



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

**Σύστημα συστάσεων επίγνωσης πλαισίου με εφαρμογή στον
τομέα της διασκέδασης στην πόλη της Αθήνας**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΝΙΚΟΛΑΟΣ Γ. ΤΖΑΓΚΑΡΑΚΗΣ

Επιβλέπων : Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων : Γεώργιος Αλεξανδρίδης

Εργαστηριακό και Διδακτικό Προσωπικό Ε.Μ.Π.

Αθήνα, Απρίλιος 2021



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Σύστημα συστάσεων επίγνωσης πλαισίου με εφαρμογή στον τομέα της διασκέδασης στην πόλη της Αθήνας

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΝΙΚΟΛΑΟΣ Γ. ΤΖΑΓΚΑΡΑΚΗΣ

Επιβλέπων : Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων : Γεώργιος Αλεξανδρίδης
Εργαστηριακό και Διδακτικό Προσωπικό Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 23η Απριλίου 2021.

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Στάμου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Απρίλιος 2021

.....
Νικόλαος Γ. Τζαγκαράκης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Νικόλαος Γ. Τζαγκαράκης, 2021.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Ένα σύστημα συστάσεων, ή ένα σύστημα προτάσεων, είναι μια υποκατηγορία συστήματος φιλτραρίσματος πληροφοριών που επιδιώκει να προβλέψει την «βαθμολογία» ή «προτίμηση» που θα έδειχνε ένας χρήστης σε ένα αντικείμενο. Τα συστήματα συστάσεων χρησιμοποιούνται σε διάφορα πεδία, με πιο χαρακτηριστικά παραδείγματα τη δημιουργία λιστών αναπαραγωγής για υπηρεσίες βίντεο και μουσικής, προτάσεις προϊόντων για διαδικτυακά καταστήματα ή προτάσεις περιεχομένου για πλατφόρμες κοινωνικών μέσων καθώς και προτάσεις τοποθεσιών που ένας χρήστης θα ενδιαφερόταν να επισκεφθεί. Αυτά τα συστήματα μπορούν να λειτουργούν χρησιμοποιώντας είσοδο ενός τύπου, όπως λ.χ. μουσική, ή πολλαπλών τύπων, όπως λ.χ. στην περίπτωση πλατφόρμας που προτείνει αντικείμενα όπως ειδήσεις, βιβλία και ερωτήματα αναζήτησης. Υπάρχουν επίσης δημοφιλή συστήματα συστάσεων για συγκεκριμένα θέματα, όπως εστιατόρια και online γνωριμίες. Στην περίπτωση των συστημάτων συστάσεων επίγνωσης πλαισίου δημιουργούνται πιο σχετικές προτάσεις προσαρμόζοντάς τις στην ειδική κατάσταση του εκάστοτε χρήστη.

Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη ενός συστήματος συστάσεων επίγνωσης πλαισίου χρησιμοποιώντας υπάρχουσες τεχνικές σύστασης σε συνδυασμό με μοντέλα μηχανικής μάθησης. Πιο συγκεκριμένα οι συστάσεις του συγκεκριμένου συστήματος αφορούν κυρίως την εστίαση (καφετέριες, εστιατόρια, μπαρ κλπ). Σκοπός του παραπάνω συνδυασμού είναι η εξάλειψη του φαινομένου της «ψυχρής εκκίνησης», το οποίο παρατηρείται στις περιπτώσεις εκείνες που ο χρήστης δεν έχει κάποια προηγούμενη αλληλεπίδραση με τα αντικείμενα, με αποτέλεσμα το σύστημα να μην είναι σε θέση να συνάγει τις προτιμήσεις του. Σε πολλές περιπτώσεις χρήστες με λίγες ή ακόμα και καθόλου αλληλεπιδράσεις αδυνατούν να λάβουν αξιολογες προτάσεις για αντικείμενα ή μέρη που τους ενδιαφέρουν. Το συγκεκριμένο πρόβλημα θα προσπαθεί να αντιμετωπιστεί σε αυτή την εργασία.

