



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

**ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗ»
ΚΑΤΕΥΘΥΝΣΗ «ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΙΣΤΟΥ»**

**Ανάλυση και Αξιολόγηση Αλγορίθμων για Συστήματα Συστάσεων
Διπλωματική Εργασία**

Καραπάνος Δημήτριος

Αθήνα, 2019



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗ»
ΚΑΤΕΥΘΥΝΣΗ «ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΙΣΤΟΥ»

Τριμελής Εξεταστική Επιτροπή

Βαρλάμης Ηρακλής (Επιβλέπων)
Αναπληρωτής Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο

Μιχαήλ Δημήτριος
Επίκουρος Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο

Τσερπές Κωνσταντίνος
Επίκουρος Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο

Ο Καραπάνος Δημήτριος

δηλώνω υπεύθυνα ότι:

- 1) Είμαι ο κάτοχος των πνευματικών δικαιωμάτων της πρωτότυπης αυτής εργασίας και από όσο γνωρίζω η εργασία μου δε συκοφαντεί πρόσωπα, ούτε προσβάλει τα πνευματικά δικαιώματα τρίτων.
- 2) Αποδέχομαι ότι η ΒΚΠ μπορεί, χωρίς να αλλάξει το περιεχόμενο της εργασίας μου, να τη διαθέσει σε ηλεκτρονική μορφή μέσα από τη ψηφιακή Βιβλιοθήκη της, να την αντιγράψει σε οποιοδήποτε μέσο ή/και σε οποιοδήποτε μορφότυπο καθώς και να κρατά περισσότερα από ένα αντίγραφα για λόγους συντήρησης και ασφάλειας.

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα εργασία αποτελεί Διπλωματική εργασία στα πλαίσια του μεταπτυχιακού προγράμματος Τεχνολογίες και Εφαρμογές Ιστού του τμήματος Πληροφορικής και Τηλεματικής. Καθώς την ολοκληρώνω θα ήθελα να ευχαριστήσω όλους όσους συνέβαλαν στην επιτυχή ολοκλήρωση της.

Καταρχάς, οφείλω ένα μεγάλο ευχαριστώ στον επιβλέποντα Καθηγητή της εργασίας αυτής κ.Ηρακλή Βαρλάμη που με στήριξε, με βοήθησε στην επιλογή θέματος, έδειξε υπομονή και κατανόηση και με κατεύθυνε αποτελεσματικά με στόχο την υλοποίησή της.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους/ες μου οι οποίοι με στήριξαν κατά την διάρκεια των σπουδών μου και ιδιαίτερα κατά την περίοδο υλοποίησης της εργασίας αυτής.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Περίληψη	7
Abstract	9
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ.....	10
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ.....	11
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ	12
ΚΕΦ.1: Εισαγωγή στα Συστήματα Συστάσεων.....	13
1.1. Η Διήθηση Πληροφορίας (Information Filtering) και η χρησιμότητα τους στα Συστήματα Συστάσεων	15
1.2. Μοντέλα Συστημάτων Συστάσεων.....	17
1.2.1. Μοντέλο Φιλτραρίσματος Συνεργασίας	17
1.2.2. Μοντέλο Φιλτραρίσματος Βάσει Περιεχομένου	19
1.3. Κατηγορίες Συστήματα Συστάσεων	21
1.3.1. Συστήματα Συστάσεων Πολλαπλών Κριτήριων	21
1.3.2. Συστήματα Συνιστώμενων Κίνδυνων.....	21
1.3.3. Κινητά Συστήματα Συστάσεων.....	22
1.3.4. Υβριδικά Συστήματα Συστάσεων	23
ΚΕΦ.2: Σύστημα Συστάσεων με Έμφαση στο Γενικό Πλαίσιο (CARs).....	24
2.1. Η Έννοια του Πλαισίου στα Συστήματα Συστάσεων.....	26
2.2. Οι Γνώσεις του Συστήματος Συστάσεων για τις Παραμέτρους Πλαισίου	26
2.3. Μεταβολή Παραγόντων Πλαισίου στο Χρόνο	27
2.4. Παρουσίαση και Μοντελοποίηση του Πλαισίου	31
ΚΕΦ.3: Αλγόριθμοι και Σύνολα Δεδομένων στα CARs.....	34
3.1. Αλγόριθμοι που χρησιμοποιούνται στα CARs	34
3.2. Σύνολα Δεδομένων που Χρησιμοποιούνται στα CARs	38
3.3. Προοπτικές για τη Βελτίωση της Λειτουργίας των CARs	42
ΚΕΦ.4: Η Μηχανή Ανοιχτού Κώδικα CARKit	46
4.1. Παλαιότεροι Αλγόριθμοι και Μηχανισμοί Συστάσεων	46
4.2. Σχεδιασμός και Χαρακτηριστικά του CAR Kit.....	48
4.2.1. Αρχιτεκτονική και Σχεδιασμός του CAR Kit	48
4.2.2. Χαρακτηριστικά του CAR Kit.....	49
ΚΕΦ.5: Αξιολόγηση με Πειραματικά Δεδομένα.....	55
5.1. Δεδομένα.....	55
5.2. Αποτελέσματα	57

ΚΕΦ.6: Συμπεράσματα.....	59
Βιβλιογραφία	61

Περίληψη

Τις τελευταίες δεκαετίες, με την άνοδο του Youtube, του Amazon, του Netflix και πολλών άλλων παρόμοιων δικτυακών υπηρεσιών, τα “Συστήματα Συστάσεων (Recommender Systems – RSs)” έχουν διαδραματίσει έναν ιδιαίτερο ρόλο τόσο στην ανάπτυξη των παραπάνω εφαρμογών όσο και στην εμπειρία που βιώνουν οι χρήστες τους. Από το ηλεκτρονικό εμπόριο όπου τα Συστήματα Συστάσεων προτείνουν στους αγοραστές προϊόντα που θα μπορούσαν να τους ενδιαφέρουν μέχρι τη διαδικτυακή διαφήμιση όπου προτείνουν στους χρήστες τα σωστά περιεχόμενα που ταιριάζουν στις προτιμήσεις τους, τα Συστήματα Συστάσεων πλέον αποτελούν ένα αναπόσπαστο κομμάτι της σύγχρονης πλοήγησης στο διαδίκτυο.

Τα Συστήματα Συστάσεων μέσω της αλγοριθμικής τους ανάπτυξης διαθέτουν την ικανότητα να προτείνουν στους χρήστες τα αντικείμενα που τους ενδιαφέρουν, τα οποία μπορεί να είναι ταινίες, τραγούδια, προϊόντα, ειδήσεις, προορισμοί, ξενοδοχεία ή οτιδήποτε άλλο ανάλογα με τον τύπο της εταιρίας που έχει αναπτύξει τα παραπάνω συστήματα. Τα τελευταία χρόνια τα Συστήματα Συστάσεων και η αποτελεσματικότητά τους αποτελούν ένα ιδιαίτερα κρίσιμο στοιχείο για τις εταιρίες, καθώς από αυτά μπορεί να κρίνεται σε μεγάλο βαθμό η κερδοφορία τους άλλα παράλληλα λειτουργούν και ως ένα εργαλείο που μπορεί να τις κάνει να ξεχωρίζουν από τους ανταγωνιστές τους.

Τα “Συστήματα Συστάσεων με Έμφαση στο Γενικό Πλαίσιο (Context - Aware Recommender Systems - CARSS)” έρχονται να κάνουν ακόμα πιο αποτελεσματικά τα Συστήματα Συστάσεων, επισημαίνοντας με αξιόπιστο τρόπο κάθε ενέργεια του χρήστη και προσαρμόζοντας αποτελεσματικά την έξοδο του συστήματος προς αυτόν σε ένα συγκεκριμένο πλαίσιο. Στα CARSS, τα συμφραζόμενα επιλέγονται συνήθως από τα δυναμικά χαρακτηριστικά της δραστηριότητας του χρήστη. Ωστόσο, δεν υπάρχουν μηχανισμοί συστάσεων που να υλοποιούν αλγορίθμους συστάσεων ενήμερους με τα συμφραζόμενα, λόγω των ιδιαίτερων χαρακτηριστικών της μορφής των δεδομένων και των μεθόδων επεξεργασίας στον τομέα αυτό. Η παρούσα εργασία έχει σκοπό να περιγράψει τη λειτουργία και τη χρήση των Συστημάτων Συστάσεων και των CARSS, των αλγορίθμων λειτουργίας τους καθώς και των συνόλων δεδομένων που χρησιμοποιούν. Στο τέλος, θα γίνει μια υλοποίηση των αλγορίθμων του CARS σε ορισμένα δεδομένα ώστε να δούμε ποιοι αλγόριθμοι θα είναι πιο αποδοτικοί.

Λέξεις κλειδιά: Σύστημα Συστάσεων, RS, Σύστημα Συστάσεων με Έμφαση στο Γενικό Πλαίσιο, CARS, CARS Kit

Abstract

In recent decades, with the rise of Youtube, Amazon, Netflix and many other similar online services, Recommender Systems (RSs) have played a significant role in both the development of the above applications and the experience of the users. From e-commerce where Recommender Systems offer shoppers products that might interest them to online advertising where they offer users the right content that suits their preferences, Recommender Systems are now an integral part of modern internet browsing.

Recommender Systems through their algorithmic development have the ability to recommend to users the items of interest to them, which can be movies, songs, products, news, destinations, hotels or anything else depending on the type of company that has developed the above systems. In recent years, RSs and their effectiveness have been a crucial factor for companies, as they may be largely judged by their profitability but also function as a tool that can make them stand out from their competitors.

Context - Aware Recommender Systems (CARs) come to make Recommender Systems even more effective by reliably tracking every action of the user and effectively adjusting the output of the system to a specific context. In CARs, contexts are usually selected from the dynamic characteristics of the user's activity. However, there are no recommendation mechanisms that implement context-aware recommendation algorithms, due to the particular characteristics of the data format and processing methods in this area. This paper aims to describe the operation and use of Recommender Systems and CARs, their operating algorithms, and the data sets they use. Finally, as a proof of concept, the CAR algorithms are trained and executed on sample data sets.

Keywords: Recommendation System, RS, General Framework Recommendation System, CARs, CAR Kit

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1. Η αρχιτεκτονική και ο σχεδιασμός του CARS Kit	49
Εικόνα 2. Δείγμα των αξιολογήσεων σε δεδομένα του DePaul Movie	54
Εικόνα 3. Εκτέλεση διαφόρων αλγορίθμων για το σύνολο δεδομένων Tijuana Restaurants	57

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1. Γνώση του Συστήματος Συστάσεων σχετικά με τους παράγοντες πλαισίου.....	28
Πίνακας 2. Σύνολα δεδομένων που χρησιμοποιούνται γενικότερα στα Συστήματα Συστάσεων και στα CARS.	42
Πίνακας 3. Σύντομη σύγκριση των πιο δημοφιλών αλγορίθμων συστάσεων και μηχανισμών συστάσεων (Zheng et. al., 2015).....	47
Πίνακας 4. Χαλαρή Μορφή.....	50
Πίνακας 5. Συμπαγής Μορφή.....	50
Πίνακας 6. Μετρικές απόδοσης των αλγορίθμων για τα δεδομέναTijuana Restaurants	58
Πίνακας 7. Μετρικές απόδοσης των αλγορίθμωνγια τα δεδομένα Japan Restaurants	58

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

BN	Bayesian Network
CAMF	Context-Aware Matrix Factorization
CARS	Context - Aware Recommender Systems
DRARS	Dynamic Risk-Aware Recommender System
FM	Factorization Machine
HLM	Hierarchical Linear Model
LDA	Latent Dirichlet Allocation
MAB	Multi-Armed Bandit
MAE	Mean Absolute Error
MAP	Mean Average Precision
MPE	Mean Percentage Error
MCDM	Multi-Criteria Decision Making
MCRS	Multi-Criteria Recommender Systems
MF	Matrix Factorization
MMR	Mean Reciprocal Rank
NDCG	Normalized Discounted Cumulative Gain
PMF	Probabilistic Matrix Factorization
POI	Points of Interest
RS	Recommender System
RMSE	Root Mean Square Error
TF	Tensor Factorization

ΚΕΦ.1: Εισαγωγή στα Συστήματα Συστάσεων

Στις μέρες μας έχουν αναπτυχθεί συστήματα που μαθαίνουν τα ενδιαφέροντα και τις ανάγκες των χρηστών του διαδικτύου και συστήνουν τις επιλογές που ταιριάζουν στα προφίλ των χρηστών αυτών. Αυτά τα Συστήματα είναι γνωστά ως Συστήματα Συστάσεων (Recommender System – RS). Ένα Σύστημα Συστάσεων αποτελεί υποκλάση του συστήματος φιλτραρίσματος πληροφοριών που επιχειρεί να προβλέψει την "αξιολόγηση" ή την "προτίμηση" που μπορεί να έχει ένας χρήστης σε σχέση με ένα αντικείμενο. (Ricci et al., 2011). Τα Συστήματα Συστάσεων χρησιμοποιούνται κυρίως σε εμπορικές εφαρμογές ποικίλων τομέων και συνήθως αναγνωρίζονται ως γεννήτριες λιστών αναπαραγωγής (playlists) για υπηρεσίες βίντεο και μουσικής όπως το Netflix, το YouTube και το Spotify, για προτάσεις προϊόντων για υπηρεσίες όπως η Amazon ή για προτάσεις περιεχομένου για πλατφόρμες κοινωνικών μέσων όπως το Facebook και το Twitter. Αυτά τα συστήματα μπορούν να λειτουργούν χρησιμοποιώντας μία μόνο είσοδο, όπως τη μουσική ή πολλαπλές εισόδους εντός και μεταξύ πλατφορμών όπως αυτές των ειδήσεων, βιβλίων και ερωτημάτων αναζήτησης. Υπάρχουν επίσης δημοφιλή Συστήματα Συστάσεων για συγκεκριμένα θέματα, όπως εστιατόρια και διαδικτυακά ραντεβού. Επιπλέον, έχουν αναπτυχθεί Συστήματα Συστάσεων για τη διερεύνηση ερευνητικών άρθρων και εμπειρογνομόνων, συνεργατών, χρηματοπιστωτικών υπηρεσιών και ασφάλισης ζωής. Τα Συστήματα Συστάσεων αναφέρθηκαν για πρώτη φορά σε μία τεχνική αναφορά του Jussi Karlgren στο Πανεπιστήμιο Columbia το 1990 (Karlgren, 1994) ως «ψηφιακές βιβλιοθήκες» και χρησιμοποιήθηκαν σε μεγάλο βαθμό σε τεχνικές αναφορές και εκδόσεις από το 1994 και έπειτα (Karlgren, 1994 ; Hill et al., 1995 ; Resnick & Varian, 1997)

Τα Συστήματα Συστάσεων συνήθως χρησιμοποιούν είτε μεμονωμένα είτε σε συνδυασμό το φιλτράρισμα συνεργασίας και το φιλτράρισμα περιεχομένου, μία τεχνική γνωστή και ως προσέγγιση βασισμένη στην προσωπικότητα, καθώς και άλλα συστήματα, όπως συστήματα βασισμένα στη γνώμη. Οι συνεργατικές προσεγγίσεις φιλτραρίσματος δημιουργούν ένα μοντέλο σύμφωνα με την παρελθούσα συμπεριφορά του χρήστη, όπως στοιχεία που έχουν αγοραστεί προηγουμένως ή έχουν επιλεγεί και / ή αριθμητική βαθμολογία που δόθηκε σε αυτά τα στοιχεία, καθώς και παρόμοιες αποφάσεις που έγιναν από άλλους χρήστες. Αυτό το μοντέλο χρησιμοποιείται στη συνέχεια για την πρόβλεψη ή αξιολόγηση στοιχείων που μπορεί να έχουν ενδιαφέρον για τον χρήστη. Οι προσεγγίσεις φιλτραρίσματος που βασίζονται στο

περιεχόμενο χρησιμοποιούν μια σειρά διακριτών χαρακτηριστικών ενός αντικειμένου, προκειμένου να προτείνουν πρόσθετα αντικείμενα με παρόμοιες ιδιότητες (Mooney & Roy, 2000). Τα σύγχρονα Συστήματα Συστάσεων συνήθως συνδυάζουν μία ή περισσότερες προσεγγίσεις σε ένα υβριδικό σύστημα.

Οι διαφορές μεταξύ φιλτραρίσματος συνεργασίας και φιλτραρίσματος περιεχομένου μπορούν να αποδοθούν με τη σύγκριση δύο Συστημάτων Συστάσεων για μουσική, του Last.fm και του Pandora Radio.

- Το Last.fm δημιουργεί ένα "σταθμό" των συνιστώμενων τραγουδιών παρατηρώντας τις ζώνες και τα μεμονωμένα κομμάτια που έχει ακούσει ο χρήστης σε τακτική βάση και συγκρίνοντας αυτά με την συμπεριφορά ακρόασης άλλων χρηστών. Το Last.fm θα παίξει κομμάτια που δεν εμφανίζονται στη βιβλιοθήκη του χρήστη, αλλά συχνά αναπαράγονται από άλλους χρήστες με παρόμοια ενδιαφέροντα. Καθώς αυτή η προσέγγιση αξιοποιεί τη συμπεριφορά των χρηστών, είναι ένα παράδειγμα μιας συνεργατικής τεχνικής φιλτραρίσματος.
- Η Pandora χρησιμοποιεί τις ιδιότητες ενός τραγουδιού ή ενός καλλιτέχνη (ένα υποσύνολο των 400 χαρακτηριστικών που παρέχει το Music Genome Project) για να δημιουργήσει έναν "σταθμό" που παίζει μουσική με παρόμοιες ιδιότητες. Τα σχόλια των χρηστών χρησιμοποιούνται για να βελτιώσουν τα αποτελέσματα του σταθμού, να υπογραμμίσουν ορισμένα χαρακτηριστικά όταν ένας χρήστης "δεν συμπαθεί" ένα συγκεκριμένο τραγούδι και να τονίσει άλλα χαρακτηριστικά όταν σε ένα χρήστη "αρέσει" ένα τραγούδι. Αυτό είναι ένα παράδειγμα μιας προσέγγισης φιλτραρίσματος με βάση το περιεχόμενο.

Κάθε τύπος συστήματος έχει τα πλεονεκτήματα και τις αδυναμίες του. Στο παραπάνω παράδειγμα, το Last.fm απαιτεί μια μεγάλη ποσότητα πληροφοριών για έναν χρήστη ώστε να κάνει ακριβείς συστάσεις. Αυτό είναι ένα παράδειγμα των προβλημάτων ψυχρής εκκίνησης (coldstart) και είναι συνηθισμένο στα συνεργατικά συστήματα φιλτραρίσματος. (Machado, 2019). Αντίθετα, η Pandora χρειάζεται ελάχιστες πληροφορίες για να ξεκινήσει, ωστόσο είναι πολύ πιο περιορισμένη στο πεδίο εφαρμογής της.

Τα Συστήματα Συστάσεων αποτελούν μία χρήσιμη εναλλακτική λύση για τους αλγόριθμους αναζήτησης, δεδομένου ότι βοηθούν τους χρήστες να ανακαλύπτουν στοιχεία που

ενδεχομένως να μην έχουν ξανασυναντήσει. Είναι αξιοσημείωτο ότι τα Συστήματα Συστάσεων συχνά υλοποιούνται με τη χρήση μηχανών αναζήτησης που καταγράφουν μη παραδοσιακά δεδομένα.

1.1. Η Διήθηση Πληροφορίας (Information Filtering) και η χρησιμότητα τους στα Συστήματα Συστάσεων

Ένα σύστημα “Διήθησης Πληροφορίας (Information Filtering)” είναι ένα σύστημα που αφαιρεί περιττές ή ανεπιθύμητες πληροφορίες από μια ροή πληροφοριών χρησιμοποιώντας ημιαυτοματοποιημένες ή υπολογιστικές μεθόδους πριν από την παρουσίασή τους στον τελικό χρήστη. Ο κύριος στόχος του είναι η διαχείριση της υπερφόρτωσης πληροφοριών και η αύξηση του λόγου του σημασιολογικού σήματος προς τον θόρυβο. Για την επίτευξη αυτού του στόχου, το προφίλ του χρήστη συγκρίνεται με ορισμένα χαρακτηριστικά αναφοράς. Αυτά τα χαρακτηριστικά μπορεί να προέρχονται από το αντικείμενο πληροφοριών, που αποτελεί την προσέγγιση που βασίζεται στο φιλτράρισμα βάσει του περιεχόμενου ή από το κοινωνικό περιβάλλον του χρήστη, που αποτελεί τη συνεργατική προσέγγιση φιλτραρίσματος.

Τα συστήματα συστάσεων και οι πλατφόρμες ανακάλυψης περιεχομένου αποτελούν συστήματα ενεργούς διήθησης πληροφορίας που προσπαθούν να παρουσιάσουν τα στοιχεία πληροφοριών του χρήστη (ταινίες, μουσική, βιβλία, ειδήσεις, ιστοσελίδες) τα οποία τον ενδιαφέρουν. Τα συστήματα αυτά προσθέτουν στοιχεία πληροφοριών στις πληροφορίες που ρέουν προς το χρήστη, σε αντίθεση με την αφαίρεση στοιχείων πληροφοριών από τη ροή πληροφοριών προς τον χρήστη. Τα συστήματα συστάσεων συνήθως χρησιμοποιούν συνεργατικές προσεγγίσεις φιλτραρίσματος ή έναν συνδυασμό συνεργατικού φιλτραρίσματος και προσεγγίσεων φιλτραρίσματος βάσει του περιεχομένου.

Ένα σύστημα φιλτραρίσματος αυτής της προσέγγισης αποτελείται από διάφορα εργαλεία που βοηθούν τους χρήστες να βρουν τις πιο πολύτιμες πληροφορίες, οπότε ο περιορισμένος χρόνος που μπορεί να αφιερώσουν για να διαβάσουν, να ακούσουν, να παρακολουθήσουν κατευθύνεται σωστά στα πιο ενδιαφέροντα και πολύτιμα αρχεία. Αυτά τα φίλτρα χρησιμοποιούνται επίσης για την οργάνωση και τη δομή των πληροφοριών με έναν σωστό και κατανοητό τρόπο. Επίσης, είναι απαραίτητα για την αποτελεσματικότητα των μηχανών αναζήτησης στο Διαδίκτυο. Οι λειτουργίες του φιλτραρίσματος βελτιώνονται καθημερινά, έτσι

ώστε οι χρήστες να λαμβάνουν αποτελεσματικότερα μηνύματα και αρχεία μέσω του διαδικτύου.

Η διήθηση πληροφορίας ασχολείται με την παροχή πληροφοριών που ο χρήστης είναι πιθανό να βρει ενδιαφέρουσα ή χρήσιμη. Ένα σύστημα διήθηση πληροφορίας βοηθά τους χρήστες φιλτράροντας μία πηγή δεδομένων και παρέχοντάς τους σχετικές πληροφορίες. Όταν η παραδοθείσα πληροφορία έρχεται με τη μορφή προτάσεων, τότε το σύστημα διήθησης πληροφορίας ονομάζεται σύστημα συστάσεων. Λόγω του ότι οι χρήστες έχουν διαφορετικά ενδιαφέροντα, το σύστημα διήθηση πληροφορίας πρέπει να είναι εξατομικευμένο ώστε να εξυπηρετεί τα συμφέροντα των μεμονωμένων χρηστών. Αυτό απαιτεί τη συλλογή ανατροφοδοτήσεων από το χρήστη προκειμένου να δημιουργηθεί ένα προφίλ των προτιμήσεων του χρήστη.

