libraries, and local geographic information systems. <https://grouplens.org/about/what-is-grouplens/>
- [14] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl. GroupLens: an open architecture for collaborative filtering of netnews. Proceedings of the 1994 ACM conference on Computer supported cooperative work, pp. 175-186, (1994).
- [15] Greg Linden, Brent Smith, and Jeremy York, Amazon.com Amazon.com Recommendations: Item-to-Item Collaborative Filtering
- [16] M. R. McLaughlin, J. L. Herlocker. A Collaborative Filtering Algorithm and Evaluation Metric that Accurately Model the User Experience. SIGIR '04: Proceedings of

- the 27th annual international ACM SIGIR conference on Research and development in information retrieval, ACM Press, (2004).
- [17] Netflix Prize and SVD, Stephen Gower, April 18th 2014
 - [18] B. Sarwar, G. Karypis, J. Konstan, J. Riedl. Item-Based Collaborative Filtering Recommendation Algorithms. Proceedings of the 10th international conference on World Wide Web, pp. 285-295, (2001).
 - [19] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. Recommender Systems Handbook. 1st edition, 2010 Chapter 4.
 - [20] Recommender Systems The Textbook, Charu C. Aggarwal, IBM T.J. Watson Research Center Yorktown Heights, NY, USA
 - [21] Schelter, S., Boden, C., Schenck, M., Alexandrov, A., Markl, V. (2013). Distributed matrix factorization with mapreduce using a series of broadcast-joins. In Proceedings of the 7th ACM conference on Recommender systems (pp. 281–284). ACM.
 - [22] M. D. Ekstrand, J. T. Riedl, J. A. Konstan. Collaborative Filtering Recommender Systems. Human-Computer Interaction, Volume 4, Issue 2, (2010).
 - [23] J. B. Schafer, D. Frankowski, J. Herlocker, S. Sen. Collaborative Filtering Recommender Systems. Conference Paper - January 2007, Springer Berlin Heidelberg,
 - [24] N. Rubens, D. Kaplan, and M. Sugiyama. Active learning in recommender systems. Recommender Systems Handbook, Springer
 - [25] Park DH, Kim HK, Choi IY, Kim JK. (2012). A review and classification of recommender systems research. 2011 International Conference on Social Science and Humanity IPEDR vol.5 (2011) © (2011) IACSIT Press, Singapore
 - [26] On Reducing the Data Sparsity in Collaborative Filtering Recommender Systems, Xin Guan Thesis, University of Warwick, Department of Computer Science, April 2017
 - [27] PolyLens: A Recommender System for Groups of Users, Mark O'Connor, Dan Cosley, Joseph A. Konstan, and John Riedl, Department of Computer Science and Engineering, University of Minnesota
 - [28] Mining semantic data for solving first-rater and cold-start problems in recommender systems, María N. Moreno García, Saddy Segrera, Vivian F. López Batista, María Dolores Muñoz, Conference: 15th International Database Engineering and Applications Symposium (IDEAS 2011), September 21 - 27, 2011, Lisbon, Portugal
 - [29] R. Rafter. Evaluation and Conversation in Collaborative Filtering. PhD Thesis, University College Dublin, College of Engineering Mathematical and Physical Sciences, (2010)
 - [30] Selecting Appropriate Metrics for Evaluation of Recommender Systems, Bhupesh Rawat, Sanjay K Dwivedi - Babasaheb Bhimrao Ambedkar University, January 2019 International Journal of Information Technology and Computer Science