Επιπρόσθετα, στη συγκεκριμένη εργασία θα αναπτυχθεί, εκτός από αλγορίθμους συστάσεων, και μια πλατφόρμα με την οποία θα αλληλεπιδρά ο χρήστης. Μέσω αυτής θα συλλέγονται βασικές πληροφορίες που απαιτούνται για την λειτουργία του συστήματος, προσφέροντας ταυτόχρονα μια ευχάριστη εμπειρία. Κίνητρο είναι οι περισσότερες διεργασίες να είναι αυτοματοποιημένες. Στα συστήματα συστάσεων το περιβάλλον που αλληλεπιδρά ο χρήστης είναι αρκετά σημαντικός παράγοντας, αφού πέρα από τις προτάσεις, το περιβάλλον καθορίζει σε μεγάλο βαθμό το βαθμό ικανοποίησής του. Στόχος δεν είναι απλά να παρέχονται καλές προτάσεις, αλλά χρειάζεται επίσης αυτό να γίνεται με έναν ομαλό τρόπο.

Λέξεις κλειδιά

Συστήματα συστάσεων, Συστήματα επίγνωσης πλαισίου, Μηχανική μάθηση, Πρόβλημα ψυχρής εκκίνησης, Pre-Filtering, Post-Filtering

Abstract

A recommender system, or a recommendation system, is a subclass of information filtering systems that seeks to predict the "rating" or "preference" a user would give to an item. Recommender systems are used in a variety of domains, with commonly recognized examples the playlist generators of video and music services, product recommenders for online stores, or content recommenders for social media platforms and open web content recommenders, as well as places suggestions that a user would be interested in to visit. These systems can operate on a single input type, e.g. music, or multiple input types within and across platforms like news, books, and search queries. There are also popular recommender systems for specific topics like restaurants and online dating. Context-aware recommender systems generate more relevant recommendations by adapting them to the specific contextual situation of the user

This diploma thesis aims to the development of a context-aware recommendation system using existing recommendation techniques in combination with machine learning models. More specifically, this system mainly targets food and coffee services (cafes, restaurants, bars, etc.). The purpose of the proposed system is to eliminate the cold start problem which is observed in those cases where the user does not have any previous interaction with the items and as a consequence, the system is not able to infer his or her preferences. In many cases users with little or no interaction fail to receive meaningful recommendations for items or places that interested them. This problem is going to be addressed in the current thesis..

Additionally, in this work, apart from recommendation algorithms, a platform is also going to be developed with which the user will interact. Through the said platform, basic information is going to be collected, required for the operation of the system, while at the same time offering pleasant experience to the end user. The overall objective is to automate the recommendation process. In recommender systems, the the environment under which the user interacts with the system, is a very important factor, as, apart from the recommendations, the environment itself largely determines the degree of user satisfaction. The aim is not to just provide good suggestions, but to do so in an unobtrusive way.

Key words

Recommender systems, Context aware systems, Machine learning, Cold start problem, Pre-Filtering, Post-Filtering

Ευχαριστίες

Η παρούσα διπλωματική εργασία σηματοδοτεί την ολοκλήρωση των μεταπτυχιακών σπουδών μου. Συνεπώς, θα ήθελα να ευχαριστήσω θερμά όλους όσους με υποστήριξαν στην εκπόνηση αυτής της διπλωματικής αλλά και συνολικά με τις σπουδές μου.

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα αυτής της διατριβής κ. Ανδρέα-Γεώργιο Σταφυλοπάτη, Καθηγητή Ε.Μ.Π., για την ευκαιρία που μου προσέφερε να εκπονήσω τη συγκεκριμένη διπλωματική καθώς και για την εμπιστοσύνη που μου έδειξε. Παράλληλα, θα ήθελα να ευχαριστήσω τον συνεπιβλέποντα κ. Γεώργιο Αλεξανδρίδη, Ε.ΔΙ.Π. ΕΜΠ, για τη συνεχή υποστήριξη και καθοδήγησή του. Οι πολύτιμες συμβουλές του και οι χρήσιμες συζητήσεις που πραγματοποιήσαμε με βοήθησαν να ξεπεράσω σημαντικά εμπόδια που συνάντησα κατά την εκπόνηση της εργασίας. Επιπροσθέτως, θα ήθελα να ευχαριστήσω τους κ. κ. Στέφανο Κόλλια, Καθηγητή Ε.Μ.Π. και Γεώργιο Στάμου, που με τίμησαν με την παρουσία τους στην τριμελή επιτροπή εξέτασης.