Ένα σύστημα συστάσεων παρέχει στο χρήστη προτάσεις βάσει των πληροφοριών που αναζήτησε. Ωστόσο, η αποτελεσματικότητα των συστημάτων συστάσεων απαιτεί την ύπαρξη ενός συστήματος διήθησης πληροφοριών, το οποίο είναι υπεύθυνο για την παρουσίαση της όσο το δυνατό περισσότερο στοχευμένης προς τα ενδιαφέροντα του χρήστη εξαγόμενης πληροφορίας, λαμβάνοντας υπόψη τη συσχέτιση των αντικειμένων που αναζητούνται με τα χαρακτηριστικά, τις συνήθειες και τις τάσεις του χρήστη.

Μια πληθώρα πληροφοριών δημιουργείται και παραδίδεται μέσω ηλεκτρονικών μέσων. Οι χρήστες κινδυνεύουν να κατακλύζονται από τη ροή των πληροφοριών, ενώ παράλληλα δεν διαθέτουν τα κατάλληλα εργαλεία για να βοηθήσουν στη διαχείριση της κατάστασης. Η Διήθηση Πληροφοριών αποτελεί μια από τις μεθόδους που εξελίσσονται ταχύτατα για τη διαχείριση μεγάλων ροών πληροφοριών. Ο στόχος της Διήθησης Πληροφοριών είναι να εκθέσει τους χρήστες μόνο σε πληροφορίες που σχετίζονται με αυτές. Πολλά συστήματα Διήθησης Πληροφοριών έχουν αναπτυχθεί τα τελευταία χρόνια για διάφορους τομείς εφαρμογών. Μερικά παραδείγματα εφαρμογών φιλτραρίσματος είναι: φίλτρα για αποτελέσματα αναζήτησης στο διαδίκτυο που χρησιμοποιούνται στο λογισμικό Διαδικτύου, φίλτρα προσωπικής αλληλογραφίας που βασίζονται σε προσωπικά προφίλ, λίστες διακομιστών ή ομάδες συζητήσεων για ομάδες ή άτομα, φίλτρα προγράμματος περιήγησης που αποκλείουν μη πολύτιμες πληροφορίες, φίλτρα που έχουν σχεδιαστεί για να έχουν πρόσβαση τα παιδιά σε κατάλληλες σελίδες, φίλτρα για εφαρμογές ηλεκτρονικού εμπορίου

που απευθύνονται μόνο σε δυνητικούς πελάτες, προϊόντα και προσφορές κ.λ.π. (Dehuri, 2012).

1.2. Μοντέλα Συστημάτων Συστάσεων

1.2.1. Μοντέλο Φιλτραρίσματος Συνεργασίας

Μια προσέγγιση για το σχεδιασμό ενός Συστήματος Συστάσεων με ευρεία χρήση είναι το συνεργατικό φιλτράρισμα. (Breese et al., 1998) Το συνεργατικό φιλτράρισμα βασίζεται στην υπόθεση ότι οι άνθρωποι που συμφώνησαν στο παρελθόν θα συμφωνήσουν στο μέλλον και ότι θα αρέσκονται σε παρόμοια είδη που τους άρεσαν και στο παρελθόν. Το σύστημα δημιουργεί προτάσεις χρησιμοποιώντας μόνο πληροφορίες σχετικά με τα προφίλ αξιολόγησης για διαφορετικούς χρήστες ή στοιχεία. Με τον εντοπισμό ομότιμων χρηστών / αντικειμένων μέσω ενός ιστορικού αξιολόγησης παρόμοιο με τον τρέχοντα χρήστη ή στοιχείο, δημιουργούν συστάσεις χρησιμοποιώντας αυτή τη σχέση. Οι αλγόριθμοι πλησιέστερων σχέσεων που βασίζονται σε χρήστες και στοιχεία μπορούν να συνδυαστούν για να αντιμετωπίσουν το πρόβλημα της ψυχρής εκκίνησης και να βελτιώσουν τα αποτελέσματα των συστάσεων χρησιμοποιώντας αυτά τα δεδομένα. Οι μέθοδοι φιλτραρίσματος συνεργασίας ταξινομούνται ως βασιζόμενοι στη μνήμη και βασιζόμενοι στο μοντέλο. Ένα πολύ γνωστό παράδειγμα των προσεγγίσεων που βασίζονται στη μνήμη είναι ο αλγόριθμος που βασίζεται στον χρήστη, ενώ αυτός των προσεγγίσεων που βασίζονται σε μοντέλο είναι ο Kernel-Mapping Recommender (Ghazanfar et al., 2012)

Ένα βασικό πλεονέκτημα της προσέγγισης φιλτραρίσματος συνεργασίας είναι ότι δεν βασίζεται σε ένα μηχανικά αναλυόμενο περιεχόμενο και ως εκ τούτου είναι σε θέση να προτείνει με ακρίβεια σύνθετα αντικείμενα όπως ταινίες χωρίς να χρειάζεται "κατανόηση" του ίδιου του αντικειμένου. Πολλοί αλγόριθμοι έχουν χρησιμοποιηθεί για τη μέτρηση της ομοιότητας των χρηστών ή της ομοιότητας των στοιχείων στα συστήματά τους. Κατά την οικοδόμηση ενός μοντέλου από τη συμπεριφορά του χρήστη, γίνεται συχνά διάκριση μεταξύ ρητών και έμμεσων μορφών συλλογής δεδομένων.

Παραδείγματα ρητής συλλογής δεδομένων πραγματοποιείται με τους ακόλουθους τρόπους:

- Ζητώντας από ένα χρήστη να αξιολογήσει ένα στοιχείο σε μια ολιστική κλίμακα.

- Ζητώντας από ένα χρήστη να ψάξει.
- Ζητώντας από ένα χρήστη να ταξινομήσει μια συλλογή αντικειμένων από τα αγαπημένα στα λιγότερο αγαπημένα.
- Παρουσιάζοντας δύο στοιχεία σε ένα χρήστη και ζητώντας του να επιλέξει το καλύτερο από αυτά.
- Ζητώντας από έναν χρήστη να δημιουργήσει μια λίστα με αντικείμενα που του αρέσουν.

Παραδείγματα σιωπηρής συλλογής δεδομένων πραγματοποιείται με τους ακόλουθους τρόπους:

- Παρατήρηση των στοιχείων που βλέπει ένας χρήστης σε ένα ηλεκτρονικό κατάστημα.
- Αναλύοντας τους χρόνους προβολής των στοιχείων. (Parsons et al., 2004)
- Διατηρώντας μια καταγραφή των στοιχείων που αγοράζει ένας χρήστης στο διαδίκτυο.
- Λαμβάνοντας μία λίστα αντικειμένων που ένας χρήστης έχει ακούσει ή έχει παρακολουθήσει στον υπολογιστή του.
- Ανάλυση του προφίλ του χρήστη στα κοινωνικά δίκτυα σχετικά με τις συμπάθειες και τις αντιπάθειές τους.

Οι συνεργατικές προσεγγίσεις φιλτραρίσματος συχνά παρουσιάζουν τρία προβλήματα που σχετίζονται με την ψυχρή εκκίνηση, την υπολογιστική ισχύ και την ανεπάρκεια ως προς τα δεδομένα:

- **Ψυχρή εκκίνηση:** Για ένα νέο χρήστη ή στοιχείο, δεν υπάρχουν αρκετά δεδομένα που να παρέχουν ακριβείς συστάσεις. (Machado, 2019)
- **Υπολογιστική Ισχύ:** Σε πολλά από τα περιβάλλοντα στα οποία αυτά τα συστήματα διατυπώνουν συστάσεις, υπάρχουν εκατομμύρια χρηστών και προϊόντων. Επομένως, για τον υπολογισμό των συστάσεων είναι συχνά απαραίτητη μια μεγάλη ποσότητα υπολογιστικής ισχύος.

- **Ανεπάρκεια Δεδομένων:** Ο αριθμός των ειδών που πωλούνται σε μεγάλους ιστότοπους ηλεκτρονικού εμπορίου είναι εξαιρετικά μεγάλος. Οι πιο ενεργοί χρήστες θα έχουν βαθμολογήσει μόνο ένα μικρό υποσύνολο της συνολικής βάσης δεδομένων. Έτσι, ακόμη και τα πιο δημοφιλή αντικείμενα έχουν πολύ λίγες αξιολογήσεις.

Ένα από τα πιο γνωστά παραδείγματα συνεργατικού φιλτραρίσματος είναι το συνεργατικό φιλτράρισμα αντικειμένου προς αντικείμενο, καθώς άνθρωποι που αγοράζουν το x αντικείμενο επίσης αγοράζουν το y , έναν δημοφιλή αλγόριθμο από το Σύστημα Συστάσεων που χρησιμοποιεί η Amazon (Linden et al., 2001).

Πολλά κοινωνικά δίκτυα χρησιμοποίησαν αρχικά συνεργατικό φιλτράρισμα για να συστήσουν νέους φίλους, ομάδες και άλλες κοινωνικές συνδέσεις εξετάζοντας το δίκτυο των συνδέσεων μεταξύ ενός χρήστη και των φίλων τους. (Ricci et al., 2011) Το συνεργατικό φιλτράρισμα εξακολουθεί να χρησιμοποιείται ως μέρος υβριδικών συστημάτων.

1.2.2. Μοντέλο Φιλτραρίσματος Βάσει Περιεχομένου

Μια άλλη κοινή προσέγγιση κατά το σχεδιασμό των Συστημάτων Συστάσεων είναι το φιλτράρισμα βάσει περιεχομένου. Οι μέθοδοι φιλτραρίσματος βάσει περιεχομένου βασίζονται σε μια περιγραφή του στοιχείου και ενός προφίλ των προτιμήσεων του χρήστη (Akhil & Joseph, 2017). Αυτές οι μέθοδοι είναι κατάλληλες για περιπτώσεις που υπάρχουν γνωστά δεδομένα για ένα στοιχείο, όπως όνομα, τοποθεσία, περιγραφή κ.λπ., αλλά όχι και για τον χρήστη. Οι μέθοδοι φιλτραρίσματος βάσει περιεχομένου αντιμετωπίζουν τη σύσταση – πρόταση ως ένα πρόβλημα ταξινόμησης συγκεκριμένων χρηστών και μαθαίνουν έναν ταξινομο για τις προτιμήσεις και τις αντιδράσεις του χρήστη ως προς τις λειτουργίες του προϊόντος.

Σε αυτό το σύστημα, οι λέξεις-κλειδιά χρησιμοποιούνται για την περιγραφή των στοιχείων, ενώ ένα προφίλ χρήστη είναι κατασκευασμένο για να υποδείξει τον τύπο αντικειμένου που αυτός ο χρήστης επιθυμεί. Με άλλα λόγια, αυτοί οι αλγόριθμοι προσπαθούν να προτείνουν αντικείμενα που είναι παρόμοια με αυτά που άρεσαν σε ένα χρήστη στο παρελθόν. Δεν βασίζεται σε μηχανισμό σύνδεσης χρηστών για τη δημιουργία αυτού του συχνά προσωρινού προφίλ. Συγκεκριμένα, τα διάφορα υποψήφια στοιχεία συγκρίνονται με στοιχεία που έχουν προηγουμένως βαθμολογηθεί από το χρήστη και προτείνονται τα καλύτερα αντικείμενα που

ταιριάζουν. Αυτή η προσέγγιση έχει τις ρίζες της στην ανάκτηση πληροφοριών και την έρευνα φιλτραρίσματος πληροφοριών.

Για τη δημιουργία ενός προφίλ χρήστη, το σύστημα επικεντρώνεται κυρίως σε δύο τύπους πληροφοριών:

- Ένα μοντέλο με τις προτιμήσεις του χρήστη.
- Ένα ιστορικό της αλληλεπίδρασης του χρήστη με το Σύστημα Συστάσεων.

Αυτές οι μέθοδοι χρησιμοποιούν ένα προφίλ αντικειμένου, δηλαδή ένα σύνολο διακεκριμένων γνωρισμάτων που χαρακτηρίζουν το στοιχείο μέσα στο σύστημα. Για την αφαίρεση των χαρακτηριστικών των αντικειμένων του συστήματος, εφαρμόζεται ένας αλγόριθμος παρουσίασης αντικειμένων. Ένας ευρέως χρησιμοποιούμενος αλγόριθμος είναι η tf-idf αναπαράσταση, γνωστή και ως διανυσματική παρουσίαση διαστήματος (vector spacer presentation). Το σύστημα δημιουργεί ένα προφίλ με βάση το περιεχόμενο των χρηστών βάσει ενός σταθμισμένου διανύσματος χαρακτηριστικών στοιχείων. Τα βάρη υποδηλώνουν τη σημασία κάθε χαρακτηριστικού για τον χρήστη και μπορούν να υπολογιστούν από μεμονωμένους διαβαθμισμένους φορείς περιεχομένου χρησιμοποιώντας μια ποικιλία τεχνικών. Οι απλές προσεγγίσεις χρησιμοποιούν τις μέσες διανυσματικές τιμές του αξιολογούμενου αντικειμένου ενώ άλλες εξελιγμένες μέθοδοι χρησιμοποιούν τεχνικές μηχανικής μάθησης, ανάλυση συμπλέγματος, δέντρα αποφάσεων και τεχνητά νευρωνικά δίκτυα για να εκτιμήσουν την πιθανότητα ότι ο χρήστης θα αρέσκεται στο στοιχείο.

Ένα βασικό ζήτημα με το φιλτράρισμα βάσει περιεχομένου είναι αν το σύστημα μπορεί να μάθει τις προτιμήσεις των χρηστών από τις ενέργειες τους σχετικά με μια πηγή περιεχομένου και να τις χρησιμοποιήσει σε άλλους τύπους περιεχομένου. Όταν το σύστημα περιορίζεται στο να προτείνει περιεχόμενο του ίδιου τύπου με αυτό που ήδη χρησιμοποιεί ο χρήστης, η τιμή από το Σύστημα Συστάσεων είναι σημαντικά μικρότερη από ότι όταν μπορούν να προτείνονται άλλοι τύποι περιεχομένου από άλλες υπηρεσίες. Για παράδειγμα, τα προτεινόμενα άρθρα ειδήσεων που βασίζονται στην περιήγηση ειδήσεων είναι χρήσιμα, αλλά θα ήταν πολύ πιο χρήσιμα όταν η μουσική, τα βίντεο, τα προϊόντα, οι συζητήσεις κ.λπ. από διαφορετικές υπηρεσίες θα μπορούσαν να προταθούν με βάση την περιήγηση ειδήσεων. Για να ξεπεραστεί αυτό, τα περισσότερα Σύστημα Συστάσεων με βάση το περιεχόμενο χρησιμοποιούν πλέον κάποια μορφή υβριδικού συστήματος.

Τα Συστήματα Συστάσεων που βασίζονται στο περιεχόμενο μπορούν επίσης να περιλαμβάνουν Συστήματα Συστάσεων που βασίζονται στη γνώμη. Σε ορισμένες περιπτώσεις, επιτρέπεται στους χρήστες να αφήνουν μία αναφορά κειμένου ή μία ανατροφοδότηση σχετικά με τα αντικείμενα. Αυτά τα κείμενα που δημιουργούνται από τους χρήστες είναι σιωπηρά δεδομένα για το Σύστημα Συστάσεων, επειδή είναι δυνητικά πλούσιοι πόροι των χαρακτηριστικών / πτυχών του αντικειμένου και της αξιολόγησης / αισθήματος των χρηστών για το στοιχείο. Τα χαρακτηριστικά που εξάγονται από τις ανασκοπήσεις που δημιουργούνται από το χρήστη αποτελούν βελτιωμένα μεταδεδομένα των αντικειμένων, επειδή, καθώς αντικατοπτρίζουν πτυχές του αντικειμένου όπως τα μεταδεδομένα, οι εξαγόμενες λειτουργίες, απασχολούν ευρέως τους χρήστες. Τα συναισθήματα που εξάγονται από τις αναθεωρήσεις μπορούν να θεωρηθούν ως βαθμολογίες χρηστών στα αντίστοιχα χαρακτηριστικά. Οι δημοφιλείς προσεγγίσεις του Συστήματος Συστάσεων με βάση τη γνώμη χρησιμοποιούν διάφορες τεχνικές, όπως εξόρυξη κειμένου, ανάκτηση πληροφοριών και ανάλυση.

1.3. Κατηγορίες Συστήματα Συστάσεων

1.3.1. Συστήματα Συστάσεων Πολλαπλών Κριτηρίων

Τα “Συστήματα Συστάσεων Πολλαπλών Κριτηρίων (Multi-Criteria Recommender Systems - MCRS)” μπορούν να οριστούν ως Συστήματα Συστάσεων που ενσωματώνουν πληροφορίες προτιμήσεων σε πολλαπλά κριτήρια. Αντί να αναπτύξουν τεχνικές συστάσεων βασισμένες στις τιμές ενός μόνο κριτηρίου, όπως τη γενική προτίμηση του χρήστη για ένα στοιχείο, τα συστήματα αυτά προσπαθούν να προβλέψουν μια βαθμολογία για ανεξερεύνητα στοιχεία του χρήστη εκμεταλλευόμενες τις πληροφορίες προτίμησης σε πολλαπλά κριτήρια που επηρεάζουν αυτή τη συνολική τιμή προτίμησης. Αρκετοί ερευνητές προσεγγίζουν το MCRS ως πρόβλημα “Πολλαπλών Κριτηρίων Λήψης Αποφάσεων (Multi-Criteria Decision Making - MCDM)” και εφαρμόζουν μεθόδους και τεχνικές MCDM για την εφαρμογή συστημάτων MCRS (Lakiotaki et al., 2011).

1.3.2. Συστήματα Συνιστώμενων Κίνδυνων

Η πλειονότητα των υπαρχουσών προσεγγίσεων στα συστήματά τους, συστήνουν να προτείνουν το πιο σχετικό περιεχόμενο στους χρήστες που χρησιμοποιούν πληροφορίες πλαισίου, αλλά δεν λαμβάνουν υπόψη τον κίνδυνο ενόχλησης του χρήστη με ανεπιθύμητες

ειδοποιήσεις. Είναι σημαντικό να εξεταστεί ο κίνδυνος ενόχλησης του χρήστη, προωθώντας συστάσεις σε ορισμένες περιπτώσεις, για παράδειγμα, κατά τη διάρκεια επαγγελματικής συνάντησης, νωρίς το πρωί ή αργά το βράδυ. Ως εκ τούτου, η απόδοση του Συστήματος Συστάσεων εξαρτάται εν μέρει από το βαθμό στον οποίο ενσωμάτωσε τον κίνδυνο στη διαδικασία της πρότασης. Μια επιλογή για τη διαχείριση αυτού του προβλήματος είναι το “Δυναμικό Σύστημα Συνιστώμενων Κινδύνων (Dynamic Risk-Aware Recommender System - DRARS)”, ένα σύστημα που διαμορφώνει τη σύσταση ανάλογα με το περιβάλλον του χρήστη. Αυτό το σύστημα συνδυάζει μια τεχνική που βασίζεται στο περιεχόμενο και έναν αλγόριθμο συμφραζομένων. (Bouneffouf, 2013)

1.3.3. Κινητά Συστήματα Συστάσεων

Τα “Κινητά Συστήματα Συστάσεων (Mobile RSs)” χρησιμοποιούν τα έξυπνα τηλέφωνα με πρόσβαση στο Διαδίκτυο για να προσφέρουν εξατομικευμένες, ευαίσθητες για το περιβάλλον συστάσεις. Αυτός είναι ένας ιδιαίτερα δύσκολος τομέας έρευνας, καθώς τα δεδομένα κινητής τηλεφωνίας είναι πιο πολύπλοκα από τα δεδομένα που συχνά πρέπει να αντιμετωπίσουν τα Συστήματα Συστάσεων. Είναι ετερογενής, απαιτεί χωρική και χρονική συσχέτιση και έχει προβλήματα επικύρωσης και γενικότητας (Ge et. al., 2010).

Υπάρχουν τρεις παράγοντες που θα μπορούσαν να επηρεάσουν τα Κινητά Συστήματα Συστάσεων και η ακρίβεια των αποτελεσμάτων πρόβλεψης: το πλαίσιο, η μέθοδος των συστάσεων και το απόρρητο. Επιπρόσθετα, τα Κινητά Συστήματα Συστάσεων πάσχουν από ένα πρόβλημα μεταφερσιμότητας, καθώς οι συστάσεις ενδέχεται να μην ισχύουν σε όλες τις περιοχές. Για παράδειγμα, θα ήταν άδικο να συστήσετε συνταγή σε μια περιοχή όπου όλα τα συστατικά ενδέχεται να μην είναι διαθέσιμα.

Ένα παράδειγμα συστήματος Κινητών Συστημάτων Συστάσεων είναι οι προσεγγίσεις εταιρειών όπως η Uber και η Lyft για τη δημιουργία διαδρομών οδήγησης για τους οδηγούς ταξί σε μια πόλη (Ge et al., 2010). Αυτό το σύστημα χρησιμοποιεί δεδομένα GPS για τις διαδρομές που λαμβάνουν οι οδηγοί ταξί κατά τη διάρκεια της εργασίας, η οποία περιλαμβάνει την τοποθεσία (γεωγραφικό πλάτος και μήκος) και την κατάσταση λειτουργίας (με ή χωρίς επιβάτες). Χρησιμοποιεί αυτά τα δεδομένα για να προτείνει μια λίστα σημείων παραλαβής κατά μήκος μιας διαδρομής, με στόχο τη βελτιστοποίηση των χρόνων και των κερδών.

Κινητά Συστήματα Συστάσεων έχουν επίσης κατασκευαστεί με επιτυχία χρησιμοποιώντας το "Web of Data" ως πηγή για δομημένες πληροφορίες. Ένα καλό παράδειγμα ενός τέτοιου συστήματος είναι το SMARTMUSEUM (Ruotsalo et al, 2013) Το σύστημα χρησιμοποιεί τεχνικές σημασιολογικής μοντελοποίησης, ανάκτησης πληροφοριών και μηχανικής μάθησης, προκειμένου να προτείνει συσχετισμό περιεχομένου με το χρήστη, ακόμη και όταν παρουσιάζεται με αραιά ή ελάχιστα δεδομένα χρήστη.