- [31] An Analysis of Memory Based Collaborative Filtering Recommender Systems with Improvement Proposals, Claudio Adrian Levinas, September 2014
- [32] KNN Model-Based Approach in Classification, Gongde Guo Fujian Normal University, Hui Wang Chinese Academy of Sciences, David A. Bell Queen’s University Belfast, Yaxin Bi Ulster University
- [33] H. Wang.: Nearest Neighbours without k: A Classification Formalism based on Probability, technical report, Faculty of Informatics, University of Ulster, N.Ireland, UK (2002)
- [34] C.M.Bishop.: Neural Networks for Pattern Recognition. Oxford University Press, UK (1995)
- [35] A Comprehensive Survey of Neighborhood-Based Recommendation Methods, Christian Desrosiers, École de Technologie Supérieure, George Karypis, University of Minnesota Twin Cities, Chapter January 2011
- [36] Amatriain Xavier, Alejandro Jaimes, Nuria Oliver (2011). Data Mining Methods for Recommender Systems. In Recommender Systems Handbook (pp. 39–71). Springer: US
- [37] Slope One Predictors for Online Rating-Based Collaborative Filtering, Daniel Lemire Anna Maclachlan† , 15 Sep (2008)
- [38] Slope One Recommendation Algorithm Based on User Clustering and Scoring Preferences, Yue Ting Song and Sheng Wu, 3rd International Conference on Mechatronics and Intelligent Robotics (ICMIR-2019)
- [39] Funk, S., “Netflix Update: Try This At Home”, <https://sifter.org/simon/journal/20061211.html>
- [40] Evaluating Collaborative Filtering Algorithms for Music Recommendations on Chinese Music Data, Yifan He, Senior Thesis 2021
- [41] Koren, Y., “Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model”, Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008.
- [42] P. Adamopoulos. Notes on Recommender Systems: A Survey of State-of-the-Art Algorithms, Beyond Rating Prediction Accuracy Approaches, and Business Value Perspectives. Leonard N. Stern School of Business, New York University, (2013).
- [43] R. Salakhutdinov and A. Mnih. Probabilistic Matrix Factorization. (2008), NIPS Proceedings.
- [44] Xiaoyuan, S., Taghi, M. K. (2009). A Survey of Collaborative Filtering Techniques. In Advances in Artificial Intelligence
- [45] Lu, J., Dianshuang, Wu, Mao, M., Wang, W., Zhang, G. (2015). Recommender system application developments: a survey in Decision Support Systems
- [46] Lathia, N. (2010). Evaluating Collaborative Filtering Over Time. PhD thesis. University of London, Department of Computer Science.

- [47] Bellogín A., Cantador I., Díez, F., Castells, P. Chavarriaga, E. (2012). An empirical comparison of social, collaborative filtering, and hybrid recommenders. *ACM Transactions on Intelligent Systems and Technology*.
- [48] Recommendation System Analysis and Evaluation, Minase Netsereab Tekleab, School of Applied Sciences of Near East University
- [49] Performance Comparison of Collaborative Filtering, Prediction Methods on Recommendation System Win Win Moea, Nang Aye Aye Htweb, Department of Computer Engineering and Information Technology, Mandalay Technological University, Mandalay, Myanmar
- [50] Barry Smyth and Paul McClave. 2001. Similarity vs. diversity. In *Proceedings of the 4th International Conference on Case-Based Reasoning*.
- [51] John Paul Kelly and Derek Bridge. 2006. Enhancing the diversity of conversational collaborative recommendations: A comparison. *Artificial Intelligence Review*
- [52] Saul Vargas and Pablo Castells. 2011. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the 5th ACM Conference on Recommender Systems*
- [53] John Paul Kelly and Derek Bridge. 2006. Enhancing the diversity of conversational collaborative recommendations: A comparison. *Artificial Intelligence Review*
- [54] Martijn C. Willemsen, Bart P. Knijnenburg, Mark P. Graus, Linda C. M. Velter-Bremmers, and Kai Fu. 2011. Using latent features diversification to reduce choice difficulty in recommendation lists
- [55] Charles L. A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Buttcher, and Ian MacKinnon. 2008. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- [56] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- [57] Andrea Barraza-Urbina, Benjamin Heitmann, Conor Hayes, and Angela Carrillo-Ramos. 2015. XPLODIV: An exploitation-exploration aware diversification approach for recommender systems. In *Proceedings of the 28th International Florida Artificial Intelligence Research Society Conference*.
- [58] Marco Tulio Ribeiro, Anisio Lacerda, Adriano Veloso, and Nivio Ziviani. 2012. Pareto-efficient hybridization for multi-objective recommender systems.
- [59] Lorraine McGinty and Barry Smyth. 2003. On the role of diversity in conversational recommender systems. In *Proceedings of the 5th International Conference on Case-Based Reasoning*.