Σε προσωπικό επίπεδο θα ήθελα πρωτίστως να ευχαριστήσω την οικογένειά μου και κυρίως τους γονείς μου. Συγκεκριμένα, ευχαριστώ από τα βάθη της καρδιάς μου τη μητέρα μου, Χρυσούλα Φασουλάκη, τον πατέρα μου, Γιώργο Τζαγκαράκη, για τη συνεχή ενθάρρυνση και υποστήριξή τους. Τέλος, θα ήθελα να ευχαριστήσω όλους μου τους φίλους, εντός και εκτός σχολής, που στάθηκαν στο πλευρό μου όλα αυτά τα χρόνια.

Νικόλαος Γ. Τζαγκαράκης,
Αθήνα, 23η Απριλίου 2021

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
Κατάλογος πινάκων	13
Κατάλογος σχημάτων	15
1. Εισαγωγή	17
1.1 Συστήματα συστάσεων και επίγνωση πλαισίου	17
1.2 Δομή της εργασίας	20
2. Συστήματα συστάσεων	23
2.1 Ορισμοί συστημάτων συστάσεων	23
2.2 Είδη και τεχνικές	25
2.2.1 Συνεργατική διήθηση	25
2.2.2 Φιλτράρισμα βάσει περιεχομένου	28
2.2.3 Υβριδικά συστήματα συστάσεων	29
2.2.4 Συστήματα συστάσεων επίγνωσης πλαισίου	31
2.2.5 Συστήματα συστάσεων πολλαπλών κριτηρίων	33
2.2.6 Συστήματα συστάσεων επίγνωσης ρίσκου	34
2.2.7 Συστήματα συστάσεων κινητών συσκευων	34
2.2.8 Συστήματα συστάσεων βασισμένα στη συνεδρία	34
2.2.9 Συστήματα συστάσεων ενισχυτικής μάθησης	34
2.3 Τρόπος αξιολόγησης	35
2.3.1 Offline αξιολόγηση	35
2.3.2 Online αξιολόγηση	36
3. Διαστάσεις επίγνωσης πλαισίου	37
3.1 Περιγραφή βασικών διαστάσεων	37
3.2 Παραδείγματα χρήσης επίγνωσης πλαισίου	38
3.2.1 Τεχνολογία πληροφορίας και τουρισμός	38
4. Περιγραφή θέματος	43
4.1 Η κεντρική ιδέα	43
4.2 LightFM	44
4.2.1 Γενική περιγραφή	44
4.2.2 Μοντέλο	45
4.3 Παραγοντοποίηση Πίνακα	47

4.4	Συνδυασμός εργαλείων	48
5.	Υλοποίηση	51
5.1	Το Σύνολο Δεδομένων	51
5.1.1	Google API	51
5.1.2	Crowling	54
5.1.3	Επεξήγηση Δεδομένων	54
5.2	Υλοποίηση ως Υπηρεσία	58
5.2.1	Αρχική σελίδα και συστάσεις	59
5.2.2	Αξιολογήσεις και χάρτες	61
5.3	Περιπτώσεις χρήσης	63
5.3.1	Χρονική στιγμή της ημέρας	63
5.3.2	Τοποθεσία χρήστη	64
5.3.3	Χρήστες	64
6.	Συμπεράσματα και Επεκτάσεις	67
6.1	Συμπεράσματα	67
6.2	Μελλοντικές Επεκτάσεις	68
	Βιβλιογραφία	71
	Παράρτημα	77
A.	Ευρετήριο Ακρωνυμίων και Συντμήσεων	77