1.3.4. Υβριδικά Συστήματα Συστάσεων

Τα περισσότερα Συστήματα Συστάσεων χρησιμοποιούν πλέον μια υβριδική προσέγγιση, συνδυάζοντας το συνεργατικό φιλτράρισμα, το φιλτράρισμα περιεχομένου και άλλες προσεγγίσεις. Δεν υπάρχει κανένας λόγος για τον οποίο αρκετές διαφορετικές τεχνικές του ίδιου τύπου δεν θα μπορούσαν να υβριδοποιηθούν. Οι υβριδικές προσεγγίσεις μπορούν να υλοποιηθούν με διάφορους τρόπους: λειτουργώντας ξεχωριστά και στη συνέχεια συνδυάζοντας τις βασισμένες σε περιεχόμενο και τις συνεργατικές προβλέψεις με την προσθήκη δυνατοτήτων βάσει περιεχομένου σε μια προσέγγιση βασισμένη στη συνεργασία (και αντίστροφα) ή με την ενοποίηση των προσεγγίσεων σε ένα μοντέλο. Αρκετές μελέτες που συγκρίνουν εμπειρικά τη συμπεριφορά του υβριδικού μοντέλου με τις καθαρές μεθόδους συνεργασίας και περιεχομένου, απέδειξαν ότι οι υβριδικές μέθοδοι μπορούν να παρέχουν ακριβέστερες συστάσεις από τις καθαρές προσεγγίσεις. Αυτές οι μέθοδοι μπορούν επίσης να χρησιμοποιηθούν για να ξεπεραστούν μερικά από τα κοινά προβλήματα στα συστήματά τους, όπως η ψυχρή εκκίνηση και το πρόβλημα της ασυμμετρίας, καθώς και η συμφόρηση της τεχνολογίας της γνώμης σε προσεγγίσεις βασισμένες στη γνώμη. (Hoekstra, 2010)

Το Netflix είναι ένα καλό παράδειγμα χρήσης υβριδικών Συστημάτων Συστάσεων. Ο ιστότοπος κάνει συστάσεις συγκρίνοντας τις συνήθειες παρακολούθησης και αναζήτησης παρόμοιων χρηστών (δηλ. συνεργατικό φιλτράρισμα) καθώς και προσφέροντας ταινίες που μοιράζονται χαρακτηριστικά με ταινίες που ο χρήστης έχει αξιολογήσει ιδιαίτερα (φιλτράρισμα βάσει περιεχομένου).

ΚΕΦ.2: Σύστημα Συστάσεων με Έμφαση στο Γενικό Πλαίσιο (CARSS)

Τα Συστήματα Συστάσεων με Έμφαση στο γενικό Πλαίσιο κάνουν συστάσεις εκμεταλλευόμενα τις πληροφορίες πλαισίου που επηρεάζουν τις προτιμήσεις των χρηστών, έχοντας ως στόχο να προτείνουν αντικείμενα που σχετίζονται με τις μεταβαλλόμενες ανάγκες των χρηστών (Villegas et al, 2018).

Τα παραδοσιακά Συστήματα Συστάσεων, όπως αυτά που βασίζονται στο φιλτράρισμα βάσει περιεχομένου και συνεργασίας, τείνουν να χρησιμοποιούν αρκετά απλά μοντέλα χρηστών. Για παράδειγμα, το συνεργατικό φιλτράρισμα που βασίζεται σε χρήστες, γενικά μοντελοποιεί τον χρήστη ως φορέα αξιολόγησης στοιχείων. Καθώς γίνονται πρόσθετες παρατηρήσεις σχετικά με τις προτιμήσεις των χρηστών, τα μοντέλα των χρηστών επεκτείνονται και η πλήρης συλλογή των προτιμήσεων των χρηστών χρησιμοποιείται για τη δημιουργία συστάσεων ή την πρόβλεψη. Η προσέγγιση αυτή, επομένως, αγνοεί την έννοια των "τοποθετημένων ενεργειών" (situate dactions) (Suchman, 1987), το γεγονός ότι οι χρήστες αλληλεπιδρούν με το σύστημα μέσα σε ένα συγκεκριμένο "πλαίσιο" και ότι οι προτιμήσεις για αντικείμενα μέσα σε ένα πλαίσιο μπορεί να διαφέρουν από αυτές ενός άλλου πλαισίου.

Σε πολλούς τομείς εφαρμογής, μια ανεξάρτητη από το περιβάλλον αναπαράσταση μπορεί να χάσει την προγνωστική δύναμή της επειδή συγκεντρώνονται δυνητικά χρήσιμες πληροφορίες από πολλαπλά περιβάλλοντα. Για παράδειγμα, όταν ένας χρήστης αγοράζει βιβλία, οι προτιμήσεις που εκφράζει ο χρήστης σε ένα πλαίσιο, όπως "βιβλία για τα παιδιά μου", μπορεί να μην έχουν προγνωστική αξία όταν ο χρήστης αναζητά συστάσεις σε διαφορετικό πλαίσιο, όπως "βιβλία που σχετίζονται με την εργασία". Το ιδανικό "Σύστημα Συστάσεων με Έμφαση στο Γενικό Πλαίσιο (Context - Aware Recommender Systems - CARS) θα ήταν επομένως σε θέση να επισημάνει με αξιόπιστο τρόπο κάθε ενέργεια του χρήστη με ένα κατάλληλο περιβάλλον και να προσαρμόσει αποτελεσματικά την έξοδο του συστήματος στο χρήστη σε αυτό το συγκεκριμένο πλαίσιο.

Η έννοια του "πλαισίου" έχει μελετηθεί εκτενώς σε διάφορους τομείς της πληροφορικής, άλλα και σε άλλους κλάδους. Για παράδειγμα, οι Bazire και Brezillon (Bazire & Brezillon, 2005) εξετάζουν και συγκρίνουν περίπου 150 διαφορετικούς ορισμούς του πλαισίου από διαφορετικά πεδία και καταλήγουν στο συμπέρασμα ότι η πολύπλευρη φύση της έννοιας καθιστά δύσκολη την εξεύρεση ενός ενωτικού ορισμού: "Το πλαίσιο είναι ένα πλαίσιο για ένα

δεδομένο αντικείμενο; Είναι το σύνολο στοιχείων που έχουν κάποια επίδραση στο αντικείμενο; Είναι δυνατόν να ορίσουμε εκ των προτέρων το πλαίσιο ή να δηλώσουμε απλώς τα αποτελέσματα εκ των υστέρων; Είναι κάτι στατικό ή δυναμικό; Πλέον, ορισμένες προσεγγίσεις εμφανίζονται σήμερα μέσω της τεχνητής νοημοσύνης. Στην ψυχολογία, γενικά μελετάμε ένα άτομο που κάνει ένα έργο σε μια δεδομένη κατάσταση. Ποιο πλαίσιο έχει σημασία για τη μελέτη μας; Το πλαίσιο του ατόμου; Το πλαίσιο της αποστολής; Το πλαίσιο της αλληλεπίδρασης; Το πλαίσιο της κατάστασης; Πότε ένα πλαίσιο αρχίζει και πού σταματάει; Ποιες είναι οι πραγματικές σχέσεις μεταξύ του πλαισίου και της γνώσης; "

Παρόλο που ο Bazire και ο Brezillon (Bazire & Brezillon, 2005) δεν καταλήγουν σε έναν συγκεκριμένο ορισμό, οι ερωτήσεις που δημιουργούν διαπερνούν ουσιαστικά όλους τους τομείς στους οποίους η συνειδητοποίηση του πλαισίου είναι απαραίτητη ή επιθυμητή, συμπεριλαμβανομένων των Συστημάτων Συστάσεων. Παρατηρώντας τις διάφορες χρήσεις του περιβάλλοντος, ο Dourish (Dourish, 2004) διακρίνει μεταξύ δύο διαφορετικών απόψεων του περιβάλλοντος: την απεικόνιση και την αλληλεπίδραση. Η παραστατική άποψη κάνει τέσσερις βασικές παραδοχές: το πλαίσιο είναι μια μορφή πληροφοριών, είναι οριοθετημένη, είναι σταθερή και είναι ανεξάρτητη από την υποκείμενη δραστηριότητα. Σε αυτή την άποψη, το πλαίσιο μπορεί να περιγραφεί χρησιμοποιώντας ένα σύνολο παρατηρήσιμων χαρακτηριστικών γνωστών εκ των προτέρων. Επιπλέον, η δομή αυτών των χαρακτηριστικών περιβάλλοντος δεν αλλάζει με την πάροδο του χρόνου.

Η αλληλεπίδραση του περιβάλλοντος, σύμφωνα με το Dourish, παίρνει μια διαφορετική στάση στις βασικές παραδοχές της αντιπροσωπευτικής άποψης. Στην αλληλεπίδραση, το πεδίο των συμφραζόμενων χαρακτηριστικών ορίζεται δυναμικά, καθώς αυτό προκαλείται μάλλον, παρά στατικά. Αντί να υποθέσουμε ότι το πλαίσιο αυτό λειτουργεί ως ένα σύνολο συνθηκών κάτω από τις οποίες συμβαίνει μια δραστηριότητα, αυτή η άποψη υποθέτει μια κυκλική σχέση μεταξύ περιβάλλοντος και δραστηριότητας, όπου η δραστηριότητα δημιουργεί το πλαίσιο και το περιβάλλον επηρεάζει τη δραστηριότητα.

Ωστόσο, δεν αποτελεί στόχο η διερεύνηση της έννοιας του πλαισίου εν γένει, αλλά η εξέταση του τρόπου με τον οποίο μπορεί να οριστεί και να χρησιμοποιηθεί το πλαίσιο στα Συστήματα Συστάσεων, προκειμένου να δημιουργηθούν πιο έξυπνες και χρήσιμες προτάσεις για τους χρήστες. Στη συνέχεια του κεφαλαίου παρουσιάζεται μία προσέγγιση για τον ορισμό του πλαισίου στα Συστήματα Συστάσεων και συζητούνται διαφορετικά παραδείγματα

ενσωμάτωσής τους στη διαδικασία των προτάσεων. Επισημαίνονται επίσης αρκετές εφαρμογές όπου έχουν χρησιμοποιηθεί αποτελεσματικά τα CARS, από ταξιδιωτικούς οδηγούς μέχρι αναζήτηση πληροφοριών και μουσικές προτάσεις.

2.1. Η Έννοια του Πλαισίου στα Συστήματα Συστάσεων

Κατά την ταξινόμηση των CARS, ακολουθούνται πολλές προηγούμενες προσεγγίσεις υποθέτοντας την ύπαρξη ορισμένων παραγόντων, όπως ο χρόνος, η τοποθεσία και ο σκοπός αγοράς, που προσδιορίζουν το πλαίσιο στο οποίο παρέχονται οι προτάσεις. Ο καθένας από αυτούς τους παράγοντες πλαισίου μπορεί να έχει μια δομή. Ο παράγοντας Χρόνος, για παράδειγμα, μπορεί να οριστεί με όρους δευτερολέπτων, λεπτών, ωρών, ημερών, μηνών και ετών. Η ταξινόμηση του πλαισίου βασίζεται στις ακόλουθες δύο πτυχές των παραγόντων πλαισίου: ποιο σύμβολο συστήματος μπορεί να γνωρίζει σχετικά με αυτούς τους παράγοντες συμφραζομένων και πώς μεταβάλλονται οι παραλληλισμένοι παράγοντες με την πάροδο του χρόνου.

2.2. Οι Γνώσεις του Συστήματος Συστάσεων για τις Παραμέτρους Πλαισίου

Ένα Σύστημα Συστάσεων μπορεί να έχει διαφορετικούς τύπους γνώσεων, οι οποίοι μπορεί να περιλαμβάνουν τον ακριβή κατάλογο όλων των σχετικών παραγόντων, τη δομή τους και τις αξίες τους, σχετικά με τους παράγοντες πλαισίου. Ανάλογα με το τι ακριβώς γνωρίζει το σύστημα (δηλαδή αυτό που παρατηρείται), μπορούμε να ταξινομήσουμε τη γνώση ενός Συστήματος Συστάσεων σχετικά με τους παράγοντες πλαισίου σε τρεις κατηγορίες: πλήρως παρατηρήσιμες, εν μέρει παρατηρήσιμες και μη παρατηρήσιμες.

Πλήρως παρατηρήσιμες: Οι γνωστοί παράγοντες που σχετίζονται με την εφαρμογή, καθώς και η δομή και οι αξίες τους κατά το χρόνο υποβολής των προτάσεων, είναι σαφώς γνωστές. Για παράδειγμα, σε περίπτωση που συνιστούμε την αγορά ενός συγκεκριμένου προϊόντος, όπως ένα πουκάμισο, το σύστημά του μπορεί να γνωρίζει ότι μόνο οι χρόνοι και ο σκοπός αγορών έχουν σημασία σε αυτήν την εφαρμογή. Επιπλέον, το σύστημά του μπορεί να γνωρίζει τη δομή όλων αυτών των παραγόντων, όπως η ύπαρξη κατηγοριών εβδομάδας, Σαββατοκύριακου και αργιών για το Χρόνο. Επιπλέον, το Σύστημα Συστάσεων μπορεί επίσης να γνωρίζει τις τιμές

των παραγόντων πλαισίου κατά το χρόνο σύστασης, για παράδειγμα, όταν πραγματοποιείται η αγορά, με ποιον και για ποιον.

Εν μέρει παρατηρήσιμες: Μόνο μερικές από τις πληροφορίες σχετικά με τους παράγοντες πλαισίου, όπως περιγράφονται παραπάνω, είναι γνωστές ρητά. Για παράδειγμα, το σύστημά του μπορεί να γνωρίζει όλους τους παράγοντες πλαισίου, όπως το χρόνο και το σκοπό, αλλά όχι τη δομή τους. Ενδεχομένως να υπάρχουν διαφορετικά επίπεδα "μερικής παρατηρησιμότητας".

Μη παρατηρήσιμες: Καμία πληροφορία σχετικά με τους παραμέτρους των συμφραζομένων δεν είναι σαφώς διαθέσιμη στο σύστημα και συστήνει τις συστάσεις χρησιμοποιώντας μόνο τη λανθάνουσα γνώση του περιβάλλοντος με σιωπηρό τρόπο. Για παράδειγμα, το Σύστημα Συστάσεων μπορεί να δημιουργήσει ένα λανθασμένο προγνωστικό μοντέλο, όπως ιεραρχικά γραμμικά ή κρυμμένα μοντέλα Markov, για να υπολογίσει άγνωστες βαθμολογίες, όπου το μη παρατηρούμενο πλαίσιο διαμορφώνεται χρησιμοποιώντας λανθάνουσες μεταβλητές.

2.3. Μεταβολή Παραγόντων Πλαισίου στο Χρόνο

Ανάλογα με το εάν οι παράγοντες πλαισίου αλλάζουν με την πάροδο του χρόνου ή όχι, έχουμε τις ακόλουθες δύο κατηγορίες: στατικοί και δυναμικοί.

Στατικοί: Οι σχετικοί παράγοντες πλαισίου και η δομή τους παραμένουν οι ίδιοι (σταθεροί) με την πάροδο του χρόνου. Για παράδειγμα, σε περίπτωση που συνιστούμε την αγορά ενός συγκεκριμένου προϊόντος, όπως ένα πουκάμισο, μπορούμε να συμπεριλάβουμε τους παράγοντες πλαισίου του χρόνου, του σκοπού και μόνο αυτούς κατά τη διάρκεια ολόκληρης της διάρκειας ζωής της εφαρμογής πρότασης αγορών.

Δυναμικοί: Αυτό συμβαίνει όταν οι παράγοντες πλαισίου αλλάζουν με κάποιο τρόπο. Για παράδειγμα, το Σύστημα Συστάσεων μπορεί να συνειδητοποιήσει με την πάροδο του χρόνου ότι ο παράγοντας του σκοπού δεν έχει πλέον σημασία για την αγορά συστάσεων και μπορεί να αποφασίσει να τον αποβάλει. Επιπλέον, η δομή κάποιων από τους παράγοντες πλαισίου μπορεί να αλλάξει με την πάροδο του χρόνου, για παράδειγμα, μπορεί να προστεθούν νέες κατηγορίες στους παράγοντες πλαισίου με την πάροδο του χρόνου.

Αυτές οι δύο διαστάσεις οδηγούν στο διάγραμμα 3x2 που παρουσιάζεται στον Πίνακα 1. Τα δύο άκρα στον Πίνακα 1 είναι οι περιπτώσεις των στατικών, πλήρως παρατηρήσιμων

συμφραζομένων παραγόντων και των δυναμικών, μη παρατηρήσιμων συμφραζομένων παραγόντων. Η πρώτη περίπτωση αφορά την κατάσταση όπου είναι γνωστοί όλοι οι παράγοντες συμφραζομένων και δεν αλλάζουν με την πάροδο του χρόνου με κανέναν τρόπο. Αυτό σημαίνει ότι το σύστημα γνωρίζει τα πάντα για το περιβάλλον, κάτι που αντιστοιχεί στην περίπτωση που ονομάζεται από το Dourish (Dourish, 2004) η αντιπροσωπευτική προβολή. Το άλλο άκρο είναι η περίπτωση των δυναμικών μη παρατηρήσιμων παραγόντων πλαισίου όταν δεν είναι γνωστό τίποτα για το πλαίσιο. Μεταξύ αυτών των δύο άκρων βρίσκονται τέσσερα ενδιάμεσα κρούσματα, όπως δυναμικοί πλήρως ή μερικώς παρατηρούμενοι παράγοντες πλαισίου και στατικοί μη παρατηρήσιμοι ή εν μέρει παρατηρήσιμοι παράγοντες.

Πίνακας 1. Γνώση του Συστήματος Συστάσεων σχετικά με τους παράγοντες πλαισίου.

Τρόπος Μεταβολής των παραγόντων πλαισίου	Γνώση του Συστήματος Συστάσεων σχετικά με τους παράγοντες πλαισίου		
	Πλήρως Παρατηρήσιμη	Μερικώς Παρατηρήσιμη	Μη Παρατηρήσιμη
Στατικοί	Τα πάντα είναι γνωστά για το πλαίσιο	Μερική και στατική γνώση του πλαισίου	Υποβόσκουσα γνώση για το πλαίσιο
Δυναμικοί	Δυναμική συσχέτιση πλαισίου	Μερική και δυναμική γνώση του πλαισίου	Τίποτα δεν είναι γνωστό για το πλαίσιο

Μία από αυτές τις περιπτώσεις ορίζεται από τη στατική μη παρατηρήσιμη πληροφορία πλαισίου, όταν η δομή ενός ή περισσότερων μη παρατηρούμενων κρυμμένων παραγόντων πλαισίου δεν αλλάζει με την πάροδο του χρόνου. Δεδομένου ότι αυτή η υποβόσκουσα δομή είναι σταθερή, μπορεί να διαμορφωθεί με κρυμμένες μεταβλητές και οι μη παρατηρούμενες πληροφορίες πλαισίου μπορούν να γνωστοποιηθούν χρησιμοποιώντας μερικές από τις μεθόδους μηχανικής μάθησης, όπως η παραγοντοποίηση πίνακα (Koren, 2008) ή τα “Ιεραρχικά Γραμμικά Μοντέλα (Hierarchical Linear Models - HLMs)”. Όσον αφορά τις

δυναμικές παρατηρούμενες ή μερικώς παρατηρήσιμες περιπτώσεις στον Πίνακα 1, αυτές σχετίζονται με την αλληλεπίδραση του πλαισίου. Ως παράδειγμα της δυναμικής παρατηρήσιμης περίπτωσης, οι Mahmood, Ricci και Venturini (Mahmood et al., 2010) παρουσιάζουν ένα σύστημα προτάσεων που προσαρμόζει το διάλογο στο πλαίσιο της αλληλεπίδρασης.

Αυτό διαμορφώνεται από ένα σύνολο δυναμικών πλαισίων που αντιπροσωπεύουν, για παράδειγμα, εάν ο χρήστης παρείχε συγκεκριμένες πληροφορίες ή έδωσε συνέχεια στις προτάσεις, όπως εάν έβαλε ένα στοιχείο στο καλάθι αγορών. Με δυναμικό τρόπο, βήμα προς βήμα, το σύστημα προσαρμόζει την αλληλεπίδραση λαμβάνοντας υπόψη μια επιλογή αυτών των παραγόντων ανάλογα με την κατάσταση της αλληλεπίδρασης. Ένα παράδειγμα της στατικής μερικώς παρατηρήσιμης περίπτωσης πλαισίου παρουσιάζεται από τους Palmisano, Tuzhilin και Gorgoglione (Palmisano et al., 2008), όπου οι πληροφορίες συμφραζόμενων καθορίζονται χρησιμοποιώντας Bayesian Network (BN). Σε αυτό το BN, το παρατηρούμενο πλαίσιο αντιστοιχεί στα εξωτερικά στρώματα του δικτύου και στο μη παρατηρούμενο υποβόσκων έως τα μεσαία στρώματα του BN. Περαιτέρω, μερικές από τις μη παρατηρούμενες παραμέτρους του BN λαμβάνονται από τα δεδομένα χρησιμοποιώντας τυπικές μεθόδους μηχανικής μάθησης (Palmisano et al., 2009).

Το δυναμικό πλαίσιο μπορεί να αλλάξει βάσει παθητικών παρατηρήσεων ή ρητών ανατροφοδοτήσεων από τους χρήστες. Για παράδειγμα, η τελευταία είναι η περίπτωση των Καθομιλούμενων Συστημάτων Συστάσεων. Στα συνηθισμένα καθομιλούμενα συστήματα, χρησιμοποιείται ανατροφοδότηση από το χρήστη για την επαναληπτική επεξεργασία του προφίλ χρήστη, οδηγώντας σε πιο κατάλληλες προτάσεις. Σε ένα σύστημα συνομιλίας CARS (Baltrunas et al., 2011), η ανατροφοδότηση από τον χρήστη μπορεί επίσης να χρησιμοποιηθεί για να τροποποιήσει διαδοχικά τους σχετικούς παράγοντες και όχι μόνο τα προφίλ των χρηστών. Για παράδειγμα, κατά τη διάρκεια συνομιλίας με έναν χρήστη, ένα προτεινόμενο εστιατόριο μπορεί να καθορίσει ότι ο χρήστης είναι σε ένα ρομαντικό ραντεβού. Αυτή η παρατήρηση, με τη σειρά της, μπορεί να οδηγήσει σε φιλτράρισμα εστιατορίων που τείνουν να είναι θορυβώδη ή χωρίς επαρκή επιλογή κρασιού. Τα Καθομιλούμενα Συστήματα Συστάσεων αποτέλεσαν ένας ενεργός τομέας έρευνας. Η επαναληπτική βελτίωση του πλαισίου στα συστήματα αυτά θα εισαγάγει ένα νέο σύνολο ερευνητικών ερωτημάτων που θα απαιτήσουν περαιτέρω διερεύνηση.