- [60] Personalizing diversity versus accuracy in session-based recommender systems, Alireza Gharahighehi, Celine Vens, KU Leuven, Campus KULAK, Department of Public Health and Primary Care, Kortrijk, Belgium
- [61] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 1999. *Modern Information Retrieval*. ACM Press / AddisonWesley
- [62] Yi Zhang, Jamie Callan, and Thomas Minka. 2002. Novelty and redundancy detection in adaptive filtering. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- [63] Yan Yang and Jian Z. Li. 2005. Interest-based recommendation in digital library. *Journal of Computer Science*
- [64] The Definition of Novelty in Recommendation System Liang Zhang, *Journal of Engineering Science and Technology Review* 6 (3) (2013)
- [65] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating collaborative filtering recommender systems.
- [66] Oscar Celma. 2009. ‘Music Recommendation and Discovery in the Long Tail. Ph.D. Dissertation. Universitat Pompeu Fabram, Chapter 4
- [67] A Survey of Long-Tail Item Recommendation Methods, Jing Qin, 29 Nov 2021
- [68] X. Huang and F. Wu, “A novel topic-based framework for recommending long tail products,” *Computers and Industrial Engineering*, vol. 137, p. 106063, 2019
- [69] A. Luke, J. Johnson, and Y. Ng, “Recommending long-tail items using extended tripartite graphs,” in *2018 IEEE International Conference on Big Knowledge*, 2018
- [70] L. Hu, L. Cao, J. Cao, Z. Gu, G. Xu, and J. Wang, “Improving the quality of recommendations for users and items in the tail of distribution,” *ACM Transactions on Information Systems*, 2017
- [71] S. Wang, M. Gong, H. Li, and J. Yang, “Multi-objective optimization for long tail recommendation,” *Knowledge-Based Systems*, 2016
- [72] J. Pang, J. Guo, and W. Zhang, “Using multi-objective optimization to solve the long tail problem in recommender system,” in *Advances in Knowledge Discovery and Data Mining -23rd Pacific-Asia Conference, PAKDD 2019*, pp. 302–313, Macau, China, 2019.
- [73] Gediminas Adomavicius and YoungOk Kwon. 2012. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*
- [74] Gediminas Adomavicius and YoungOk Kwon. 2011. Maximizing aggregate recommendation diversity: A graph-theoretic approach. In *Proceedings of the 1st International Workshop on Novelty and Diversity in Recommender Systems*

- [75] Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 2012
- [76] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the 5th ACM Conference on Recommender Systems*.
- [77] Oscar Celma. 2009. ‘ Music Recommendation and Discovery in the Long Tail. Ph.D. Dissertation. Universitat Pompeu Fabra
- [78] Rong Hu and Pearl Pu. 2011. Enhancing recommendation diversity with organization interfaces. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*
- [79] Zhou, T., Kuscsik, Z., Liu, J. G., Medo, M., Wakeling, J. R., and Zhang, Y. C. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10), 4511-4515.
- [80] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 5, 4, Article 19 (December 2015), 19 pages. DOI=<http://dx.doi.org/10.1145/2827872>
- [81] Castells, P., Vargas, S., and Wang, J. (2011). Novelty and diversity metrics for recommender systems: choice, discovery and relevance. In *Proceedings of the fifth ACM conference on Recommender systems* (pp. 109-116).
- [82] McNee, S. M., Riedl, J., and Konstan, J. A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. April 2006
- [83] Ziegler, C. N., McNee, S. M., Konstan, J. A., and Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web* (pp. 22-32).
- [84] Alejandro Bellogín, Ivan Cantador, and Pablo Castells. 2013. A comparative study of heterogeneous item recommendations in social systems. *Information Sciences*
- [85] ”F1-Micro and F1-Macro: A Simultaneous Evaluation of Precision, Recall, and F1 Score” Authors: Sokolova, M., and Lapalme, G. Published in: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016*