Κατάλογος πινάκων

3.1 Ποσοστό διαθέσιμων δεδομένων ανά λειτουργία χρήστη για ένα σύνολο εγγεγραμμένων πελατών	40
---	----

Κατάλογος σχημάτων

1.1	Επένδυση επιχειρήσεων στην εξατομίκευση (Πηγή [1])	19
1.2	Υπερφόρτωση πληροφοριών [2]	20
2.1	Συνεργατική διήθηση [3]	26
2.2	Παράδειγμα βαθμολόγησης με εικονίδια	28
2.3	Φιλτράρισμα βάσει περιεχομένου [3]	29
2.4	Υβριδικό σύστημα συστάσεων [3]	31
2.5	Πολυδιάστατο μοντέλο για $User \times Item \times Time$ recommendation space [4]	32
3.1	Επωνυμίες που ανήκουν στην AccorHotels[5]	39
3.2	Περιβαλλον κρατήσεων της Accor[5]	40
3.3	Google Maps [6]	41
3.4	Αναζήτηση και φίλτρα [7]	41
5.1	Αρχιτεκτονική υλοποίησης	52
5.2	Χωρισμός Αττικής σε περιοχές	53
5.3	Ιστόγραμμα πλήθους κριτικών χρηστών	55
5.4	Ιστόγραμμα διακύμανσης κριτικών χρηστών	56
5.5	Πλήθος κριτικών ανά επιχείρηση	56
5.6	Κατηγορίες επιχειρήσεων	57
5.7	Επίπεδα τιμών επιχειρήσεων	58
5.8	Αρχική σελίδα	59
5.9	Αποτελέσματα συστάσεων	60
5.10	Χρήση χαρτών για προβολή και μετάβαση	62
5.11	Περιβάλλον αξιολόγησης	63

Κεφάλαιο 1

Εισαγωγή

Τα τελευταία χρόνια, τα συστήματα συστάσεων έχουν ενταχθεί αρκετά στη ζωή της πλειοψηφίας των ανθρώπων και μάλιστα αρκετοί τα χρησιμοποιούν ακόμα και σε περιπτώσεις που δεν το αντιλαμβάνονται. Συστάσεις υπήρχαν πάντα, απλά όχι με τη μορφή που έχουν πάρει σήμερα. Πιο συγκεκριμένα ενώ παλαιότερα οι συστάσεις γινόντουσαν μόνο μεταξύ ατόμων, πλέον μεγάλη είναι και η συμβολή του λογισμικού, αυτοματοποιώντας έτσι όλη τη διαδικασία.

Καθημερινά είναι σύνηθες φαινόμενο να γίνονται προτάσεις από άτομα του κοινωνικού μας περιγύρου για μια σειρά από ζητήματα, όπως λ.χ. την επίσκεψη σε ένα κατάστημα ή σε ένα εστιατόριο. Ειδικά στη δεύτερη περίπτωση, εδώ και χρόνια δημοσιεύονται λίστες που περιέχουν καταλόγους συστάσεων για πολλές κατηγορίες εστιατορίων, όπως είναι λ.χ. η Zagat [8] και η Michelin Red Guide [9]. Σήμερα, πέρα των προαναφερόμενων λιστών, αυτό τον ρόλο τον έχουν αναλάβει διάφορες εφαρμογές στο Διαδίκτυο που προσφέρουν ακριβώς αυτή την υπηρεσία. Πρόκειται, δηλαδή, για συστήματα συστάσεων που βοηθούν τους χρήστες στην περιήγηση στον συνεχώς αυξανόμενο αριθμό διαθέσιμων επιλογών. Κλασικό παράδειγμα ενός τέτοιου συστήματος είναι το σύστημα της διαδικτυακής πλατφόρμας ταινιών και σειρών Netflix [10], το οποίο βοηθάει τους χρήστες του να αποφασίσουν ποια ταινία η σειρά θα παρακολουθήσουν στη συνέχεια.