Οι περισσότερες από τις εφαρμογές CARS ακολουθούν τη στατική, πλήρως παρατηρήσιμη (ή αντιπροσωπευτική) άποψη του πλαισίου (Palmisano et al., 2008). Ορισμένες αξιοσημείωτες εξαιρέσεις, πέραν αυτών που προαναφέρθηκαν, περιλαμβάνουν τις εργασίες των Jin, Zhou και Mobasher (Jin et al., 2005) και των Anand και Mobasher (Anand & Mobasher, 2007), οι οποίες εξέτασαν προσεγγίσεις στο πλαίσιο μοντελοποίησης βασισμένες στην αλληλεπίδραση, κυρίως στον τομέα του εξατομικευμένου διαδικτύου. Μία από τις κύριες προκλήσεις της αναπαραστατικής άποψης του πλαισίου είναι το γνωστό πρόβλημα τεχνικής στον τομέα που συσχετίζεται τυπικά με συστήματα βασισμένα στη γνώση. Από τη φύση τους, τα αντιπροσωπευτικά συστήματα που έχουν επίγνωση του πλαισίου, είναι ειδικά για το συγκεκριμένο τομέα και το σύνολο των μεταβλητών πλαισίου που είναι σχετικό με την υπολογιστική εργασία, όπως η δημιουργία προτάσεων, πρέπει να καθοριστούν ως μέρος του σχεδιασμού του συστήματος. Συχνά είναι δύσκολο να προσδιοριστεί εκ των προτέρων ποιοι είναι οι σχετικοί παράγοντες πλαισίου. Για παράδειγμα, σε έναν τομέα συστάσεων εστιατορίων, μπορεί να είναι δύσκολο να καθοριστεί αν η ενδυμασία του χρήστη θα ήταν σχετική με την επιλογή του εστιατορίου. Μια άλλη πρόκληση των αντιπροσωπευτικών συστημάτων είναι ότι οι πληροφορίες που σχετίζονται με συναφείς παράγοντες πλαισίου πρέπει να προσδιοριστούν ως μέρος της συλλογής δεδομένων και τέτοιες ιστορικές πληροφορίες πλαισίου συχνά δεν είναι διαθέσιμες κατά το σχεδιασμό του συστήματος. Για παράδειγμα, στους τομείς των προτάσεων ταινιών, σπάνια συμβαίνει οι χρήστες να έχουν τη δυνατότητα ή την επιθυμία να αξιολογήσουν το ίδιο στοιχείο πολλές φορές σε διαφορετικά περιβάλλοντα.

Οι προσεγγίσεις που υποθέτουν ένα μη παρατηρήσιμο ή δυναμικό πλαίσιο έχουν και τις δικές τους προκλήσεις. Αυτές οι προκλήσεις σχετίζονται κυρίως με τον προσδιορισμό της αποτελεσματικότητας των μοντέλων, όπως οι κρυφοί παράγοντες, που μπορεί να χαρακτηρίζουν διαφορετικά πλαίσια. Επιπρόσθετα, σε τέτοια συστήματα που έχουν επίγνωση του πλαισίου, οι πληροφορίες πλαισίου πρέπει να εξάγονται από τις συνεχιζόμενες δραστηριότητες των χρηστών και οι αλλαγές στο πλαίσιο πρέπει να καταγράφονται. Χωρίς ένα ρητό μοντέλο ή αναπαράσταση του πλαισίου, είναι συχνά δύσκολο ή μη πρακτικό να προσδιοριστεί τι χαρακτηρίζει ένα συγκεκριμένο πλαίσιο της δραστηριότητας ενός χρήστη. Σε πολλές εφαρμογές, ο χαρακτηρισμός των συμφραζόμενων σημείων μπορεί να πρέπει να πραγματοποιηθεί με βάση ειδικά ευρετήρια για το συγκεκριμένο τομέα. Αυτό είναι ακόμη μεγαλύτερη πρόκληση αν το πλαίσιο θεωρηθεί δυναμικό. Σε αυτή την περίπτωση, το σύστημα

πρέπει να είναι σε θέση να προσδιορίσει πότε πρέπει να "μεταβεί" σε ένα διαφορετικό υποκείμενο μοντέλο πλαισίου.

Στη συνέχεια περιγράφεται ο τρόπος με τον οποίο το πλαίσιο μπορεί να ενσωματωθεί στη διαδικασία παραγωγής των προτάσεων, εστιάζοντας κυρίως στη στατική, πλήρως παρατηρήσιμη άποψη του πλαισίου, καθώς αποτελεί ένα κυρίαρχο έως σήμερα πρότυπο CARS.

2.4. Παρουσίαση και Μοντελοποίηση του Πλαισίου

Το πρόβλημα της σύστασης αντιμετωπίζεται παραδοσιακά ως πρόβλεψη προβλήματος στην οποία, δεδομένου ενός χρήστη προφίλ και ενός στοιχείου στόχου, η λειτουργικότητα του συστήματος συνιστάται να προβλέψει την αξιολόγηση αυτού του στοιχείου από τον χρήστη, αντανακλώντας το βαθμό προτιμήσεων του χρήστη για αυτό το είδος. Συγκεκριμένα, το σύστημα προτάσεων προσπαθεί να εκτιμήσει τη λειτουργία βαθμολόγησης:

$$R: \text{Χρήστες} \times \text{Στοιχεία} \rightarrow \text{Αξιολόγηση} \quad (1)$$

που αντιστοιχεί τα ζεύγη στοιχείων χρήστη με ένα ταξινομημένο σύνολο τιμών αξιολόγησης. Σημειώνεται, ότι το R μπορεί να θεωρηθεί ως μέτρο χρησιμότητας γενικής χρήσης (ή προτίμησης) για τα ζεύγη χρηστών - στοιχείων. Η άποψη της σύστασης ως πρόβλημα πρόβλεψης προέρχεται από το γεγονός ότι το R είναι μερική συνάρτηση: οι αξιολογήσεις για όλα τα ζεύγη στοιχείων χρήστη δεν είναι γνωστές και συνεπώς πρέπει να συγκεντρωθούν.

Για το υποσύνολο ζευγαριών χρηστών-αντικειμένων για τα οποία είναι γνωστές οι τιμές αξιολόγησης, αυτές οι τιμές αξιολόγησης λαμβάνονται ρητά από τους χρήστες (για παράδειγμα, βαθμολογίες ταινιών) ή σιωπηρά (για παράδειγμα, βάσει παρατηρήσεων συμπεριφοράς χρηστών, όπως συναλλαγές αγοράς). Μόλις καθοριστεί ένα αρχικό σύνολο αξιολογήσεων, ένα Σύστημα Συστάσεων προσπαθεί να εκτιμήσει τις τιμές βαθμολόγησης για ζεύγη στοιχείων - χρήστη που δεν έχουν ακόμη βαθμολογηθεί από τους χρήστες. Αυτά τα παραδοσιακά συστήματα ονομάζονται δύο διαστάσεων (2D), δεδομένου ότι θεωρούν μόνο τις διαστάσεις χρηστών και αντικειμένων ως εισροές στη διαδικασία εξαγωγής της πρότασης. Σημειώνεται ότι η αναπαράσταση των χρηστών και των αντικειμένων στα Συστήματα Συστάσεων δεν περιορίζεται από τα αναγνωριστικά τους (όπως γίνεται από πολλούς

αλγορίθμους συνεργατικού φιλτραρίσματος) αλλά μπορεί να λάβει υπόψη τα συνολικά προφίλ χρηστών και τις λειτουργίες περιεχομένου των αντικειμένων.

Σε αντίθεση με το παραδοσιακό μοντέλο, τα CARS προσπαθούν να ενσωματώσουν ή να χρησιμοποιήσουν πρόσθετα στοιχεία (πέραν των πληροφοριών σχετικά με τους χρήστες και τα αντικείμενα) για να εκτιμήσουν τις προτιμήσεις των χρηστών σε αντικείμενα που δεν βλέπουν. Όταν τέτοιου είδους συσχετιστικά στοιχεία μπορούν να ενσωματωθούν ως μέρος της εισόδου στο Σύστημα Συστάσεων, η λειτουργία βαθμολόγησης μπορεί να θεωρηθεί ως "πολυδιάστατη":

$$R: \text{Χρήστες} \times \text{Στοιχεία} \times \text{Πλαίσια} \rightarrow \text{Αξιολογήσεις} \quad (2)$$

όπου τα Πλαίσια αντιπροσωπεύουν ένα σύνολο παραγόντων που περιγράφουν περαιτέρω τους όρους υπό τους οποίους το ζεύγος χρήστη-στοιχείου διαθέτει συγκεκριμένη βαθμολογία. Η υποκείμενη υπόθεση αυτού του εκτεταμένου μοντέλου είναι ότι οι προτιμήσεις των χρηστών για τα στοιχεία δεν είναι μόνο συνάρτηση των ίδιων των αντικειμένων, αλλά και συνάρτηση του πλαισίου στο οποίο εξετάζονται τα στοιχεία. Θα πρέπει να σημειωθεί ότι η παραπάνω χαρτογράφηση δεν είναι μοναδική στη στατική, πλήρως παρατηρήσιμη άποψη του πλαισίου, αλλά, για παράδειγμα - ακόμη και στην περίπτωση που δεν παρατηρείται - οι αξιολογήσεις μπορεί να εξαρτώνται από ένα σύνολο κρυφών παραγόντων που προέρχονται από παρατηρήσεις της δραστηριότητας του χρήστη και χρησιμοποιούνται σιωπηρά ως μέρος της εκτίμησης των προτιμήσεων των χρηστών.

Στη στατική, πλήρως παρατηρήσιμη άποψη, ωστόσο, τα Πλαίσια αντιπροσωπεύουν ένα σύνολο ρητών μεταβλητών που μοντελοποιούν τους σχετικούς παράγοντες στον υποκείμενο τομέα, όπως για παράδειγμα, το χρόνο, τη θέση, το περιβάλλον, τη συσκευή, την κατάσταση κ.λ.π. Ανεξάρτητα από την αναπαράσταση των πλαισίων, τα CARS που έχουν επίγνωση του πλαισίου πρέπει να μπορούν να αποκτούν πληροφορίες περιβάλλοντος που αντιστοιχούν στη δραστηριότητα του χρήστη, όπως για παράδειγμα, κάνοντας μια αγορά ή βαθμολογώντας ένα στοιχείο. Από τη μία πλευρά, αυτές οι πληροφορίες σχετικές με το γενικό πλαίσιο χρειάζονται δυνητικά ως μέρος της διαδικασίας μάθησης και μοντελοποίησης όπως η ανακάλυψη κανόνων ή η κατάτμηση χρηστών. Από την άλλη πλευρά, για ένα δεδομένο χρήστη-στόχο και ένα στοιχείο στόχο, το σύστημα πρέπει να είναι σε θέση να προσδιορίσει τις τιμές συγκεκριμένων παραμέτρων συμφραζομένων ως μέρος της συνεχιζόμενης αλληλεπίδρασης του χρήστη με το σύστημα.

Οι πληροφορίες πλαισίου μπορούν να ληφθούν με πολλούς διαφορετικούς τρόπους. Μπορεί να ληφθούν ρητά ή σιωπηρά. Οι ρητές πληροφορίες συμφραζομένων μπορούν να ληφθούν από τους ίδιους τους χρήστες ή από αισθητήρες που έχουν σχεδιαστεί για τη μέτρηση συγκεκριμένων φυσικών ή περιβαλλοντικών πληροφοριών. Για παράδειγμα, η αίτηση μπορεί να ζητήσει από ένα άτομο που αναζητά μια σύσταση εστιατορίου να διευκρινίσει εάν πρόκειται για ρομαντικό ραντεβού ή για ένα επαγγελματικό δείπνο. Εάν το Σύστημα Συστάσεων εστιατορίου βρίσκεται σε κινητή συσκευή, μπορείτε να λάβετε επιπλέον πληροφορίες σχετικά με τα πλαίσιο μέσω GPS ή άλλων αισθητήρων σχετικά με τη θέση, το χρόνο και τις καιρικές συνθήκες.

Σε ορισμένες περιπτώσεις, ωστόσο, οι πληροφορίες συμφραζομένων πρέπει να εξαχθούν ή να συναχθούν από άλλα παρατηρούμενα δεδομένα. Για παράδειγμα, ένα σύστημα ηλεκτρονικού εμπορίου μπορεί να επιχειρήσει, χρησιμοποιώντας επαναλαμβανόμενα μαθήματα συμπεριφοράς χρηστών, να κάνει διάκριση μεταξύ του πλαισίου στο οποίο ο χρήστης αγοράζει ένα δώρο για έναν σύζυγο και το πλαίσιο αγοράς ενός βιβλίου που σχετίζεται με την εργασία. Οι προσεγγίσεις για την υπονοούμενη συμπερίληψη των πληροφοριών πλαισίου απαιτούν τυπικά μοντέλα πρόβλεψης από ιστορικά δεδομένα (Palmisano et al.,2008).

ΚΕΦ.3: Αλγόριθμοι και Σύνολα Δεδομένων στα CARSs

3.1. Αλγόριθμοι που χρησιμοποιούνται στα CARSs

Οι παλαιότερες εργασίες πάνω στα CARS πραγματοποιήθηκαν σε ένα αλγοριθμικό επίπεδο, όπου οι ερευνητές προσπάθησαν να χρησιμοποιήσουν τις πληροφορίες πλαισίου για τη βελτίωση των υφιστάμενων αλγορίθμων (Adomavicius & Tuzhilin, 2011). Η έρευνα σχετικά με αυτό το ζήτημα έχει ως αποτέλεσμα την παραγωγή πολλών παραλλαγών των παραδοσιακών αλγορίθμων Συστημάτων Συστάσεων που χρησιμοποιούνται στα CARS. Γενικότερα, διαπιστώνεται ότι υπάρχουν ακόμη πολλά περιθώρια για τον ορισμό της έννοιας του πλαισίου (Dey, 2001) και πρέπει να υπάρξει μια αυστηρότερη έρευνα για το μοντέλο αυτών των πληροφοριών στα Συστήματα Συστάσεων. Με αυτόν τον τρόπο, ενδεχομένως να παρουσιαστεί ένα παράθυρο ευκαιρίας για τους ακαδημαϊκούς, ώστε να επιδείξουν τους καλύτερους αλγόριθμους για την κατασκευή εφαρμογών CARS.

Το γενικό πλαίσιο μπορεί να ενσωματωθεί σε διάφορα στάδια της διαδικασίας σύστασης, συμπεριλαμβανομένου του προ-φιλτραρίσματος και μετά-φιλτραρίσματος με βάση τα συμφραζόμενα και της μοντελοποίησης των συμφραζομένων (Adomavicius & Tuzhilin, 2011). Στο πλαίσιο του προ-φιλτραρίσματος και του μετά-φιλτραρίσματος με βάση τα συμφραζόμενα, οι εκτιμήσεις προβλέφθηκαν χρησιμοποιώντας οποιαδήποτε δισδιάστατη λειτουργία βαθμολόγησης. Το πλαίσιο χρησιμοποιείται εκ των προτέρων στο προ-φιλτράρισμα για την κατασκευή των αρχείων δεδομένων ή των αξιολογήσεων, ενώ στο μετά-φιλτράρισμα το πλαίσιο εφαρμόζεται στο σύνολο των συστάσεων που προκύπτουν. Η σχεδίαση με βάση τα συμφραζόμενα εφαρμόζει τις πληροφορίες πλαισίου απευθείας ως μέρος της εκτίμησης αξιολόγησης. Σε αυτή την ενότητα γίνεται μία παρουσίαση των επιλεγμένων δημοσιεύσεων ως προς τον τρόπο με τον οποίο ενσωματώνουν τις πληροφορίες συμφραζομένων και την προσέγγιση των συστάσεων που χρησιμοποιούνται στα CARS.

Η “Παραγοντοποίηση Μήτρας (Matrix Factorization - MF)” (Lee & Seung, 1999) θεωρείται ως ένας de-facto αλγόριθμος συστάσεων, ο οποίος απέκτησε την πρώτη αναγνώριση και την ευρύτερη έκθεση σε Σύστημα Συστάσεων με το βραβείο Netflix (Takács et al., 2008) για τις συστάσεις ταινιών. Όπως υποδηλώνει το όνομά του, η MF είναι ο τρόπος αποσύνθεσης μιας μήτρας σε προϊόντα δύο μητρών, έτσι ώστε η αρχική μήτρα να διατηρείται όταν πολλαπλασιάζεται. Στο Σύστημα Συστάσεων, η MF μπορεί να χρησιμοποιηθεί για την εύρεση

λανθασμένων χαρακτηριστικών που εμφανίζουν αλληλεπιδράσεις μεταξύ δύο διαφορετικών τύπων οντοτήτων (χρηστών και αντικειμένων). Με δεδομένο ένα σύνολο M χρηστών και ένα σύνολο N αντικειμένων, το R είναι μία μήτρα διαβάθμισης μεγέθους $|M| \times |N|$, μπορούμε να ανακαλύψουμε K κρυφά χαρακτηριστικά βρίσκοντας δύο μήτρες $P(|M| \times K)$ και $Q(|N| \times K)$ έτσι ώστε η προσέγγιση τους να είναι $R \approx P \times Q^T$ (Koren et al., 2009).

Σε μία από τις αρχικές προσπάθειες να περιγραφούν οι μέθοδοι MF, υπάρχουν τρία μοντέλα “Παραγοντοποίησης Μήτρας Βάσει του Πλαισίου (Context-Aware Matrix Factorization – CAMF)” που βασίζονται στην αλληλεπίδραση των στοιχείων και των παραγόντων του περιβάλλοντος. Αυτά είναι τα CAMF-C με περιεχόμενο ανεξάρτητο από το στοιχείο, τα CAMF-CI με ζεύγος συνθηκών στοιχείων – συμφραζομένων και τα CAMF-CC με μια παράμετρο μοντέλου για κάθε κατηγορία περιβάλλοντος - πλαισίου συνθηκών για την παροχή συνιστώμενων χώρων ενδιαφέροντος και μουσικής στους χρήστες (Baltrunas et al., 2011). Οι προτιμήσεις των σπουδαστών σε σχέση με τις ερωτήσεις ενσωματώθηκαν ως αντικειμενοστρεφείς όροι υπό τη μορφή μεροληψίας στοιχείων σε βασικό μοντέλο CAMF (Yang et al., 2014).

Οι βασικές μέθοδοι CAMF (Baltrunas et al., 2011) αντιπροσωπεύουν το μη πιθανολογικό γραμμικό μοντέλο που παραγοντοποιεί τη μήτρα βαθμολόγησης σε ένα προϊόν δύο μικρότερων λανθάνουσων μητρών (χρηστών και αντικειμένων) με συμφραζόμενα χαρακτηριστικά. Η “Πιθανολογική Παραγοντοποίηση Μήτρας (Probabilistic Matrix Factorization - PMF)”, από την άλλη πλευρά, περιγράφει κάθε χρήστη και στοιχείο από ένα μικρό σύνολο χαρακτηριστικών σε n -διάστατους λανθάνοντες φορείς οι οποίοι δεν είναι γνωστοί συναρτήσει του χρόνου αλλά ανακαλύπτονται αυτόματα από τον αλγόριθμο εκμάθησης (Mnih & Salakhutdinov, 2008). Η PMF είχε χρησιμοποιηθεί για την εκμάθηση ποικίλων κοινωνικών περιεχομένων ως κρυμμένοι λανθάνοντες φορείς για το πρόβλημα πρόβλεψης της αξιολόγησης της χρησιμότητας αναθεώρησης (Tang et al., 2013) πρόβλημα διαδικτυακής κοινωνικής πρότασης (Jiang et al., 2014) και συστάσεις για το σημείο ενδιαφέροντος (Ren et al., 2017). Η “Αραιή Γραμμική Μέθοδος (Sparse Linear Method - SLIM)” αποτελεί την παραδοσιακή προσέγγιση των MF για τις κορυφαίες συστάσεις, οι οποίες συσσωρεύουν αραιές αξιολογήσεις για άλλα στοιχεία που αξιολογούνται από ένα χρήστη, για να αντιμετωπίσουν την υψηλή ακεραιότητα και να μειώσουν το χρόνο μάθησης του μοντέλου (Ning & Karypis, 2011). Η SLIM είχε επεκταθεί σε CSLIM (Contextual SLIM) χρησιμοποιώντας βασικά μοντέλα CAMF για να συμπεριλάβει itemKNN (βασισμένο στο στοιχείο k -

πλησιέστερου γείτονα) και userKNN και των προτιμήσεων των χρηστών αντίστοιχα. Οι ευρετικές επιστήμες χρησιμοποιήθηκαν επίσης για την εναλλαγή μεταξύ των βασικών μοντέλων CAMF και των παραλλαγών τους (Baltrunas et al., 2011) καθώς και για την αντιμετώπιση της κατάστασης εκκίνησης με ψυχρή εκκίνηση στα CARS (Braunhofer, 2014). Τα τυπικά μοντέλα MF και CAMF συνήθως προβλέπουν για ένα μοναδικό πλέγμα χρηστών-αντικειμένων κάτω από ένα ενιαίο πλαίσιο το οποίο είχε επεκταθεί σε πολλαπλά περιβάλλοντα στα CARS (Man et al., 2015). Οι συστάσεις για τα συμβάντα αντιμετωπίστηκαν αντικαθιστώντας τη βασική βαθμολογία ενός χρήστη σε ένα γεγονός στον πρότυπο κανόνα πρόβλεψης CAMF με τη λειτουργία βαθμολόγησης ενός μοντέλου γραμμικών συμφραζόμενων χαρακτηριστικών που περιλαμβάνει μια ποικιλία συμφραζόμενων χαρακτηριστικών μεταξύ του χρήστη και του στοιχείου για ένα κοινωνικό δίκτυο (Gu et al., 2016).

Όπως παρατηρήθηκε, το CAMF βρίσκει τα λανθάνοντα πλαίσια εξετάζοντας μόνο τις αλληλεπιδράσεις μεταξύ του χρήστη και των στοιχείων - οντοτήτων. Οι αλγόριθμοι συστάσεων που έχουν σχεδιαστεί κατά κύριο λόγο για να λειτουργούν σε μήτρες δεν μπορούν να χειριστούν την τριμερή σχεσιακή φύση δεδομένων, όπως σενάρια με ετικέτες, χρόνου ή θέσης (Frolov & Oseledets, 2017). Η MF χρησιμοποιείται συχνότερα στις τεχνικές μοντελοποίησης που εξαρτώνται από τα συμφραζόμενα.

Η “Λανθάνουσα Κατανομή Ντίριχλετ (Latent Dirichlet Allocation - LDA)”, αποτελεί έναν ευρέως χρησιμοποιούμενο αλγόριθμο γενικότερα για το Σύστημα Συστάσεων, που είναι εξίσου εφαρμόσιμος για τις συμφραζόμενες προσεγγίσεις προ-φιλτραρίσματος και μετα-φιλτραρίσματος. Η LDA συνάγει δύο μήτρες διανομής από τα δεδομένα, όπου η μία μήτρα αποτελεί την κατανομή πάνω σε θέματα για ένα συγκεκριμένο έγγραφο και η άλλη ως κατανομή ανά θέμα στο λεξιλόγιο (Blei et al., 2003). Τα πειραματικά αποτελέσματα από τον Stevens (Stevens et al., 2012) έδειξαν ότι τα LDA και MF είναι παρόμοια σε μια μέση περίπτωση για να εξαγάγουν λανθάνοντα θέματα, αλλά η LDA έχει καλύτερη συνοχή θεμάτων (ικανότητα αξιολόγησης θεμάτων από μεγάλη συλλογή κειμένων) σε σύγκριση με ένα μοντέλο MF.

Το LDA είναι ένα γενετικό πιθανολογικό μοντέλο που χρησιμοποιεί Bayesian στατιστικά για να ανακαλύψει λανθάνοντα θέματα στο σώμα μέσω παρατηρήσιμων μεταβλητών των λέξεων (Blei et al., 2003). Στο αρχικό μοντέλο LDA (Blei et al., 2003), δεδομένου ενός συνόλου εγγράφων, υπάρχουν ορισμένα λανθάνοντα θέματα που δεν έχουν ακόμη παρατηρηθεί. Κάθε

έγγραφο διαθέτει διανομή σχετικά με αυτά τα θέματα και κάθε θέμα διαθέτει διανομή σχετικά με τις λέξεις στο λεξιλόγιο. Αυτές οι αναλογίες θέματος προσδιορίζονται σε μεμονωμένες λέξεις στα έγγραφα που οδηγούν σε κατανομές θεμάτων-λέξεων-κλειδιών. Επίσης, είναι άκυρες από τις προτιμήσεις των χρηστών και τα συμφραζόμενα χαρακτηριστικά στην απλοϊκή προσέγγιση. Στα μοντέλα LDA που έχουν επίγνωση του πλαισίου, οι κατανομές (αναλογίες) του πρότυπου μοντέλου LDA εμπλουτίστηκαν με διάφορους τρόπους, όπως: συμπεριλαμβάνοντας τα περιβάλλοντα που βασίζονται στην παραπομπή ως τη δεξαμενή των λέξεων σε RS (He et al., 2010), με την ενσωμάτωση των κοινωνικών πλαισίων ως μία δεξαμενή λέξεων από τα κινητά αρχεία καταγραφής των χρηστών σε ένα Σύστημα Συστάσεων που βασίζεται σε κινητά (Yu et al., 2012), με τη συμπερίληψη των προηγούμενων προτιμήσεων των χρηστών για τα αντικείμενα σε μια ερώτηση που βασίζεται στο Σύστημα Συστάσεων (Hariri et al., 2013), με την ενσωμάτωση των προσωπικών προτιμήσεων των χρηστών και των περιβαλλοντικών πλαισίων σε τηλεοπτικό περιεχόμενο Συστήματος Συστάσεων (Yuan et al., 2015), με την εξαγωγή πολλαπλών πλαισίων (tweets, emails, αναρτήσεις ή προσωπικά χαρακτηριστικά των χρηστών) από διάφορα αντικείμενα (ειδήσεις, κοινωνική δικτύωση) και ψηφιακά ίχνη (Facebook, Twitter, ηλεκτρονικά ταχυδρομεία), σε κοινωνικές βάσεις (Hsieh et al., 2016) και με την εξαγωγή των πλαισίων ως σχέσεις συσχέτισης ή συν-εμφάνισης μέσα στο στοιχείο παρουσιάζει μια βάση γνώσεων σε ένα σημασιολογικό CARS (Allahyari & Kochut, 2016).

Παρόλο που δεν είναι τόσο δημοφιλής όσο οι προαναφερόμενες τεχνικές στα CARS, οι αλγόριθμοι Bandit μπορούν να βοηθήσουν στη βελτιστοποίηση των αποτελεσμάτων όταν δεν υπάρχουν επαρκή δεδομένα για να δημιουργηθεί ένα αυστηρό στατιστικό μοντέλο. Το πρόβλημα Bandit είναι ένα πρόβλημα στατιστικής απόφασης όπου ένας εκπρόσωπος αποσκοπεί στη βελτιστοποίηση των αποφάσεών του βελτιώνοντας παράλληλα τις γνώσεις του (Robbins, 1952) (Berry, 2012). Ο αλγόριθμος Bandit αποτελεί ένα πλαίσιο για τη μεγιστοποίηση της συμμετοχής των συνολικών χρηστών, με την εξισορρόπηση του συνδυασμού μεταξύ εκμετάλλευσης (εμφάνιση καλύτερων στοιχείων) και εξερεύνησης (εμφάνιση νέων στοιχείων) (Bubeck & Cesa-Bianchi, 2012). Το δίλημμα εξερεύνησης / εκμετάλλευσης έγκειται στο γεγονός ότι πρέπει να γίνει χρήση της εξερεύνησης για την απόκτηση νέων γνώσεων σχετικά με ένα περιβάλλον, ενώ η χρήση της εκμετάλλευσης βοηθάει στη λήψη αποφάσεων ή στη δημιουργία προβλέψεων, χρησιμοποιώντας τις υπάρχουσες γνώσεις (Bubeck & Cesa-Bianchi, 2012). Αυτό το δίλημμα εξερεύνησης-εκμετάλλευσης μπορεί

να σχετίζεται στενά με ένα Σύστημα Συστάσεων το οποίο πρέπει να δίνει έμφαση στις υπάρχουσες πληροφορίες για να αυξήσει το ενδιαφέρον των χρηστών, ενώ ταυτόχρονα ανακαλύπτει νέες πληροφορίες για τη βελτίωση της εμπειρίας των χρηστών. Τα περισσότερα από αυτά τα σενάρια βασίζονται στην υπολογιστική συμπεριφορά των χρηστών και στο περιβάλλον που δρουν, επομένως μπορεί να αντιμετωπιστεί καλύτερα από τα CARS. Στα CARS η εξερεύνηση είναι χρήσιμη για τη διατύπωση συστάσεων με βάση τη συνεχιζόμενη κατάσταση, η οποία μερικές φορές μπορεί να μην φέρει τα αναμενόμενα αποτελέσματα, αλλά μακροπρόθεσμα, η εκμετάλλευση συγκεντρώνει περισσότερα οφέλη για την παραγωγή ικανοποιητικών αποτελεσμάτων. Ο Multi-Armed Bandit (MAB), συμπεριλαμβανομένων των LinucB, ε-greedy, και Bernoulli & Thomson Sampling αποτελούν σημαντικούς αλγόριθμους Bandit (Bubeck & Cesa-Bianchi, 2012) που χρησιμοποιούνται ως επί το πλείστον σε σενάρια συστάσεων. Στον αλγόριθμο MAB για ένα CARS, ένας εκπρόσωπος επιλέγει μια ενέργεια σε κάθε στάδιο και λαμβάνει μια ανταμοιβή από αυτό, ενώ ο υπεύθυνος λήψης αποφάσεων αποφασίζει ποιο σκέλος θα χρησιμοποιήσει προκειμένου να έχει τη μέγιστη συνολική ανταμοιβή. Μερικοί κλασσικοί αλγόριθμοι που χρησιμοποιούνται στα CARS περιλαμβάνουν τον αλγόριθμο PageRank της Google (El Helou et al., 2010), τα δίκτυα Bayesian (Ono et al., 2009), (Park et al., 2006), και το Support Vector Machine (SVM) (Oku et al., 2006).

3.2. Σύνολα Δεδομένων που Χρησιμοποιούνται στα CARSs

Στην εποχή των Big Data, το Σύστημα Συστάσεων αποτελεί ένα από τα πιο περιζήτητα ερευνητικά θέματα που εφαρμόζονται τόσο από ακαδημαϊκούς όσο και για την επίλυση προβλημάτων στην πραγματική ζωή. Τα σύνολα δεδομένων συγκριτικής αξιολόγησης για τα παραδοσιακά Συστήματα Συστάσεων, κατηγοριοποιούνται σύμφωνα με τους τομείς εφαρμογής τους. Έτσι, υπάρχουν διαθέσιμα τα Συστήματα Συστάσεων για ταινίες όπως Movielens, Yahoo, Prize Netflix και Movie Tweetings, τα Συστήματα Συστάσεων για μουσική όπως Last.fm, Yahoo music, Amazon audio και Audioscrobbler, τα Συστήματα Συστάσεων για εφαρμογές κινητών συσκευών όπως Mobile App Retrieval, frappe, τα Συστήματα Συστάσεων για βιβλία όπως Book-crossing, Amazon books dataset, τα Συστήματα Συστάσεων για φαγητό όπως Chicago Entree, Yelp, τα Συστήματα Συστάσεων για αγορές εμπορευμάτων όπως Amazon, τα Συστήματα Συστάσεων για την υγεία όπως Hospital Ratings, Texas healthcare information, NHS diabetes και τα Συστήματα Συστάσεων για ακαδημαϊκά άρθρα όπως CiteULike, Bibsonomy, CiteSeer, ACM, DBLP.

Ομοίως, τα συστήματα CARS εφαρμόζονται τόσο σε εφαρμογές έρευνας όσο και σε πραγματικό χρόνο σε διαφορετικά σύνολα δεδομένων και εφαρμογών όπως τα κοινωνικά δίκτυα με βάση τα γεγονότα (Macedo et al., September), οι εφαρμογές για κινητά, η ηλεκτρονική μάθηση, το σύστημα επισήμανσης (Frolov & Oseledets, 2017). Κάθε ένα από τα συστήματα αυτά έχει διαφορετικό στόχο, διαφορετικό κοινό και κατά συνέπεια οδηγεί σε διαφορετικές απαιτήσεις για τη χρήση του Συστήματος Συστάσεων. Όπως έχουν συζητηθεί από τους Adomavicius και Tuzhilin (Adomavicius & Tuzhilin, 2011), οι τομείς εφαρμογής για το CARS είναι περιορισμένοι. Αυτό ισχύει ιδιαίτερα επειδή υπάρχει έλλειψη διαθέσιμων στο κοινό δεδομένων για το CARS (Yujie & Licai, 2010). Μεγάλες εταιρείες όπως το Facebook, το Google, το Youtube και το Bing επενδύουν σε μεγάλο βαθμό στις υπηρεσίες συστάσεων, αλλά έχουν αποστείλει ελάχιστες ή καθόλου πληροφορίες σχετικά με την εφαρμογή των υπηρεσιών σύστασης που έχουν αναπτύξει. Ως εκ τούτου, η ερευνητική κοινότητα του Συστήματος Συστάσεων δεν μπορεί να επωφεληθεί πλήρως από αυτές.

Υπάρχουν τρία είδη συνόλων δεδομένων στον τομέα των CARS, τα σημεία αναφοράς πραγματικού κόσμου, τα ημισυνθετικά και τα πλήρως συνθετικά σύνολα δεδομένων. Μεγάλο μέρος του πειραματισμού και της αξιολόγησης των CARS είναι διαπιστευμένο στα σύνολα δεδομένων σημείων αναφοράς που χρησιμοποιούνται στη γενική ρύθμιση του Συστήματος Συστάσεων. Όμως, αυτά τα δημόσια σύνολα δεδομένων περιέχουν πολύ περιορισμένες αξιολογήσεις σχετικών με το περιβάλλον και δεν επαρκούν για την πολύ καλή αξιολόγηση των αλγορίθμων CARS. Μια πιθανή λύση αποτελεί η ανίχνευση τέτοιων δεδομένων από τους γνωστούς ιστότοπους που υποστηρίζουν κάθε έναν ξεχωριστό τομέα εφαρμογής. Αλλά αυτή είναι μια πολύ χρονοβόρα και περίπλοκη διαδικασία καθώς τίθενται ζητήματα πνευματικής ιδιοκτησίας από τις εταιρείες ή ζητήματα προστασίας της ιδιωτικής ζωής από τους χρήστες. Μια άλλη πρακτική που ακολουθείται στην έρευνα των CARS είναι η δημιουργία από πλευράς ερευνητών και προγραμματιστών εφαρμογών, των δικών τους συνθετικών ή ημισυνθετικών συνόλων δεδομένων για τα CARS, προσομοιάζοντας τα συμφραζόμενα χαρακτηριστικά που θέλουν ώστε στη συνέχεια να αξιολογήσουν τους αλγορίθμους τους. Εκτός από αυτές τις επιλογές, υπάρχουν ελάχιστα διαθέσιμα δεδομένα CARS προς το κοινό. Για παράδειγμα, το CARSKit3 (Zheng et al., 2015) είναι ένας μηχανισμός CARS ανοιχτού κώδικα σε γλώσσα Java που υλοποιεί αλγορίθμους και παρέχει σύνολα δεδομένων και μετρήσεις αξιολόγησης ειδικά για το CARS, οι οποίες είναι ελεύθερα χρησιμοποιήσιμες, τροποποιήσιμες και διανεμητές υπό

την GNU General Public License.⁴ Ωστόσο, ο αριθμός των διαθέσιμων συνόλων δεδομένων μέσω του CARSKit είναι πολύ περιορισμένος.

Οι ταινίες webscope του Yahoo και το MovieAT είναι δύο πραγματικές πλατφόρμες που έχουν αναγεννηθεί για να δημιουργήσουν ημιτεχνητά δεδομένα σε βασικές μεθόδους CAMF. Ένα σύνολο δεδομένων κινηματογραφικών ταινιών που αποτελείται από δεδομένα που σχετίζονται με το χρόνο, τον τόπο και το είδος είχε δημιουργηθεί αρχικά (Adomavicius et al., 2005) ώστε ένα Σύστημα Συστάσεων ταινιών να μπορέσει να χρησιμοποιηθεί σε διάφορες προσεγγίσεις CARS (Baltrunas et al., 2011). Συγκεντρωτικά σύνολα δεδομένων ταινιών δημιουργήθηκαν επίσης από τα σύνολα δεδομένων των Moviepilot και Filmtipset (Braunhofer, 2014) με την ενσωμάτωση δυνητικών τεμαχίων πληροφοριών που σχετίζονται με τα δεδομένα. Το σύνολο δεδομένων Benchmark δημιουργήθηκε συνθετικά για την προσθήκη πλαισίων με τη μορφή χρονικών (Liu et al., 2015) και επίκαιρων πληροφοριών (Allahyari & Kochut, 2016). Το σύνολο δεδομένων Lastfm αναδιαμορφώθηκε με την ενσωμάτωση χαρακτηριστικών χρονοσήμανσης (Wu et al., 2016). Ομοίως, τα σύνολα δεδομένων LastFM, MovieLens, Netflix και VOD διαμορφώθηκαν για να συμπεριλάβουν την αλληλουχία και την εποχή ως συσχετιζόμενες πληροφορίες. Τα αρχεία δεδομένων μουσικής Yahoo και Yelp αναδιαμορφώθηκαν επίσης για να συμπεριλάβουν τις αξιολογήσεις των χρηστών κάτω από συγκεκριμένα πλαίσια για τη διαχείριση των συστάσεων μουσικής (Yuan et al., 2016). Ένα σύνολο δεδομένων ανασκόπησης προϊόντος δημιουργήθηκε από την ανίχνευση ενός ιστοτόπου ανασκόπησης προϊόντων Ciao με τους συγγραφείς, τους κριτές και τις μεταξύ τους σχέσεις να αποτελούν τις συσχετιζόμενες πληροφορίες. Η άδεια για διεξαγωγή έρευνας σχετικά με τα σύνολα δεδομένων Coursera MOOC με συγκεκριμένα χαρακτηριστικά συμφραζομένων λήφθηκε για το πρόβλημα της σύστασης ερωτήσεων (Yang et al., 2014).

Τα σύνολα δεδομένων που σχετίζονται με τα τρόφιμα (Ono et al., 2009) χρησιμοποιήθηκαν για την αξιολόγηση των κορυφαίων N συστάσεων στα CARS (Shi et al., 2014) και για την καταγραφή της σημασιολογικής σχέσης μεταξύ των περιβαλλόντων και των οντοτήτων (Liu et al., 2015). Το σύνολο δεδομένων των εστιατορίων δημιουργήθηκε αποκλειστικά για το CARS μέσω μελετών που διενέργησαν οι χρήστες. Η αξιολόγηση των χρηστών με τις συσχετιζόμενες επιπτώσεις στις προτιμήσεις της μουσικής είχε καταγραφεί μέσω εφαρμογών ιστού για τη δημιουργία του συνόλου δεδομένων InCarMusic (Baltrunas et al., 2011) που χρησιμοποιήθηκε για τις N κορυφαίες συστάσεις (Zheng et al., 2014) εντός των MF μεθόδων (Karatzoglou et al., 2010) στο CARS. Ένα άλλο σύνολο δεδομένων με βάση τα περιεχόμενα δημιουργήθηκε από

την εφαρμογή Appzaar για το πρόβλημα της πρότασης των κορυφαίων N επιλογών (Shi et al., 2012).

Τα σύνολα δεδομένων κοινωνικής δικτύωσης δημιουργήθηκαν από την ανίχνευση ιστότοπων Renren (www.renren.com) και Tencent Weibo (t.qq.com) και από το κοινωνικό δίκτυο Glue (Wermser et al., 2011) με την ενσωμάτωση προσωπικών και διαπροσωπικών συσχετιζόμενων παραγόντων για το πρόβλημα των κοινωνικών συστάσεων. Το Tuenti (<https://www.tuenti.com/>), ένας ιστότοπος κοινωνικής δικτύωσης είχε ανιχνευθεί για τη λήψη των κλικ του χρήστη-βίντεο ώστε να χρησιμοποιηθούν στις προτάσεις των κορυφαίων N επιλογών (Shi et al., 2014). Το Medium.com που χρησιμοποιήθηκε ευρέως σε τυπικά Συστήματα Συστάσεων είχε χρησιμοποιηθεί για προβλήματα CARS κοινωνικών συμβάντων-ενεργειών.

Τα σημεία ενδιαφέροντος και τα συσχετιζόμενα χαρακτηριστικά περιεχομένου εξήχθησαν μέσω μιας εφαρμογής με τη βοήθεια χρηστών που συμμετείχαν στο πείραμα (Baltrunas et al., 2011 ; Elahi et al., 2013). Τα σύνολα δεδομένων για τη χρήση εφαρμογών κινητής τηλεφωνίας δημιουργήθηκαν με πληροφορίες σχετικά με τα περιεχόμενα που συλλέχθηκαν από την εφαρμογή Frappe (Shi et al., 2014) και μέσω μελετών που διενέργησαν χρήστες (Yu et al., 2012). Το σύνολο δεδομένων του Twitter με ετικέτες με γεωγραφικά χαρακτηριστικά και τα ψηφιακά ίχνη των χρηστών είχαν συγκεντρωθεί για CARS. Το CiteSeerX είχε ανιχνευθεί για να δημιουργήσει ένα σύνολο δεδομένων CARS για ένα σύστημα βασισμένο σε παραπομπές. Ένα σύνολο δεδομένων σχετικών με το αρχείο καταγραφής των κλικ των χρηστών δημιουργήθηκε από την ανίχνευση ενός ιστότοπου πωλήσεων προϊόντων μόδας σε απευθείας σύνδεση, προσθέτοντας δημογραφικά στοιχεία και προσωπικά χαρακτηριστικά συμφραζόμενων σε αυτό (Tavakol & Brefeld, 2014). Η βαθμολογία χρήστη του Yahoo και ένα σύνολο δεδομένων πλοήγησης για το διαδίκτυο έχουν συγκεντρωθεί για διάφορα συσχετιζόμενα χαρακτηριστικά για ένα διαδραστικό σύστημα συστάσεων (Hariri et al., 2014).

Τα συστήματα CARS για ειδήσεις χρησιμοποιούν συνήθως τυπικά σύνολα δεδομένων βασισμένα σε Συστήματα Συστάσεων για την παραγωγή ημι-τεχνητών συνθετικών δεδομένων. Το σύνολο δεδομένων της Yahoo! Today (Tang et al., 2014), το KDD Cup 2012 Online Advertising (Elahi et al., 2013) και τα New York Times Annotated Corpus αναδιαμορφώθηκαν ώστε να αποκτήσουν συσχετιζόμενες ροές κλικ για το πρόβλημα σύστασης των ειδήσεων. Παρόλο που παρατηρείται κάποια εξέλιξη σε σύνολα δεδομένων ειδικά για το CARS,

ταυτόχρονα μπορεί να φανεί ότι αυτά τα συνθετικά σύνολα δεδομένων συνήθως δεν είναι διαθέσιμα στο κοινό. Υπάρχουν ορισμένα σύνολα δεδομένων CARS διαθέσιμα στο CARS Kit5 σύμφωνα με τους όρους της Γενικής Δημόσιας Άδειας GNU όπως InCarMusic (Baltrunas et al., 2011), Foods (Ono et al., 2009), Restaurant (Oku et al., 2006), Frappe (Baltrunas et al., 2015), TripAdvisor (Zheng et al., 2012) και Adoms (Adomavicius & Tuzhilin, 2011) τα οποία αναφέρονται επίσης σε πολλές εκδόσεις του CARS.

Στον Πίνακα 2, παρατίθενται σύνολα δεδομένων (γενικά και συγκεκριμένου πλαισίου) που χρησιμοποιούνται στα Συστήματα Συστάσεων και στα CARS. Όπως φαίνεται στον Πίνακα 2, τα σύνολα δεδομένων για το γενικό Σύστημα Συστάσεων είναι ως επί το πλείστον ανοιχτού κώδικα και είναι διαθέσιμα στο κοινό. Ωστόσο, τα σύνολα δεδομένων εννοιολογικού πλαισίου είναι κυρίως ιδιοκτησιακά (ιδιωτικά) και συνήθως δεν είναι διαθέσιμα στο ευρύ κοινό. Συνήθως δημιουργούνται για τον τομέα για τον οποίο πρόκειται να αναπτυχθεί ένα σύστημα CARS όπως τα συστήματα που βασίζονται στα “Σημεία Ενδιαφέροντος (Points of Interest - POIs)” ή τα συστήματα που βασίζονται σε κινητά.

Πίνακας 2. Σύνολα δεδομένων που χρησιμοποιούνται γενικότερα στα Συστήματα Συστάσεων και στα CARS.

Γενικής Χρήσης	Εννοιολογικού Πλαισίου
MovieLens, Netflix, Meetup, Grocery, LastFM, VOD, Medium.com, Twitter, Wikipedia dumps, facebook, Playlists, CiteULike, Yahoo Yelp, New York Times, KDD Cup 2012, Delicios.	Tourism, InCarMusic, Appazaar.net, Product review, Foursquare, Foods, Restaurant, Tencent Weibo.com, Renren.com, MOOC, Adom (movies), Frappe, getGlue, CiteSeerX, mobile logs, Zalando click logs, TripAdvisor, Adom, Frappe.

3.3. Προοπτικές για τη Βελτίωση της Λειτουργίας των CARSs

Τα CARS έχουν τεράστιο αντίκτυπο στον κλάδο της έρευνας και της ανάπτυξης, καθώς διαδραματίζουν ένα σημαντικό ρόλο στην πρόβλεψη και στις προσδοκίες των αναγκών των χρηστών ως απάντηση στη συμπεριφορά τους σε σενάρια από τον πραγματικό κόσμο. Με την δημιουργία του πλαισίου στα Συστήματα Συστάσεων, πολλοί ερευνητές άρχισαν να

βελτιώνουν τους κλασσικούς αλγόριθμους σύστασης, ενσωματώνοντας διάφορα είδη πλαισίων στη διαδικασία των συστάσεων. Αυτό έχει οδηγήσει σε μια σημαντική μεταβολή των παραδειγμάτων στον τομέα των Συστημάτων Συστάσεων από το να είναι γενική προς την εστίαση στις συνθήκες και στο γενικό πλαίσιο (Adomavicius & Tuzhilin, 2011). Παρ'όλα αυτά, η αναγνώριση του σωστού αλγορίθμου για τον τύπο του τομέα εφαρμογής για τον οποίο πρόκειται να κατασκευαστεί το Σύστημα Συστάσεων παραμένει μια πρόκληση για τους προγραμματιστές εφαρμογών (Champiri et al., 2015). Στις προηγούμενες ενότητες αναδείχθηκαν οι πληροφορίες σχετικά με το πώς οι αλγόριθμοι που χρησιμοποιούνται στα CARS διαφέρουν από αυτούς που χρησιμοποιούνται στα παραδοσιακά Συστήματα Συστάσεων. Επιπλέον, αναλύθηκε η εξέλιξη των αλγορίθμων που χρησιμοποιούνται στα CARS κατά τη διάρκεια του χρόνου. Εκτός αυτού, τονίστηκε επίσης ο τρόπος ενσωμάτωσης σχετικών πλαισίων από διάφορες πηγές σε ένα μοντέλο συστάσεων που απαιτεί τη διαμόρφωση των σωστών μέτρων απόδοσης για τη δοκιμή τέτοιων περιπτώσεων χρήσης.