Γενικότερα, τα *συστήματα συστάσεων* (recommender systems - RS) μπορούν να ταξινομηθούν σε 3 κατηγορίες [11]: (i) *συνεργατικής διήθησης* (collaborative filtering - CF), (ii) *βάσει περιεχομένου* (content-based - CBF) και (iii) υβριδικά (hybrid). Η CF τεχνική μοιάζει με αυτό που αναφέρθηκε παραπάνω: συστάσεις από στόμα σε στόμα (φίλοι, οικογένεια), στην οποία οι απόψεις των άλλων χρηστών χρησιμοποιούνται για τον προσδιορισμό της συνάφειας μιας σύστασης. Σε αυτήν την περίπτωση, το σύστημα συλλογικής σύστασης χρησιμοποιεί τις βαθμολογίες που παρέχονται από τους χρήστες του είτε για να προτείνει ένα ενδιαφέρον αντικείμενο είτε για τον εντοπισμό παρόμοιων χρηστών. Για παράδειγμα, το Netflix χρησιμοποιεί αυτήν την τεχνική για να προτείνει βίντεο με υψηλή βαθμολογία από άτομα που, στο παρελθόν, αξιολόγησαν βίντεο σε παρόμοιο τρόπο με τον χρήστη. Η σύσταση βάσει περιεχομένου επικεντρώνεται στη χρήση του περιεχομένου ενός αντικειμένου για να επιβεβαιώσει τη σχετικότητά του. Με άλλα λόγια, μια εφαρμογή προτάσεων μουσικής που χρησιμοποιεί την προσέγγιση αυτή μπορεί να χρησιμοποιήσει το είδος ενός μουσικού κομματιού για να προτείνει άλλες μουσικές επιλογές που μοιράζονται το ίδιο είδος. Τα υβριδικά συστήματα χρησιμοποιούν και τις δύο τεχνικές κατά τη δημιουργία μιας σύστασης. Παραδείγματος χάριν, το ίδιο σύστημα συστάσεων μουσικής που χρησιμοποιεί την προσέγγιση βάσει περιεχομένου, μπορεί επίσης να επεκταθεί ώστε να ενσωματωθεί η CF τεχνική. Σε αυτή την περίπτωση, οι συστάσεις περιέχουν κομμάτια μουσικής του ίδιου είδους που βρήκαν ενδιαφέροντα χρήστες με παρόμοιο μουσικό γούστο.

1.1 Συστήματα συστάσεων και επίγνωση πλαισίου

Έχει καταβληθεί μεγάλη προσπάθεια για την ανάπτυξη νέων μεθόδων σύστασης που εστιάζουν τόσο στο αντικείμενο όσο και στο χρήστη [2, 12, 13]. Ωστόσο, σε πολλές εφαρμογές δεν λαμβάνεται υπόψη το πλαίσιο υπό το οποίο αυτές πραγματοποιούνται. Για παράδειγμα, μια ταξιδιωτική εφαρμογή μπορεί να έχει συναγάγει ορθά ότι σε έναν επισκέπτη αρέσει το σκι και να του προτείνει σχετικά πακέτα διακοπών. Ωστόσο, αυτές οι προτάσεις πιθανότητα να μην είναι ορθές αν γίνουν λ.χ. κατά

τη διάρκεια του καλοκαιριού. Συνεπώς, είναι σημαντικό να ενσωματωθεί το πλαίσιο επίγνωσης στη διαδικασία σύστασης.

Τα τελευταία χρόνια, με τον πολλαπλασιασμό των έξυπνων συσκευών (τηλεφώνων, ρολογιών κλπ), οι εφαρμογές έχουν πρόσβαση σε όλο και περισσότερες πληροφορίες επίγνωσης πλαισίου από τους χρήστες. Παρόλα αυτά, τα συστήματα σύστασης δεν χρησιμοποιούν ρητά αυτές τις πληροφορίες κατά την παροχή συστάσεων. Ένας τρόπος αντιμετώπισης αυτού του ζητήματος είναι να δημιουργηθούν νέα μοντέλα που μπορούν να ενσωματώσουν αυτή τη πληροφορία του χρήστη, βελτιώνοντας έτσι την ποιότητα των συστάσεων τους [14, 15].