Οι αλγόριθμοι τελευταίας τεχνολογίας στα CARS προέρχονται κυρίως από τους παραδοσιακούς αλγορίθμους συστάσεων. Εξελίσσονται με την πάροδο του χρόνου και μπορούν να υλοποιηθούν σε διαφορετικούς τομείς εφαρμογών. Εντούτοις, θα μπορούσε να διευκολύνει σημαντικά το έργο των συστάσεων εννοιολογικού πλαισίου εάν υπήρχε ένα ενιαίο και τυποποιημένο πλαίσιο στα CARS που να είναι αρκετά ευέλικτο ώστε να φιλοξενήσει διαφοροποιημένα περιβάλλοντα και τομείς εφαρμογής. Με αυτόν τον τρόπο, η ερευνητική κοινότητα μπορεί να συμβάλει στη βελτίωση αυτού του ενιαίου πλαισίου αντί να προτείνει διαφορετικές παραλλαγές που να ταιριάζουν σε κάθε συγκεκριμένο τομέα εφαρμογής.

Το ζήτημα των διαστάσεων είναι ένα εγγενές πρόβλημα στο CARS που εμποδίζει την ερευνητική κοινότητα να εργαστεί σημαντικά στον τομέα. Οι γνωστές τεχνικές μείωσης των διαστάσεων που χρησιμοποιούνται στο CARS περιλαμβάνουν μεθόδους παραγοντοποίησης, PCA, LSI και το πιο γενετικό μοντέλο LDA που αρχικά προέρχεται από το SVD με ελαφρούς μετασχηματισμούς στην ελαχιστοποίηση του λιγότερο τετραγωνισμένου σφάλματος έναντι της εισόδου (Udell et al., 2016). Εκτός από αυτές τις τεχνικές, συναντάται επίσης η χρήση ευρετικών ώστε να αντιμετωπιστεί ο αυξανόμενος αριθμός διαστάσεων στο CARS. Ωστόσο, οι μέθοδοι παραγοντοποίησης λαμβάνουν μεγαλύτερες εκθέσεις σε σύγκριση με παρόμοιες τεχνικές στα ίδια σύνολα δεδομένων, κυρίως για λανθάνουσα μεταβλητή αποσύνθεση. Η μείωση των διαστάσεων συνεπάγεται ένα κόστος, συνεπώς, πρέπει να υπάρχει

δικαιολογημένη ισορροπία μεταξύ του μικρότερου αριθμού διαστάσεων και της διατήρησης των αναγκαίων πληροφοριών.

Οι εφαρμογές CARS που αναπτύχθηκαν για να χειριστούν την οπτική απεικόνιση στο πλαίσιο των προφίλ μάθησης των σπουδαστών (Adomavicius et al., 2005) παραμελούν τη δυναμική των πλαισίων. Παρόλο που υπάρχουν ορισμένοι αλγόριθμοι όπως ο αλγόριθμος Bandit και η διαδικασία Markov που δείχνουν μία πληρέστερη τάση προς την αλληλεπίδραση με το περιβάλλον, που ωστόσο, σπάνια χρησιμοποιούνται σε καθαρή εφαρμογή CARS. Τα μοντέλα χρηστών που βασίζονται σε μερικές πληροφορίες για τους χρήστες και τα συμφραζόμενα οδηγούν σε κακή απόδοση του CARS. Αυτό συνεπάγεται την ανάγκη διερεύνησης πιθανών μηχανισμών για την ταυτοποίηση συσχετιζόμενων μεταβολών και την προσαρμογή τους στα πρότυπα υποδείξεων κατά τη διάρκεια του χρόνου εκτέλεσης. Για να ανταπεξέλθει κανείς με την έρευνα διαχείρισης περιβάλλοντος, είναι πολύ σημαντικό να συγκεντρωθούν ετερογενή περιβάλλοντα από το περιβάλλον και να υποστηριχθούν κανόνες συλλογισμού για να συμπεράνουμε τα γεγονότα του πλαισίου από τα ανεπεξέργαστα δεδομένα. Αυτό απαιτεί τη δημιουργία ενός προ-υλοποιούμενου πλαισίου μοντέλου χρήστη για το CARS (Braunhofer, 2014) που μπορεί να χρησιμεύσει ως πλαίσιο αναφοράς για τη δημιουργία πιο εξειδικευμένων μοντέλων.

Φαίνεται ότι υπάρχει επείγουσα ανάγκη να δημιουργηθούν σύνολα δεδομένων ειδικά για το CARS. Παρόλο που υπάρχουν μερικά εκτός από τα σύνολα δεδομένων αναφοράς που χρησιμοποιούνται στη γενική ρύθμιση του Συστήματος Συστάσεων, τα περισσότερα από αυτά τα σύνολα δεδομένων CARS δεν είναι διαθέσιμα στο κοινό, ίσως λόγω των θεμάτων ιδιωτικού απορρήτου και πνευματικών δικαιωμάτων. Η παροχή κατάλληλων συνόλων δεδομένων στους ακαδημαϊκούς μπορεί να μετακινήσει την περιοχή CARS προς εφαρμογές πραγματικού κόσμου. Μια άλλη κοινή πρακτική που παρατηρείται στο CARS είναι ότι δοκιμάζεται η γενικότερη αποτελεσματικότητα των εφαρμογών τους χρησιμοποιώντας έναν παραδοσιακό μηχανισμό αξιολόγησης χωρίς σύνδεση. Τέτοια μέτρα μπορεί να παρέχουν εύχρηστα αποτελέσματα, αλλά αργότερα θα μπορούσαν να αντικρουστούν τα αποτελέσματα που προκύπτουν από τις ρυθμίσεις στο διαδίκτυο. Επομένως, η ανάπτυξη κατάλληλων στρατηγικών αξιολόγησης είναι πολύ σημαντική για την προώθηση της τεχνολογίας των CARS.

Από τη μελέτη αυτών των δημοσιεύσεων βρέθηκε ότι τα περιβάλλοντα συνήθως κωδικοποιούνται στη διαδικασία των συστάσεων χωρίς καμία επίσημη έρευνα. Οι

συναισθηματικές καταστάσεις των χρηστών όπως η πείνα, η συντροφικότητα εξετάστηκαν στο πλαίσιο του μοντέλου (Karatzoglou et al., 2010 ; Liu et al., 2015), αλλά υπάρχουν πολύ περιορισμένα παραδείγματα τέτοιων εργασιών. Ομοίως, υπάρχουν πολύ λίγα παραδείγματα (Allahyari & Kochut, 2016) που χρησιμοποιούν τις βάσεις γνώσης για να συμπεράνουν πλουσιότερα πλαίσια. Η ενσωμάτωση άλλων τεχνολογιών όπως η ανάλυση συναισθημάτων, η οντολογία, η ομαδοποίηση ή η εξόρυξη κανόνων σύνδεσης στην έρευνα CARS μπορεί να βελτιώσει σημαντικά τις συστάσεις που αφορούν τις πρακτικές σεναρίων.

ΚΕΦ.4: Η Μηχανή Ανοιχτού Κώδικα CARSKit

Το σύστημα παροχής συστάσεων έχει καταδειχθεί ως ένα από τα πιο χρήσιμα εργαλεία για να βοηθηθούν οι αποφάσεις των χρηστών. Διάφοροι αλγόριθμοι συστάσεων έχουν αναπτυχθεί και εφαρμοστεί τόσο από τις εμπορικές βιβλιοθήκες προγραμματισμού όσο και από τις βιβλιοθήκες συστάσεων ανοικτού κώδικα. Το από τα συμφραζόμενα ενήμερο σύστημα παροχής συστάσεων (CARS) προέκυψε ως μία νέα ερευνητική κατεύθυνση κατά την τελευταία δεκαετία και έχουν προταθεί πολλοί συναφείς αλγόριθμοι παροχής συστάσεων

Δυστυχώς, κανένας μηχανισμός συστάσεων δεν άρχισε να ενσωματώνει αυτούς τους αλγορίθμους χαρακτηριστικά τους, λόγω των ιδιαίτερων χαρακτηριστικών της μορφής των δεδομένων και των μεθόδων επεξεργασίας στον τομέα του CARS. Στην εργασία των Zheng και των συνεργατών του, (Zheng et. al., 2015), εισάγεται μια μηχανή σύστασης ανοικτού κώδικα βασισμένη σε Java που λαμβάνει γνώση των συμφραζόμενων, με την ονομασία CARS Kit, η οποία αναγνωρίζεται ως η πρώτη βιβλιοθήκη παροχής συστάσεων ανοικτού πηγαίου κώδικα που είναι ειδικά σχεδιασμένη για το CARS. Εφαρμόζει τους πιο σύγχρονους αλγορίθμους συστάσεων που είναι σχετικοί με το περιεχόμενο και στα επόμενα θα παρουσιαστεί η ευκολία με την οποία το CARS Kit επιτρέπει στις μηχανές παροχής συστάσεων να διαμορφώνονται και να αξιολογούνται.

4.1. Παλαιότεροι Αλγόριθμοι και Μηχανισμοί Συστάσεων

Υπάρχουν τουλάχιστον τρεις λόγοι για την ανάπτυξη μιας μηχανής συστάσεων: 1) οι κλασικοί αλγόριθμοι συστάσεων είναι οι συχνότεροι που πρέπει να αξιολογηθούν στην πράξη. Μια βιβλιοθήκη συστάσεων είναι σε θέση να παρέχει τυποποιημένες εφαρμογές και να αποφεύγει επανειλημμένα την κατάρτιση αυτών των αλγορίθμων από καιρό σε καιρό. 2) Αποτελεί επίσης μια τυποποιημένη πλατφόρμα για δείκτες αναφοράς ή αξιολογήσεις. 3) Είναι επίσης χρήσιμη τόσο για σκοπούς έρευνας όσο και για βιομηχανικές πρακτικές, καθώς και αποτελεί εκπαιδευτικό εργαλείο στη διδασκαλία και τη μάθηση.

Κατά τη διάρκεια των τελευταίων δεκαετιών, έχουν προταθεί πάρα πολλοί αλγόριθμοι συστάσεων και αρκετοί μηχανισμοί συστάσεων, όπως οι Mahout¹, Duine², Cofi³, EasyRec⁴,

¹Mahout, <http://mahout.apache.org/>

²Duine, <http://www.duineframework.org/>

³Cofi, <http://www.nongnu.org/cofi/>

GraphLab To Create⁵, LensKit⁶, LibRec⁷, MyMediaLite⁸, άρχισαν να ενσωματώνουν αυτούς τους αλγορίθμους στα χαρακτηριστικά τους. Στην εργασία των Zheng και των συνεργατών του, (Zheng et. al., 2015), επιλέχθηκαν οι πιο δημοφιλείς και παρήχθη μια σύντομη σύγκριση που παρουσιάζεται στον Πίνακα 3.

Πίνακας 3. Σύντομη σύγκριση των πιο δημοφιλών αλγορίθμων συστάσεων και μηχανισμών συστάσεων (Zheng et. al., 2015).

	Mahout	GraphLab Create	MyMediaLite	EasyRec	LensKit	LibRec	CARSKit
Initial release (date)	04/2007	2014	10/2010	03/2011	03/2011	2014	09/2015
Latest release (date)	05/2015	2015	09/2013	05/2013	05/2015	06/2015	09/2015
Latest version	0.10.1	1.4.1	3.10	0.98	2.2	1.3.2	0.1.0
Active updates	Yes	Yes	No	No	Yes	Yes	Yes
License	Apache	BSD, AGPL	GPL	N/A	GPL	GPL	GPL
Programming language	Java	Python, C++	C#	WebService	Java	Java	Java
Running platform	JVM	Any	.NET	WebApp	JVM	JVM	JVM
Command line	Yes	Yes	Yes	No	Yes	Yes	Yes
Documentation	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Online updates	Yes	Yes	Yes	No	No	No	No
Algorithm diversity	Classical	Classical	State-of-the-art	Classical	Classical	State-of-the-art	CARS
Distributed computing	Partial	Yes	No	No	No	No	No
Open source	Yes	Partial	Yes	Yes	Yes	Yes	Yes

Οι Mahout και GraphLab Create είναι δύο μεγάλες πλατφόρμες εξαγωγής δεδομένων και μηχανικής μάθησης, όπου το GraphLab ήταν μια βιβλιοθήκη ανοιχτού κώδικα, που πρόσφατα αναπτύχθηκε ως ένα από τα εμπορικά προϊόντα του DATO⁹. Το MyMediaLite ήταν μία από τις πιο δημοφιλείς βιβλιοθήκες συστάσεων αλλά δεν ενημερώθηκε για κάποιο διάστημα ενώ τα Rival¹⁰ και WrapRec¹¹ σχεδιάστηκαν πρόσφατα ως «περιτύλιγμα» του MyMediteLite για να παρέχουν ένα πρότυπο δείκτη αναφοράς και αξιολογήσεις.

Σκοπός τους είναι να παρέχουν έναν απλό και ενοποιημένο τρόπο για τη δημιουργία συγκρίσιμων αποτελεσμάτων αξιολόγησης σε διαφορετικά πλαίσια συστάσεων. Το EasyRec

⁴EasyRec, <http://easyrec.org/>

⁵GraphLab Create, <https://dato.com/products/create/>

⁶LensKit, <http://lenskit.org/>

⁷LibRec, <http://www.librec.net/>

⁸MyMediaLite, <http://www.mymedialite.net/>

⁹DATO, <https://dato.com/>

¹⁰Rival, <http://rival.recommenders.net/>

¹¹WrapRec, <https://github.com/babakx/WrapRec/>

είναι ένα ειδικό περιβάλλον που σχεδιάστηκε ως υπηρεσία Web που παραδίδεται στις εφαρμογές Web, αλλά ξεπεράστηκε και δεν ενημερώθηκε πια. Το LensKit και το LibRec είναι δύο σύγχρονες βιβλιοθήκες συστάσεων βασισμένες σε java, όπου το LensKit παρέχει μόνο την εφαρμογή κλασικών αλγορίθμων συστάσεων, όπως το φιλτράρισμα που βασίζεται στο χρήστη και το επαναλαμβανόμενο φιλτράρισμα kNN, τη κλίση ενός συνιστώμενου και τη παραγοντοποίηση μήτρας. Στην πραγματικότητα, μόνο οι MyMediaLite και το LibRec είναι σε θέση να παράσχουν τους πλέον σύγχρονους αλγόριθμους συστάσεων εκτός από τους κλασσικούς και το LibRec τρέχει πιο γρήγορα με βάση την αξιολόγηση αναφοράς, (Guo et. al., 2015). Σε αντίθεση με τους υπάρχοντες μηχανισμούς συστάσεων, το CARS Kit έχει σχεδιαστεί ειδικά για συστάσεις σχετικές με τα συμφραζόμενα, όπου οι πιο σύγχρονοι αλγόριθμοι συστάσεων στο CARS υλοποιούνται με τρόπο εύκολης διαμόρφωσης και ευέλικτης επέκτασης.

4.2. Σχεδιασμός και Χαρακτηριστικά του CARS Kit

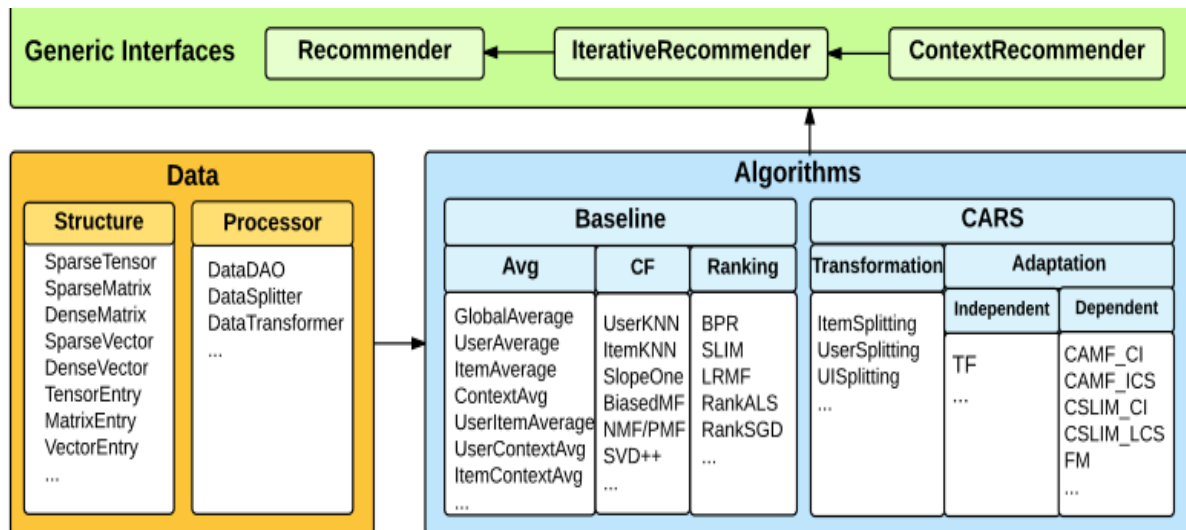
Το CARS Kit¹² είναι ένα ελεύθερο λογισμικό ανοιχτού κώδικα, όπου μπορεί να χρησιμοποιηθεί, να τροποποιηθεί και να διανεμηθεί σύμφωνα με τους όρους της Γενικής Δημόσιας Άδειας GNU. Στα επόμενα θα παρουσιαστούν το συγκεκριμένο σχέδιο και τα χαρακτηριστικά του CARS Kit.

4.2.1. Αρχιτεκτονική και Σχεδιασμός του CARS Kit

Το CARSKit παρέχει μια ευέλικτη αρχιτεκτονική έτσι ώστε να είναι εύκολο να επεκταθεί το εύρος των αλγορίθμων συμφραζομένων και να δημιουργηθούν χώροι για την ανάπτυξη νέων αλγορίθμων στο μέλλον.

Ολόκληρος ο σχεδιασμός μπορεί να απεικονιστεί από την Εικόνα 1. Η ροή εργασίας είναι απλή στο σχεδιασμό: οι διαφορετικοί αλγόριθμοι συστάσεων είναι οι συγκεκριμένες υλοποιήσεις και επεκτάσεις από τις γενικές διεπαφές όπου ορίζονται οι μοιρασμένες και οι κοινές λειτουργίες, όπως η πρόβλεψη βαθμολόγησης ή η βαθμολόγηση για ένα χρήστη σε ένα στοιχείο σε ένα συγκεκριμένο συμφραζόμενο. Οι αξιολογήσεις για τις προβλέψεις αξιολόγησης και οι κορυφαίες N συστάσεις ενσωματώνονται στη μηχανή Σύστασης.

¹²CARSKit, <https://github.com/irecsys/CARSKit/>



Εικόνα 1. Η αρχιτεκτονική και ο σχεδιασμός του CARS Kit

4.2.2. Χαρακτηριστικά του CARS Kit

Λόγω των ιδιαίτερων χαρακτηριστικών της μορφής των δεδομένων και των μεθόδων επεξεργασίας στο CARS, υπάρχουν πολλές προκλήσεις στο σχεδιασμό. Στα επόμενα θα εισαχθούν τα βασικά χαρακτηριστικά που ειδικά σχεδιάστηκαν για το CARS, τα οποία μπορούν να αναφερθούν περαιτέρω από αυτές τις πτυχές: μετασχηματιστής δεδομένων, δομή δεδομένων, αλγόριθμοι συστάσεων, διαμόρφωση και αξιολογήσεις.

Μετασχηματιστής Δεδομένων: Στο παραδοσιακό Σύστημα Συστάσεων, το πρόβλημα πρόβλεψης αξιολόγησης ξεκινά με ένα δισδιάστατο πίνακα βαθμολογιών:

$$\text{Χρήστες} \times \text{Στοιχεία} \rightarrow \text{Αξιολογήσεις.}$$

Στο CARS, τα συμφραζόμενα θεωρούνται ως πρόσθετες εισροές, με αποτέλεσμα τη δημιουργία μιας πολυδιάστατης λειτουργίας βαθμολόγησης:

$$\text{Χρήστες} \times \text{Στοιχεία} \times \text{Συμφραζόμενα} \rightarrow \text{Αξιολογήσεις, (Adomavicius et. al., 2011).}$$

Αυτή η ειδική μορφή των δεδομένων δημιουργεί προκλήσεις για τη δημιουργία ενός τέτοιου εξοπλισμού συστάσεων, όπου η αρχική δομή σε άλλες μηχανές συστάσεων δεν μπορεί να επαναχρησιμοποιηθεί εύκολα για συστάσεις ενήμερες από συμφραζόμενα. Πριν από τη συζήτηση της δομής δεδομένων, εισάγεται σε αυτό το σημείο ο μετασχηματιστής δεδομένων. Συνήθως, τα δεδομένα με βάση τα συμφραζόμενα μπορούν να αποθηκευτούν σε δύο μορφές: σε χαλαρή μορφή και σε συμπαγή μορφή, όπως φαίνεται στους παρακάτω πίνακες.

Πίνακας 4: Χαλαρή Μορφή

Χρήστης	Στοιχείο	Αξιολόγηση	Διάσταση	Κατάσταση
U1	T1	3	Χρόνος	Σαββατοκύριακο
U1	T1	3	Τοποθεσία	Δουλειά
U2	T2	4	Χρόνος	Καθημερινή
U2	T2	4	Τοποθεσία	Σπίτι

Πίνακας 5: Συμπαγής Μορφή

Χρήστης	Στοιχείο	Αξιολόγηση	Χρόνος	Τοποθεσία
U1	T1	3	Σαββατοκύριακο	Δουλειά
U2	T2	4	Καθημερινή	Σπίτι
U1	T1	4	Σαββατοκύριακο	Σπίτι
U2	T2	2	Καθημερινή	Δουλειά

Η διάσταση των συμφραζομένων είναι ίδια με τη μεταβλητή με βάση τα συμφραζόμενα, π.χ. Χρόνος και τοποθεσία, όπως φαίνεται στο παραπάνω παράδειγμα. Οι συνθήκες των συμφραζομένων αφορούν συγκεκριμένες τιμές σε μια διάσταση, π.χ. Σαββατοκύριακο και ημέρα εβδομάδας είναι δύο συνθήκες στη διάσταση Χρόνος. Η χαλαρή μορφή υποθέτει ότι υπάρχει μόνο μία αξιολόγηση για κάθε ζεύγος «χρήστη, στοιχείο» σε συσχετισμένα συμφραζόμενα, ενώ η συμπαγής μορφή επιτρέπει την αποθήκευση πολλαπλών αξιολογήσεων σε ένα ίδιο ζευγάρι «χρήστη, στοιχείο» σε διαφορετικές καταστάσεις με βάση τα

συμφραζόμενα. Σε συμπαγή μορφή, κάθε σειρά αντιπροσωπεύει ένα συγκεκριμένο προφίλ αξιολόγησης συμφραζομένων, δηλαδή, υπάρχουν μόνο δύο προφίλ αξιολόγησης συμφραζομένων στη χαλαρή μορφή αλλά τέσσερα προφίλ αξιολόγησης σε συμπαγή μορφή σε αυτό το παράδειγμα.