Αν και αυτές οι προσπάθειες είναι πολλά υποσχόμενες, πολλά ζητήματα προκύπτουν κατά την ενσωμάτωση του πλαισίου στη διαδικασία σύστασης. Το πρώτο ζήτημα είναι ο όγκος και η ταχύτητα με την οποία τα δεδομένα επίγνωσης πλαισίου δημιουργούνται, καθιστώντας σημαντική πρόκληση την ενσωμάτωσή τους στα συστήματα συστάσεων. Ειδικότερα, η υπολογιστική ισχύς που απαιτείται για την εκπαίδευση ενός μοντέλου χρησιμοποιώντας τέτοια σύνολα δεδομένων είναι τεράστια [16]. Επομένως, στην παρούσα εργασία αξιοποιούνται τεχνικές μηχανικής μάθησης για να καταναίμει το φορτίο σε διάφορα μοντέλα αντί για ένα, μειώνοντας τον αριθμό των καταχωρήσεων που χρησιμοποιούνται για την εκπαίδευση κάθε μοντέλου.

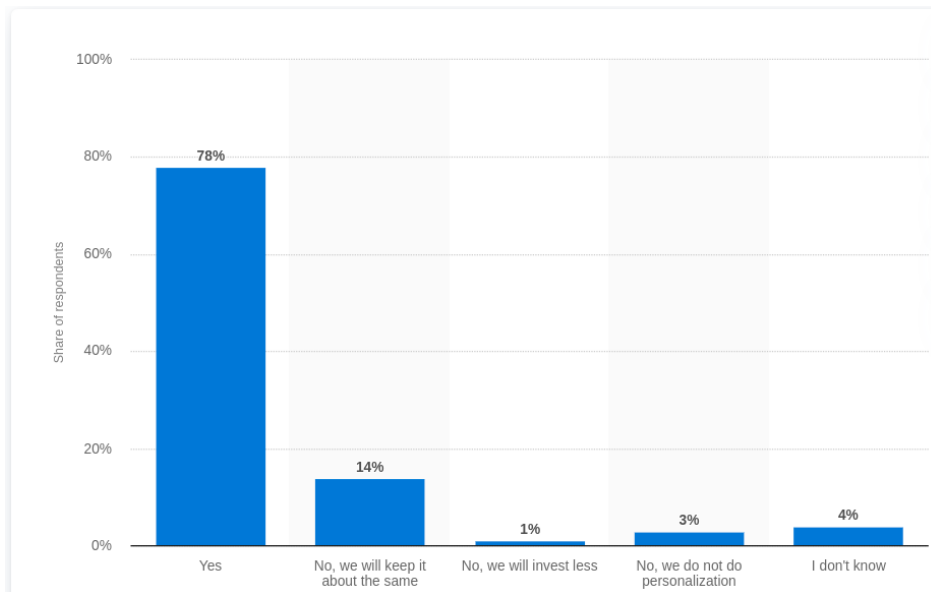
Επιπλέον, η δυσκολία που παρουσιάζεται στην ενσωμάτωση του περιβάλλοντος χρήστη σε νέα μοντέλα, παρουσιάζει μια ακόμη πρόκληση επειδή προσθέτει νέες διαστάσεις στο μοντέλο. Αυτή η εργασία προτείνει μια αρχιτεκτονική που βασίζεται σε υπάρχουσες CF τεχνικές για την παροχή συστάσεων επίγνωσης πλαισίου. Υποθέτοντας ότι τα υπάρχοντα συλλογικά μοντέλα δεν λαμβάνουν υπόψιν το πλαίσιο του χρήστη, η αρχιτεκτονική χρησιμοποιεί τις πληροφορίες επίγνωσης πλαισίου ως κριτήρια φιλτραρίσματος.