Οι περισσότερες πληροφορίες σχετικά με τα συμφραζόμενα έχουν τη μορφή κατηγορικών δεδομένων. Τόσο η χαλαρή όσο και η συμπαγής μορφή θα αυξήσουν την πίεση αποθήκευσης και το υπολογιστικό κόστος. Στο CARSKit, αποθηκεύεται η ταξινόμηση συμφραζομένων σε δυαδική μορφή, κάτι που είναι σε θέση να ενισχύσει σημαντικά την απόδοση της εκτέλεσης. Για να βοηθηθούν οι τελικοί χρήστες να προετοιμάσουν τα δεδομένα βαθμολόγησης, παρέχονται δύο μέθοδοι, ο Μετασχηματισμός από χαλαρή μορφή σε δυαδική και ο μετασχηματισμός από συμπαγή μορφή σε δυαδική ως μετασχηματιστές δεδομένων στο toolkit του περιβάλλοντος εργασίας.

Δομή των Δεδομένων: Δεδομένου ότι τα συμφραζόμενα θεωρούνται ως πρόσθετες εισροές, η δομή των δεδομένων γίνεται το πιο σημαντικό μέρος που αφήνει άμεσα την επίδραση του στην ευελιξία και την απόδοση λειτουργίας.

Υπάρχουν δύο παράγοντες που εμπλέκονται στο σχεδιασμό της δομής δεδομένων: η αποθήκευση των δεδομένων και οι λειτουργίες των δεδομένων. Διαισθητικά, τα δεδομένα που είναι ενήμερα με τα συμφραζόμενα μπορούν να εκπροσωπούνται σε ένα χώρο N-διαστάσεων ως τανυστές, όπου η κάθε διάσταση συμφραζόμενου θεωρείται ως μεμονωμένη διάσταση στον χώρο βαθμολόγησης. Σε αυτήν την περίπτωση, δημιουργούνται οι τανυστές συχνότητας (σπανιότητας) και καταχώρησης για να καταγραφούν οι δείκτες κάθε χρήστη, στοιχείου, διάστασης συμφραζόμενου και η σχετική βαθμολογία.

Αυτή η δομή είναι καλή για τους αλγόριθμους συστάσεων που λαμβάνουν γνώση των συμφραζόμενων χρησιμοποιώντας N-διαστατικές λειτουργίες, όπως τον αλγόριθμο συστάσεων multiverse που περιγράφεται στην εργασία των (Karatzoglou et. al., 2010). Εν τω μεταξύ, υπάρχουν πολλοί άλλοι αλγόριθμοι σύστασης συμφραζομένων που διερευνούν τις εξαρτήσεις μεταξύ των συμφραζόμενων και των διαστάσεων του χρήστη / στοιχείου, όπου η δισδιάστατη λειτουργία εξακολουθεί να είναι η πιο συχνή που υιοθετήθηκε σε αυτούς τους αλγόριθμους. Σε αυτή την περίπτωση, συνεχίζεται η χρήση των Sparse Matrix και Sparse Vector, οι οποίες είναι οι καλά αναγνωρισμένες και αποδοτικές αναπαραστάσεις δεδομένων σε υπάρχουσες βιβλιοθήκες συστάσεων, όπως το Lens Kit και το LibRec. Το Sparse Matrix

χρησιμοποιεί την συμπιεσμένη αποθήκευση γραμμών και στηλών¹³, η οποία επίσης αποδείχθηκε ότι ενισχύει την αποτελεσματικότητα λειτουργίας στο σχεδιασμό του Lib Rec,(Guo et. al., 2015).

Αλγόριθμοι Σύστασης: Όπως αναφέρθηκε προηγουμένως, το CARSKit είναι η 1η βιβλιοθήκη που είναι ειδικά σχεδιασμένη για το CARS. Όπως φαίνεται στην Εικόνα 1, οι αλγόριθμοι συμφραζομένων διαιρούνται σε δύο κατηγορίες: Τους Αλγόριθμους μετασχηματισμού και τους αλγόριθμους προσαρμογής.

Οι αλγόριθμοι μετασχηματισμού προσπαθούν να προ-επεξεργάζονται τα δεδομένα και να μετατρέπουν το σύνολο των δεδομένων των συμφραζομένων σε ένα δισδιάστατο πίνακα βαθμολογίας που περιέχει μόνο χρήστες, στοιχεία και βαθμολογίες, έτσι ώστε να μπορούν να εφαρμοστούν οποιοδήποτε παραδοσιακοί αλγόριθμοι συστάσεων. Μια από τις πιο αποτελεσματικές τεχνικές που εμπίπτουν σε αυτή την κατηγορία είναι οι προσεγγίσεις διάσπασης που έχουν επίγνωση των συμφραζόμενων,(Baltrunas & Ricci, 2013).

Οι αλγόριθμοι προσαρμογής ενσωματώνουν άμεσα τα συμφραζόμενα στη λειτουργία της πρόβλεψης. Υπάρχουν δύο υποκατηγορίες που περιλαμβάνονται: η ανεξάρτητη μοντελοποίηση (π.χ. TF,(Karatzoglou et. al., 2010)) η οποία υποθέτει ότι τα συμφραζόμενα είναι ανεξάρτητα από τους χρήστες (και στοιχεία) και η εξαρτώμενη μοντελοποίηση που εκμεταλλεύεται τις εξαρτήσεις μεταξύ των χρηστών, των αντικειμένων και των συμφραζόμενων, όπως το CAMF,(Baltrunas et. al., 2011), και η αραιή γραμμική μέθοδος συμφραζόμενων (CSLIM), (Zheng et. al., 2014). Η εξαρτώμενη μοντελοποίηση μπορεί να κατασκευαστεί με δύο τρόπους: με τη μοντελοποίηση των αποκλίσεων των διαβαθμίσεων των συμφραζόμενων,(Baltrunas et. al., 2011), και με την εκμάθηση των ομοιοτήτων των συμφραζόμενων,(Zheng et. al., 2015).

Οι μηχανές παραγοντοποίησης (FM), (Rendle et. al., 2011), είναι ένας εκλεπτυσμένος αλγόριθμος που εκμεταλλεύεται τα ζεύγη σχέσεων στη διαδικασία εκμάθησης του. Μεταξύ αυτών των αλγορίθμων, οι TF και CAMF είναι δύο δημοφιλείς αλγόριθμοι που έχουν αναγνωριστεί ως οι βασικοί αλγόριθμοι στο CARS.

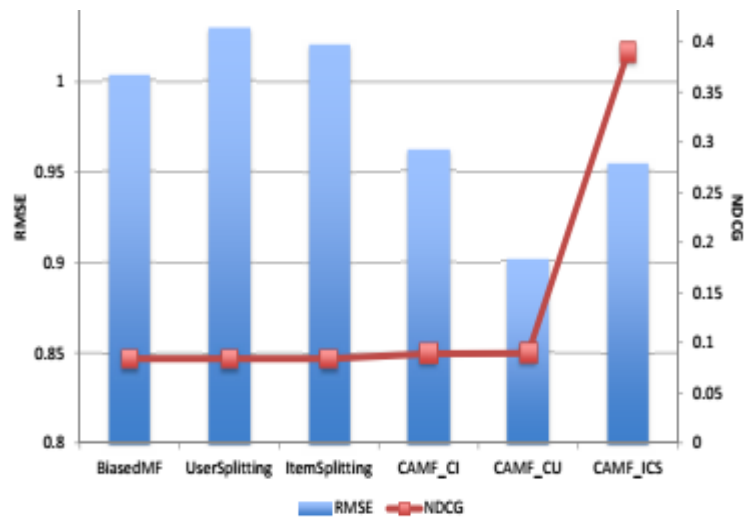
Εκτός από εκείνους τους σύγχρονους αλγορίθμους συστάσεων με συμφραζόμενα, στην εργασία των Zheng και των συνεργατών του, (Zheng et. al., 2015), συμπεριλαμβάνονται επίσης

¹³Sparse Matrix Storage, http://netlib.org/linalg/html_templates/node90.html

ορισμένοι παραδοσιακοί αλγόριθμοι συστάσεων σαν βασική γραμμή του πακέτου. Στην εργασία των Zheng και των συνεργατών του, (Zheng et. al., 2015) δεν ξανασυναρμολογήθηκαν αυτοί οι αλγόριθμοι και δεν επαναχρησιμοποιήθηκαν άμεσα οι κλασικές μηχανές σύστασης που παρέχονται από το Lib Rec. Υπάρχουν δύο κύριοι λόγοι για να συμπεριληφθούν αυτοί οι παραδοσιακοί αλγόριθμοι συστάσεων. Από τη μια πλευρά, αυτοί οι αλγόριθμοι μπορούν να εφαρμοστούν μετά τον μετασχηματισμό των δεδομένων (π.χ. λειτουργίες διαίρεσης), το οποίο είναι ένα ουσιαστικό βήμα στους αλγόριθμους μετασχηματισμού που έχουν επίγνωση των συμφραζόμενων. Από την άλλη πλευρά, συνήθως είναι κοινός ο ανταγωνισμός ενός αλγορίθμου συμφραζόμενων με αλγορίθμους μη-συμφραζόμενων για να κριθεί αν το αποτέλεσμα είναι σημαντικό ή το ότι είναι απαραίτητος ο αλγόριθμος συστάσεων που λαμβάνει υπόψη τα συμφραζόμενα.

Διαμόρφωση και αξιολογήσεις: Για σκοπούς αξιολόγησης, παρέχεται στην εργασία των Zheng και των συνεργατών του, (Zheng et. al., 2015), το Data Splitter το οποίο επιτρέπει στους χρήστες να υιοθετήσουν είτε την αξιολόγηση με δοκιμή είτε τις εγκάρσιες Νεπικυρώσεις. Οι περισσότεροι από τους αλγορίθμους συστάσεων που είναι ενσωματωμένοι στο CARSKit είναι ικανοί να εκτελέσουν τις δύο εργασίες σύστασης: πρόβλεψη αξιολόγησης και σύσταση στοιχείου, εκτός από εκείνες που σχεδιάστηκαν ειδικά για σύσταση top-N, όπως το CSLIM. Αλλά η αξιολόγηση είναι διαφορετική από τις παραδοσιακές αξιολογήσεις, καθώς τα συμφραζόμενα αποτελούν πρόσθετες εισροές στη διαδικασία αξιολόγησης.

Συνήθως, η πρόβλεψη της αξιολόγησης μπορεί να αξιολογηθεί με διαφορετικά σφάλματα πρόβλεψης, όπως το μέσο απόλυτο σφάλμα (MAE), το μέσο τετραγωνικό σφάλμα ρίζας (RMSE) και το μέσο σφάλμα πρόβλεψης (MPE). Η σύσταση του στοιχείου μπορεί να αξιολογηθεί με μετρήσεις συνάφειας, όπως η ακρίβεια και η ανάκληση, και με μετρήσεις κατάταξης, όπως η μέση τιμή της μέσης ακρίβειας (MAP), το κανονικοποιημένο αποκομμένο σωρευτικό κέρδος (NDCG) και η μέση αμοιβαιότητα (MMR).



Εικόνα 2. Δείγμα των αξιολογήσεων σε δεδομένα του DePaul Movie

Εξερευνήθηκε μια πενταπλάσια διασταύρωση επικύρωσης στα δεδομένα DePaul Movie¹⁴ χρησιμοποιώντας αλγορίθμους συστάσεων που βασίζονται στην παραγοντοποίηση πίνακα. Τα αποτελέσματα της αξιολόγησης με βάση τις προβλέψεις αξιολόγησης (από το RMSE) και τις κορυφαίες 5 συστάσεις (από το NDCG) παρουσιάζονται στην Εικόνα 2. Μεταξύ αυτών των αλγορίθμων, το Biased MF είναι μια τυποποιημένη τεχνική παραγοντοποίησης μήτρας και επίσης ένας αλγόριθμος μη συμφραζομένων που εφαρμόζεται στα δεδομένα που είναι ενήμερα με συμφραζόμενα.

Ο διαχωρισμός των χρηστών και η διάσπαση του στοιχείου είναι δύο προσεγγίσεις διαχωρισμού συμφραζομένων,(Zheng et. al., 2014), στις οποίες εφαρμόζεται Biased MF μετά από μετασχηματισμό των δεδομένων. Επιπροσθέτως, εξετάζονται επίσης ο CAMF με βάση την απόκλιση (π.χ. CAMF CI και CAMF CU),(Baltrunas et. al., 2011), και ο CAMF με βάση την ομοιότητα (π.χ. CAMF ICS),(Zheng et. al., 2015). Τα αποτελέσματα αποκαλύπτουν ότι το CAMF ICS είναι σε θέση να βοηθήσει στην απόκτηση του καλύτερου NDCG σε αυτό το σύνολο δεδομένων, ενώ δεν υπάρχουν στατιστικές σημαντικές διαφορές στο RMSE μεταξύ αυτών των αλγορίθμων. Επιπλέον, το CARSKit παρέχει ευέλικτες διαμορφώσεις από ένα μόνο αρχείο το οποίο περιλαμβάνει και διαμόρφωση αλγορίθμου (π.χ. παραμέτρους αλγορίθμου) και πειραματική διαμόρφωση, π.χ. εισροές και έξοδοι, στρατηγική αξιολόγησης και μετρήσεις κ.λπ.

¹⁴ Υπάρχουν 5029 αξιολογήσεις (βαθμοί από 1 έως 5) από 97 χρήστες σε 79 ταινίες, με τα συμφραζόμενα «χρόνος, τοποθεσία, συντροφιά»

ΚΕΦ.5: Αξιολόγηση με Πειραματικά Δεδομένα

5.1. Δεδομένα

Αυτήν η εργασία θα εστιάσει σε δεδομένα αξιολόγησης εστιατορίων, τα οποία παρέχουν επίσης πληροφορίες για το context. Χρησιμοποιούνται δύο σύνολα δεδομένων, το Tijuana Restaurants (Ramirez-Garcia & García-Valdez, 2014) και το Japan Restaurants (Oku et al., 2006). Τα σύνολα δεδομένων παρέχονται από το CARSKit, αλλά χρειάζονται προεπεξεργασία για να δημιουργηθεί το τελικό αρχείο που θα δοθεί ως είσοδος στους αλγορίθμους.

Από τα δεδομένα Tijuana Restaurants δημιουργήθηκε ένα csv αρχείο, με στήλες User, Restaurant, Rating, Time, Location. Οι δύο τελευταίες στήλες δημιουργήθηκαν από την αρχική μεταβλητή context, η οποία έχει τιμή 1-6, με 1-3 να είναι εργάσιμη μέρα και 4-6 αργία, ενώ η τοποθεσία είναι School (τιμές 1,4), Home(τιμές (2,5), Work (3,6), ενώ απορρίφθηκε μια εγγραφή με ελλιπή δεδομένα. Τα δεδομένα έχουν συνολικά 1421 αξιολογήσεις από 50 χρήστες για 40 εστιατόρια.

Για τα δεδομένα Japan Restaurants, συνδυάστηκαν οι αξιολογήσεις των 8 χρηστών σε ένα αρχείο. Αυτά τα δεδομένα δεν περιγράφουν ποιο εστιατόριο αξιολογείται, ωστόσο δίνουν 17 παραμέτρους για το κάθε ένα, οπότε συνδυάστηκαν και βρέθηκαν 69 μοναδικά «μοτίβα», τα οποία χρησιμοποιήθηκαν για να βρεθεί σε ποιο εστιατόριο ανήκει η αξιολόγηση. Οπότε, τα δεδομένα περιέχουν 800 αξιολογήσεις 69 εστιατορίων από 8 χρήστες, ενώ έχουν τις 14 μεταβλητές context:

- Time:
 - Month {1=Jan. to 12=Dec.}
 - Week {0=Monday to 6=Sunday}
 - Hour {0 to 23}
- Schedule:
 - Area Type {0=None, 1=Entertainment District, 3=Near Station, 4=Tourist Resort}
 - Budget (Yen) {0 to 10000}
 - Holiday {0=None, 1=A Day OFF, 2=Recess, 3=Before Holiday}
- Partner:
 - Num. of Male {0 to 10}

- Num. of Female {0 to 10}
- Lowest Age {0 to 100}
- Highest Age {0 to 100}
- Relation {0=None, 1=Family, 2=Boy/Girlfriend, 3=Friend, 4=Boss, 5=Subordinate}
- Status {0=None, 1=Student, 2=Working}
- External Factor:
 - Weather {0=Fine, 1=Cloudy, 2=Rainy}
 - Temperature (Celsius) {-5 to 40}

Μετρικές Αξιολόγησης Αποτελεσμάτων

Έστω ότι ο αλγόριθμος προβλέπει N αξιολογήσεις $pred_i$ με $i=1, 2, \dots, N$, ενώ οι πραγματικές τιμές των αξιολογήσεων είναι $rating_i$ με $i=1, 2, \dots, N$. Οι μετρικές που θα χρησιμοποιηθούν για να αξιολογηθούν οι προβλέψεις των διαφορετικών αλγορίθμων είναι οι εξής:

- MAE: Mean absolute error

$$MAE = \frac{1}{N} \sum_{i=1}^N |(pred_i - rating_i)|$$

- RMSE: Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (pred_i - rating_i)^2}$$

- NMAE: Normalized MAE

$$NMAE = \frac{1}{N} \left(\sum_{i=1}^N |(pred_i - rating_i)| \right) / (maxRate - minRate)$$

Στη συνέχεια, υπολογίζονται το MAE και το MSE, αλλά οι προβλέψεις των αλγορίθμων στρογγυλοποιούνται ώστε να αντιστοιχούν στα επίπεδα των πραγματικών τιμών:

$$rPred_i = round(pred_i / minRate) * minRate,$$

- rMAE: rounded MAE

$$rMAE = \frac{1}{N} \sum_{i=1}^N |(rPred_i - rating_i)|$$

- rRMSE: rounded MSE

$$rRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (rPred_i - rating_i)^2}$$

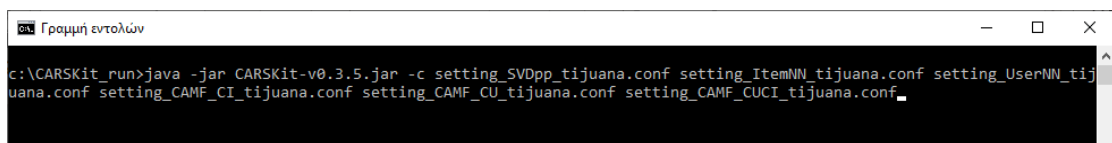
5.2. Αποτελέσματα

Στη συνέχεια, δημιουργήθηκαν τα κατάλληλα configuration files, ένα για κάθε αλγόριθμο και σύνολο δεδομένων, ώστε να τρέξουν οι αλγόριθμοι. Ως task, ορίστηκε η πρόβλεψη της αξιολόγησης, η οποία γίνεται με cross-validation (K=5). Επίσης, τα αρχεία δεδομένων όπως περιεγράφηκαν προηγουμένως είναι στην «compact» μορφή, οπότε ορίστηκαν να μετατραπούν σε binary (δηλ. μια στήλη ανά τιμή, π.χ. location).

Κρατήθηκαν τις προεπιλεγμένες τιμές για κάθε αλγόριθμο, οι οποίες είναι:

- numFactors:10
- numIter:100
- lrate:0.02
- maxlrate:-1.0
- regB:1.0E-4
- regU:1.0E-4
- regI:1.0E-4
- regC:0.001

Ένα παράδειγμα εκτέλεσης του αλγορίθμου φαίνεται στην Εικόνα 3.



Εικόνα 3. Εκτέλεση διαφόρων αλγορίθμων για το σύνολο δεδομένων Tijuana Restaurants

Στον Πίνακα 6 φαίνονται τα αποτελέσματα της εκτέλεσης των διαφορετικών αλγορίθμων για το σύνολο δεδομένων Tijuana Restaurants. Μιας και οι μετρικές σχετίζονται με σφάλμα πρόβλεψης, όσο μικρότερη η τιμή, τόσο καλύτερη η απόδοση του αλγορίθμου. Παρατηρείται ότι σε όλες τις περιπτώσεις, οι αλγόριθμοι που λαμβάνουν υπόψη τους το context είναι πιο αποδοτικοί. Συγκεκριμένα, ο CAMF_CU έχει την καλύτερη απόδοση από όλους τους αλγορίθμους σε όλες τις μετρικές. Από αυτούς που δεν το λαμβάνουν υπόψη, ο SVD++ έχει το

χαμηλότερο σφάλμα σε όλες τις μετρικές, εκτός από το RMSE, στο οποίο ο ItemKNN είναι πιο αποδοτικός.

Πίνακας 6. Μετρικές απόδοσης των αλγόριθμων για τα δεδομένα Tijuana Restaurants

Μέθοδος	MAE	RMSE	NMAE	rMAE	rRMSE
SVD++	0,737378	1,159896	0,184344	0,719046	1,167279
ItemKNN	0,834045	1,144764	0,208511	0,791913	1,172747
UserKNN	1,134132	1,350692	0,283533	1,090607	1,352743
CAMF_CI	0,757313	1,090457	0,189328	0,728980	1,120061
CAMF_CU	0,635986	0,971674	0,158996	0,598008	0,988785
CAMF_CUCI	0,677166	0,987350	0,169292	0,656086	1,032673

Στη συνέχεια εκτελέστηκαν οι αλγόριθμοι για το σύνολο δεδομένων JapanRestaurants. Τα αποτελέσματα φαίνονται στον Πίνακα 7. Αρχικά, παρατηρείται ότι όλοι οι αλγόριθμοι δίνουν το ίδιο rMAE και rRMSE, ενώ στις υπόλοιπες μετρικές, ο καλύτερος είναι ο ItemKNN. Από τους context-aware αλγόριθμους, ο CAMF_CU δίνει το καλύτερο αποτέλεσμα. Εδώ φαίνεται ότι ίσως ο αλγόριθμος μπερδεύεται επειδή οι μεταβλητές context είναι πάρα πολλές και πιθανότατα αρκετές από αυτές δεν παίζουν ρόλο στην αξιολόγηση.

Πίνακας 7. Μετρικές απόδοσης των αλγόριθμων για τα δεδομένα Japan Restaurants

Μέθοδος	MAE	RMSE	NMAE	rMAE	rRMSE
SVD++	0,214806	0,264605	0,214806	0,576023	0,629568
ItemKNN	0,211873	0,250658	0,211873	0,576023	0,629568
UserKNN	0,235574	0,270273	0,235574	0,576023	0,629568
CAMF_CI	0,335467	0,413353	0,335467	0,576023	0,629568
CAMF_CU	0,251838	0,311689	0,251838	0,576023	0,629568
CAMF_CUCI	0,304778	0,374656	0,304778	0,576023	0,629568

Οπότε, και στις δύο περιπτώσεις, ο καλύτερος context-aware αλγόριθμος είναι ο CAMF_CU, ο οποίος πραγματοποιεί επεξεργασία των δεδομένων και παραγοντοποίηση του πίνακα (matrix factorization) με βάση τους χρήστες. Ωστόσο, στο δεύτερο παράδειγμα, οι κλασικοί αλγόριθμοι φαίνεται να αποδίδουν καλύτερα, πιθανόν επειδή δεν υπάρχουν οι κατάλληλες μεταβλητές για το context.