Ένα άλλο πρόβλημα είναι η πολυπλοκότητα της κατανόησης και της εφαρμογής νέων συστημάτων συστάσεων, καθώς και το σχετικό κόστος. Οι μεγάλες εταιρείες έχουν αρκετούς πόρους για να επενδύσουν στα δικά τους μοντέλα και ιδιαιτερότητες. Για παράδειγμα, το 2007 η Netflix πλήρωσε 1 εκατομμύριο δολάρια ώστε να βελτιωθεί ο αλγόριθμος Cinematch κατά 10, 06% [17]. Ταυτόχρονα, μικρότερες εταιρείες, προγραμματιστές και ερευνητές γενικά αγωνίζονται να βρουν μια ευέλικτη λύση που αφαιρεί από αυτούς την ανάγκη κατανόησης των εσωτερικών λειτουργιών κάθε μοντέλου. Επομένως, ένα άλλο κίνητρο αυτής της εργασίας είναι να παρουσιάσει ένα σύστημα που επιτρέπει στους μικρούς παίκτες να εφαρμόσουν τα δικά τους συστήματα συστάσεων που χρησιμοποιούν υπάρχοντα μοντέλα που ταιριάζουν καλύτερα στις ανάγκες τους.

Από τότε που το Διαδίκτυο χρησιμοποιήθηκε για πρώτη φορά αποτελεί αναπόσπαστο κομμάτι της κοινωνίας και του τρόπου ζωής ενός μεγάλου ποσοστού ανθρώπων. Στατιστικά στοιχεία που παρουσιάζονται από τη Διεθνή Ένωση Τηλεπικοινωνιών δείχνουν ότι ο αριθμός των παγκόσμιων χρηστών του Διαδικτύου έχει αυξηθεί αρκετά και απαρτίζεται από το 59.5% της ανθρωπότητας. Αυτό το ποσοστό μεταφράζεται σε 4.66 δισεκατομμύρια ανθρώπους, και παρουσιάζει μια αυξητική τάση της τάξης 7.3% ετησίως (316 εκατομμύρια). Κάθε χρήστης κατά μέσο όρο ξοδεύει περίπου 7 ώρες ημερησίως, ενώ το 92.6% αυτών, για την περιήγηση του χρησιμοποιεί κινητές συσκευές. Αυτό, προέκυψε από την εμπορευματοποίηση του Διαδικτύου και την ψηφιοποίηση της κοινωνίας, που τελικά οδήγησε σε αυτό που σήμερα αναφερόμαστε ως Big Data - μεγάλες δεξαμενές δεδομένων που μπορούν να καταγράφουν, κοινοποιηθούν, επεξεργαστούν, αποθηκευτούν και αναλυθούν [18].

Καθώς οι χρήστες περνάνε όλο και περισσότερο χρόνο στο Διαδίκτυο τόσο στην επαγγελματική όσο και στην ιδιωτική τους ζωή, η ποσότητα των διαθέσιμων δεδομένων πολλαπλασιάζεται με ταχείς ρυθμούς. Σύμφωνα με τον Merivis[19], 1.2 zettabytes νέων δεδομένων δημιουργούνται κάθε χρόνο. Οι εταιρείες σε όλο τον κόσμο συλλέγουν τεράστιες ποσότητες δεδομένων για τους πελάτες τους με την ελπίδα ότι θα μπορέσουν να βελτιώσουν τις εμπειρίες των χρηστών τους και τελικά να αυξήσουν τα έσοδά τους. Θα περίμενε κανείς αυτό να είναι επωφελές για όλους, αφού οι εταιρείες θα είναι σε θέση να στοχεύουν τους πελάτες τους με μεγαλύτερη ακρίβεια και οι χρήστες θα έχουν πρόσβαση σε περισσότερες πληροφορίες. Ωστόσο, αυτό απέχει από την αλήθεια.