ΚΕΦ.6: Συμπεράσματα

Τα Συστήματα παροχής συστάσεων έχουν χρησιμοποιηθεί για την επιτυχή αντιμετώπιση του προβλήματος της υπερφόρτωσης πληροφοριών παρέχοντας συστάσεις στους τελικούς χρήστες. Τα τελευταία χρόνια, οι ερευνητές συνειδητοποίησαν τη σημασία των συμφραζόμενων και προσπαθούν να ενσωματώσουν τα περιεχόμενα στα Συστήματα Συστάσεων για να δημιουργήσουν συστήματα παροχής συστάσεων με γνώμονα τα συμφραζόμενα (CARS) τα οποία είναι ικανά να προσαρμόζουν τις συστάσεις στις συναφείς συνθήκες των χρηστών. Το συμφραζόμενο ορίζεται ως κάθε πληροφορία που μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει κάθε οντότητα, (Dey, 2001). Στο CARS, τα συμφραζόμενα επιλέγονται συνήθως από τα δυναμικά χαρακτηριστικά της ίδιας της δραστηριότητας, (Zheng et. al., 2015).

Για παράδειγμα, ο χρήστης μπορεί να επιλέξει ένα διαφορετικό τύπο ταινίας εάν πρόκειται να παρακολουθήσει την ταινία με παιδιά και όχι με συνεργάτες, όπου η συντροφιά είναι ένα χαρακτηριστικό από τη δραστηριότητα "βλέποντας μια ταινία" και η τιμή του μπορεί να αλλάξει συχνά όταν η ίδια δραστηριότητα συμβαίνει επανειλημμένα. Επιπλέον, τα περιορισμένα χαρακτηριστικά από τα προφίλ των χρηστών μπορούν να θεωρηθούν επίσης ως συμφραζόμενα, όπως η πρόθεση του χρήστη και οι συναισθηματικές καταστάσεις. Τα περισσότερα χαρακτηριστικά του στοιχείου, όπως το είδος της ταινίας, ο τραγουδιστής μουσικής ή οι υπηρεσίες δωματίων του ξενοδοχείου, θεωρούνται ως περιεχόμενο στα Συστήματα Συστάσεων. Τα συστήματα CARS έχουν αναπτυχθεί αρκετά την τελευταία δεκαετία. Πολλοί κλασσικοί αλγόριθμοι σύστασης από τα συμφραζόμενα, όπως οι προσεγγίσεις διάσπασης με βάση τα συμφραζόμενα (CASA), (Baltrunas & Ricci, 2013), η παραγοντοποίηση τανυστή (TF), (Karatzoglou et. al., 2010), και η παραγοντοποίηση πίνακα από τα συμφραζόμενα (CAMF), (Baltrunas et. al., 2011), έχουν πλέον αναγνωριστεί σαν οι βασικές γραμμές σε αυτόν τον τομέα.

Ωστόσο, δεν υπάρχουν μηχανισμοί συστάσεων που να υλοποιούν αλγορίθμους συστάσεων ενήμερους με τα συμφραζόμενα, λόγω των ιδιαίτερων χαρακτηριστικών της μορφής των δεδομένων και των μεθόδων επεξεργασίας στον τομέα αυτό. Οι ερευνητές πρέπει να μεταφράσουν αυτούς τους αλγορίθμους επανειλημμένα, γεγονός που παρεμποδίζει περαιτέρω την εξέλιξη της ανάπτυξης του CARS. Ως εκ τούτου, υπάρχει μεγάλη ανάγκη για μια τέτοια βιβλιοθήκη συστάσεων, ώστε όχι μόνο να ικανοποιηθεί ο σκοπός του βασικού

σχεδιασμού και της αξιολόγησης, αλλά και για να ενισχυθεί η περαιτέρω ανάπτυξη των συστάσεων που είναι ενήμερες με τα συμφραζόμενα.

Στην εργασία των Zheng και των συνεργατών του, (Zheng et. al., 2015), παρουσιάζεται το CARSKit - μια μηχανή σύστασης ανοικτού κώδικα βασισμένη σε Java που είναι ενήμερη με τα συμφραζόμενα και που εφαρμόζει τους πιο σύγχρονους αλγορίθμους συμφραζομένων για παροχή συστάσεων και παρέχει μια τυποποιημένη πλατφόρμα για αξιολογήσεις και πρακτική χρήση. Από όσο γνωρίζουμε, είναι η πρώτη βιβλιοθήκη ανοικτού κώδικα που είναι ειδικά σχεδιασμένη για το CARS και είναι σε θέση να παρέχει ευέλικτες διαμορφώσεις και συμβατούς χώρους για επέκταση με νέους αλγόριθμους. Το CARSKit συμβάλλει στον τομέα των συστημάτων συστάσεων με την παροχή: 1) της 1ης βιβλιοθήκης ειδικά σχεδιασμένης για συστάσεις ενήμερες με συμφραζόμενα. 2) των εφαρμογών των αλγορίθμων σύγχρονων συστημικών συστάσεων τελευταίας τεχνολογίας. 3) της αποτελεσματικής δομής δεδομένων και της ευέλικτης διαμόρφωσης, καθώς και τις αξιολογήσεις πολλαπλών πτυχών. Στο μέλλον, θα προστεθούν περισσότεροι αλγόριθμοι συστάσεων ή μέρη από τα συμφραζόμενα, στο CARSKit. Είναι επίσης δυνατό να συμπεριληφθούν αλγόριθμοι προτάσεων των συμφραζομένων, όπου η πρόταση συμφραζόμενου ή η σύσταση συμφραζόμενου, (Zheng, 2015), είναι μια νέα κατεύθυνση που προέρχεται από το CARS και έχει ως στόχο να προτείνει κατάλληλα συμφραζόμενα για χρήση από τους χρήστες στα στοιχεία τους.

Βιβλιογραφία

- Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In Recommender systems handbook (pp. 217-253). Springer, Boston, MA.
- Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems (TOIS)*, 23(1), 103-145.
- Aghdam, M. H., Hariri, N., Mobasher, B., & Burke, R. D. (2015). Adapting Recommendations to Contextual Changes Using Hierarchical Hidden Markov Models. *RecSys*, 15, 241-244.
- Akhil, P. V., & Joseph, S. (2017). A SURVEY OF RECOMMENDER SYSTEM TYPES AND ITS CLASSIFICATION. *International Journal of Advanced Research in Computer Science*, 8(9).
- Allahyari, M., & Kochut, K. (2016, November). Semantic context-aware recommendation via topic models leveraging linked open data. In *International Conference on Web Information Systems Engineering* (pp. 263-277). Springer, Cham.
- Anand, S. S., & Mobasher, B. (2007). Contextual Recommendation, From Web to Social Web: Discovering and Deploying User and Content Profiles: Workshop on Web Mining, WebMine 2006, Berlin, Germany, September 18, 2006. Revised Selected and Invited Papers.
- Baltrunas, L., & Ricci, F. (2014). Experimental evaluation of context-dependent collaborative filtering using item splitting. *User Modeling and User-Adapted Interaction*, 24(1-2), 7-34.
- Baltrunas, L., Church, K., Karatzoglou, A., & Oliver, N. (2015). Frappe: Understanding the usage and perception of mobile app recommendations in-the-wild. *arXiv preprint arXiv:1505.03014*.
- Baltrunas, L., Kaminskas, M., Ludwig, B., Moling, O., Ricci, F., Aydin, A., ... & Schwaiger, R. (2011, August). Incarmusic: Context-aware music recommendations in a car. In *International Conference on Electronic Commerce and Web Technologies* (pp. 89-100). Springer, Berlin, Heidelberg.

- Baltrunas, L., Ludwig, B., & Ricci, F. (2011, October). Matrix factorization techniques for context aware recommendation. In Proceedings of the fifth ACM conference on Recommender systems (pp. 301-304). ACM.
- Baltrunas, L., Ludwig, B., Peer, S., & Ricci, F. (2011, July). Context-aware places of interest recommendations for mobile users. In International Conference of Design, User Experience, and Usability (pp. 531-540). Springer, Berlin, Heidelberg.
- Bazire, M., & Brézillon, P. (2005, July). Understanding context before using it. In International and Interdisciplinary Conference on Modeling and Using Context (pp. 29-40). Springer, Berlin, Heidelberg.
- Berry, D. M. (2012). Introduction: Understanding the digital humanities. In Understanding digital humanities (pp. 1-20). Palgrave Macmillan, London.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.
- Bouneffouf, D. (2013). DRARS, a dynamic risk-aware recommender system (Doctoral dissertation).
- Bouneffouf, D., Bouzeghoub, A., & Gançarski, A. L. (2012, November). A contextual-bandit algorithm for mobile context-aware recommender system. In International Conference on Neural Information Processing (pp. 324-331). Springer, Berlin, Heidelberg.
- Braunhofer, M. (2014, October). Hybridisation techniques for cold-starting context-aware recommender systems. In Proceedings of the 8th ACM Conference on Recommender systems (pp. 405-408). ACM.
- Breese, J. S., Heckerman, D., & Kadie, C. (1998, July). Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence (pp. 43-52). Morgan Kaufmann Publishers Inc..
- Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Foundations and Trends® in Machine Learning, 5(1), 1-122.

- Champiri, Z. D., Shahamiri, S. R., & Salim, S. S. B. (2015). A systematic review of scholar context-aware recommender systems. *Expert Systems with Applications*, 42(3), 1743-1758.
- Consolvo, S., Arnstein, L., & Franza, B. R. (2002, September). User study techniques in the design and evaluation of a ubicomp environment. In *International Conference on Ubiquitous Computing* (pp. 73-90). Springer, Berlin, Heidelberg.
- Dehuri, S. (Ed.). (2012). *Intelligent Techniques in Recommendation Systems: Contextual Advancements and New Methods: Contextual Advancements and New Methods*. IGI Global.
- Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing*, 5(1), 4-7.
- Dourish, P. (2004). What we talk about when we talk about context. *Personal and ubiquitous computing*, 8(1), 19-30.
- Elahi, M., Braunhofer, M., Ricci, F., & Tkalcic, M. (2013, December). Personality-based active learning for collaborative filtering recommender systems. In *Congress of the Italian Association for Artificial Intelligence* (pp. 360-371). Springer, Cham.
- Frolov, E., & Oseledets, I. (2017). Tensor methods and recommender systems. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(3), e1201.
- Ge, Y., Xiong, H., Tuzhilin, A., Xiao, K., Gruteser, M., & Pazzani, M. (2010, July). An energy-efficient mobile recommender system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 899-908). ACM.
- Ghazanfar, M. A., Prügell-Bennett, A., & Szedmak, S. (2012). Kernel-mapping recommender system algorithms. *Information Sciences*, 208, 81-104.
- Gu, Y., Song, J., Liu, W., Zou, L., & Yao, Y. (2016, October). Context aware matrix factorization for event recommendation in event-based social networks. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 248-255). IEEE.
- Guo, G., Zhang, J., Sun, Z., & Yorke-Smith, N. (2015, June). LibRec: A Java Library for Recommender Systems. In *UMAP Workshops* (Vol. 4).

- Hariri, N., Mobasher, B., & Burke, R. (2013, October). Query-driven context aware recommendation. In Proceedings of the 7th ACM conference on Recommender systems (pp. 9-16). ACM.
- Hariri, N., Mobasher, B., & Burke, R. (2014, October). Context adaptation in interactive recommender systems. In Proceedings of the 8th ACM Conference on Recommender systems (pp. 41-48). ACM.
- Haruna, K., Akmar Ismail, M., Suhendroyono, S., Damiasih, D., Pierewan, A., Chiroma, H., & Herawan, T. (2017). Context-aware recommender system: a review of recent developmental process and future research direction. *Applied Sciences*, 7(12), 1211.
- He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, L. (2010, April). Context-aware citation recommendation. In Proceedings of the 19th international conference on World wide web (pp. 421-430). ACM.
- Hidasi, B., & Tikk, D. (2016). General factorization framework for context-aware recommendations. *Data Mining and Knowledge Discovery*, 30(2), 342-371.
- Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995, May). Recommending and evaluating choices in a virtual community of use. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 194-201). ACM Press/Addison-Wesley Publishing Co..
- Hoekstra, R. (2010). The knowledge reengineering bottleneck. *Semantic Web*, 1(1, 2), 111-115.
- Hsieh, C. K., Yang, L., Wei, H., Naaman, M., & Estrin, D. (2016, April). Immersive recommendation: News and event recommendations using personal digital traces. In Proceedings of the 25th International Conference on World Wide Web (pp. 51-62). International World Wide Web Conferences Steering Committee.
- Jiang, M., Cui, P., Wang, F., Zhu, W., & Yang, S. (2014). Scalable recommendation with social contextual information. *IEEE Transactions on Knowledge and Data Engineering*, 26(11), 2789-2802.
- Jin, X., Zhou, Y., & Mobasher, B. (2005, July). Task-oriented web user modeling for recommendation. In International Conference on User Modeling (pp. 109-118). Springer, Berlin, Heidelberg.

- Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010, September). Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In Proceedings of the fourth ACM conference on Recommender systems (pp. 79-86). ACM.
- Karlgren, J. (1994). Newsgroup Clustering Based On User Behavior-A Recommendation Algebra. SICS Research Report.
- Koren, Y. (2008, August). Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 426-434). ACM.
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, (8), 30-37.
- Lakiotaki, K., Matsatsinis, N. F., & Tsoukias, A. (2011). Multicriteria user modeling in recommender systems. *IEEE Intelligent Systems*, 26(2), 64-76.
- Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788.
- Linden, G. D., Jacobi, J. A., & Benson, E. A. (2001). U.S. Patent No. 6,266,649. Washington, DC: U.S. Patent and Trademark Office.
- Liu, Q., Ma, H., Chen, E., & Xiong, H. (2013). A survey of context-aware mobile recommendations. *International Journal of Information Technology & Decision Making*, 12(01), 139-172.
- Liu, Q., Wu, S., & Wang, L. (2015). Cot: Contextual operating tensor for context-aware recommender systems.
- Macedo, A. Q., Marinho, L. B., & Santos, R. L. (2015, September). Context-aware event recommendation in event-based social networks. In Proceedings of the 9th ACM Conference on Recommender Systems (pp. 123-130). ACM.
- Machado, L. (2019). Fair team recommendations for multidisciplinary projects (Master's thesis).

- Mahmood, T., Ricci, F., & Venturini, A. (2009). Improving recommendation effectiveness: Adapting a dialogue strategy in online travel planning. *Information Technology & Tourism*, 11(4), 285-302.
- Man, T., Shen, H., Huang, J., & Cheng, X. (2015, October). Context-adaptive matrix factorization for multi-context recommendation. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (pp. 901-910). ACM.
- Mnih, A., & Salakhutdinov, R. R. (2008). Probabilistic matrix factorization. In *Advances in neural information processing systems* (pp. 1257-1264).
- Mooney, R. J., & Roy, L. (2000, June). Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries* (pp. 195-204). ACM.
- Ning, X., & Karypis, G. (2011, December). Slim: Sparse linear methods for top-n recommender systems. In *2011 IEEE 11th International Conference on Data Mining* (pp. 497-506). IEEE.
- Odić, A., Tkalič, M., Tasič, J. F., & Košir, A. (2013). Predicting and detecting the relevant contextual information in a movie-recommender system. *Interacting with Computers*, 25(1), 74-90.
- Oku, K., Nakajima, S., Miyazaki, J., & Uemura, S. (2006, May). Context-aware SVM for context-dependent information recommendation. In *Proceedings of the 7th international Conference on Mobile Data Management* (p. 109). IEEE Computer Society.
- Ono, C., Takishima, Y., Motomura, Y., & Asoh, H. (2009, June). Context-aware preference model based on a study of difference between real and supposed situation data. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 102-113). Springer, Berlin, Heidelberg.
- Palmisano, C., Tuzhilin, A., & Gorgoglione, M. (2008). Using context to improve predictive modeling of customers in personalization applications. *IEEE transactions on knowledge and data engineering*, 20(11), 1535-1549.
- Panniello, U., Tuzhilin, A., Gorgoglione, M., Palmisano, C., & Pedone, A. (2009, October). Experimental comparison of pre-vs. post-filtering approaches in context-aware

- recommender systems. In Proceedings of the third ACM conference on Recommender systems (pp. 265-268). ACM.
- Park, H. S., Yoo, J. O., & Cho, S. B. (2006, September). A context-aware music recommendation system using fuzzy bayesian networks with utility theory. In International conference on Fuzzy systems and knowledge discovery (pp. 970-979). Springer, Berlin, Heidelberg.
- Parsons, J., Ralph, P., & Gallagher, K. (2004). Using viewing time to infer user preference in recommender systems.
- Ramirez-Garcia, X., & García-Valdez, M. (2014). Post-filtering for a restaurant context-aware recommender system. In Recent Advances on Hybrid Approaches for Designing Intelligent Systems (pp. 695-707). Springer, Cham.
- Ren, X., Song, M., Haihong, E., & Song, J. (2017). Context-aware probabilistic matrix factorization modeling for point-of-interest recommendation. *Neurocomputing*, 241, 38-55.
- Rendle, S., Gantner, Z., Freudenthaler, C., & Schmidt-Thieme, L. (2011, July). Fast context-aware recommendations with factorization machines. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 635-644). ACM.
- Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-59.
- Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1-35). Springer, Boston, MA.
- Ruotsalo, T., Haav, K., Stoyanov, A., Roche, S., Fani, E., Deliai, R., ... & Hyvönen, E. (2013). SMARTMUSEUM: A mobile recommender system for the Web of Data. *Web semantics: Science, services and agents on the world wide web*, 20, 50-67.
- Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M., & Hanjalic, A. (2014, November). Cars2: Learning context-aware representations for context-aware recommendations. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (pp. 291-300). ACM.

- Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M., Hanjalic, A., & Oliver, N. (2012, August). TFMAP: optimizing MAP for top-n context-aware recommendation. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 155-164). ACM.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012, July). Exploring topic coherence over many models and many topics. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (pp. 952-961). Association for Computational Linguistics.
- Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication. Cambridgeuniversitypress.
- Takács, G., Pilászy, I., Németh, B., & Tikk, D. (2008, October). Matrix factorization and neighbor based algorithms for the netflix prize problem. In Proceedings of the 2008 ACM conference on Recommender systems (pp. 267-274). ACM.
- Tang, J., Gao, H., Hu, X., & Liu, H. (2013, October). Context-aware review helpfulness rating prediction. In Proceedings of the 7th ACM conference on Recommender systems (pp. 1-8). ACM.
- Tang, L., Jiang, Y., Li, L., & Li, T. (2014, October). Ensemble contextual bandits for personalized recommendation. In Proceedings of the 8th ACM Conference on Recommender Systems (pp. 73-80). ACM.
- Tavakol, M., & Brefeld, U. (2014, October). Factored MDPs for detecting topics of user sessions. In Proceedings of the 8th ACM Conference on Recommender Systems (pp. 33-40). ACM.
- Udell, M., Horn, C., Zadeh, R., & Boyd, S. (2016). Generalized low rank models. Foundations and Trends® in Machine Learning, 9(1), 1-118.
- Unger, M., Bar, A., Shapira, B., & Rokach, L. (2016). Towards latent context-aware recommendation systems. Knowledge-Based Systems, 104, 165-178.
- Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., & Duval, E. (2012). Context-aware recommender systems for learning: a survey and future challenges. IEEE Transactions on Learning Technologies, 5(4), 318-335.

- Villegas, N., Sanchez, C., Diaz-Cely, J., Tamura, G., (2018, January). Characterizing context-aware recommender systems a systematic literature review. *Knowledge-Based Systems*, 140, 173-200.
- Wermser, H., Rettinger, A., & Tresp, V. (2011, July). Modeling and learning context-aware recommendation scenarios using tensor decomposition. In *2011 International Conference on Advances in Social Networks Analysis and Mining* (pp. 137-144). IEEE.
- Wu, Q., Wang, H., Gu, Q., & Wang, H. (2016, July). Contextual bandits in a collaborative environment. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 529-538). ACM.
- Yang, D., Adamson, D., & Rosé, C. P. (2014, October). Question recommendation with constraints for massive open online courses. In *Proceedings of the 8th ACM Conference on Recommender systems* (pp. 49-56). ACM.
- Yu, K., Zhang, B., Zhu, H., Cao, H., & Tian, J. (2012, May). Towards personalized context-aware recommendation by mining context logs through topic models. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 431-443). Springer, Berlin, Heidelberg.
- Yuan, F., Guo, G., Jose, J. M., Chen, L., Yu, H., & Zhang, W. (2016, October). Lambdafm: learning optimal ranking with factorization machines using lambda surrogates. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 227-236). ACM.
- Yuan, J., Sivrikaya, F., Hopfgartner, F., Lommatzsch, A., & Mu, M. (2015). Context-aware LDA: Balancing relevance and diversity in TV content recommenders.
- Yujie, Z., & Licai, W. (2010, August). Some challenges for context-aware recommender systems. In *2010 5th International Conference on Computer Science & Education* (pp. 362-365). IEEE.
- Zheng, Y. (2015, March). A revisit to the identification of contexts in recommender systems. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion* (pp. 133-136). ACM.

- Zheng, Y. (2015, November). Context suggestion: Solutions and challenges. In 2015 IEEE International Conference on Data Mining Workshop (ICDMW) (pp. 1602-1603). IEEE.
- Zheng, Y., Burke, R., & Mobasher, B. (2012, September). Differential context relaxation for context-aware travel recommendation. In International Conference on Electronic Commerce and Web Technologies (pp. 88-99). Springer, Berlin, Heidelberg.
- Zheng, Y., Burke, R., & Mobasher, B. (2014, March). Splitting approaches for context-aware recommendation: An empirical study. In Proceedings of the 29th Annual ACM Symposium on Applied Computing (pp. 274-279). ACM.
- Zheng, Y., Mobasher, B., & Burke, R. (2014, August). Context recommendation using multi-label classification. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02 (pp. 288-295). IEEE Computer Society.
- Zheng, Y., Mobasher, B., & Burke, R. (2014, November). Deviation-based contextual SLIM recommenders. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (pp. 271-280). ACM.
- Zheng, Y., Mobasher, B., & Burke, R. (2014, October). CSLIM: Contextual SLIM recommendation algorithms. In Proceedings of the 8th ACM Conference on Recommender Systems (pp. 301-304). ACM.
- Zheng, Y., Mobasher, B., & Burke, R. (2015, June). Integrating context similarity with sparse linear recommendation model. In International Conference on User Modeling, Adaptation, and Personalization (pp. 370-376). Springer, Cham.
- Zheng, Y., Mobasher, B., & Burke, R. (2015, November). Carskit: A java-based context-aware recommendation engine. In 2015 IEEE International Conference on Data Mining Workshop (ICDMW) (pp. 1668-1671). IEEE.
- Zheng, Y., Mobasher, B., & Burke, R. (2015, November). Similarity-based context-aware recommendation. In International Conference on Web Information Systems Engineering (pp. 431-447). Springer, Cham.