Για να γίνει ακόμη πιο κατανοητή η μεγάλη σημασία της εξατομίκευσης που συνεπάγεται με συστήματα συστάσεων παρουσιάζεται στο Σχήμα 1.1. Σχετική έρευνα έγινε στις επιχειρήσεις το 2020



Σχήμα 1.1: Επένδυση επιχειρήσεων στην εξατομίκευση (Πηγή [1])

[1], με το ερώτημα να είναι αν οι επιχειρήσεις σκοπεύουν να επενδύσουν στην εξατομίκευση. Η συντριπτική πλειοψηφία του 78% απάντησε θετικά, δείχνοντας έτσι την κομβική σημασία της εξατομίκευσης στην επιχειρήσεις. Το 14% σημείωσε ότι θα κρατήσει τις επενδύσεις του στο συγκεκριμένο θέμα αμετάβλητες. Ενώ τέλος το υπόλοιπο 7% μοιράζεται σε εκείνες τις επιχειρήσεις που είτε δεν έχουν καθόλου ή θα μειώσουν την εξατομίκευση καθώς και σε εκείνες που δεν γνώριζαν να απαντήσουν.

Πράγματι, στην εποχή των Big Data, γίνεται όλο και πιο δύσκολο για τους απλούς Χρήστες Διαδικτύου να πλοηγηθούν με επιτυχία στη θάλασσα πληροφοριών στον Παγκόσμιο Ιστό αποτελεσματικά. Για παράδειγμα, ένας χρήστης που περιηγείται σε μια διαδικτυακή υπηρεσία ταινιών δεν επιθυμεί να πάει μέσα από δεκάδες, ή ίσως εκατοντάδες, μη ενδιαφέρουσες ταινίες πριν βρει μια που θα του κεντρίσει το ενδιαφέρον. Η ογκώδης ποσότητα επιλογών καθιστά πιο δύσκολο για τον χρήστη να βρει ακριβώς το επιθυμητό στοιχείο, μια κατάσταση που ονομάζεται υπερφόρτωση πληροφοριών [2] (διάγραμμα 1.2). Αυτό είναι ήδη πραγματικότητα σε διάφορους τομείς, για παράδειγμα πολυμέσα και τουρισμός. Για να είναι επιτυχημένοι και ανταγωνιστικοί οι εμπλεκόμενοι σε αυτούς τους τομείς, χρειάζονται οι χρήστες τους να είναι ικανοποιημένοι με τις υπηρεσίες τους. Τροφοδοτούμενοι από αυτήν την ανάγκη για την καταπολέμηση της υπερφόρτωσης πληροφοριών, τα τελευταία χρόνια παρατηρήθηκε αύξηση της έρευνας στον τομέα της σύστασης συστημάτων [20].

Στον τομέα του τουρισμού οι περισσότεροι άνθρωποι γνωρίζουν ιστότοπους όπως η Expedia, η Hotels και το TripAdvisor. Αυτοί οι ιστότοποι περιλαμβάνουν όλες τις μορφές συστήματος σύστασης για την παρουσίαση του χρήστη με δημοφιλή προϊόντα και καλές προσφορές, αλλά προς το παρόν δεν παρέχουν στον χρήστη πραγματικά εξατομικευμένες συστάσεις. Αντί να βλέπει τον χρήστη ως άτομο με τα δικά του ενδιαφέροντα, υποθέτει ότι όλοι οι χρήστες είναι παρόμοιοι. Ένας λόγος για αυτό είναι η εγγενής πολυπλοκότητα αντικειμένων σε αυτόν τον τομέα. Οι απόψεις των ανθρώπων για αντικείμενα όπως ξενοδοχεία και εστιατόρια εξαρτάται από μεγάλο αριθμό παραγόντων, καθιστώντας δύσκολο να προσφερθούν καλές προσωπικές προτάσεις.

Μερικοί λόγοι για τους οποίους οι καθιερωμένες τεχνικές προτάσεων δεν μπορούν να εφαρμοστούν άμεσα στον τομέα του τουρισμού αναφέρονται από τους Felfernig, Gordea, Jannach, Terran και Zanker [21]. Το συλλογικό φιλτράρισμα είναι μια ευρέως χρησιμοποιούμενη προσέγγιση σε συστήματα σύστασης, αλλά λειτουργεί καλύτερα όταν υπάρχει μια ευρεία κοινότητα χρηστών και κάθε χρήστης έχει ήδη αξιολογήσει σημαντικό αριθμό αντικειμένων. Δεδομένου του γεγονότος ότι οι δραστηριότητες προγραμματισμού ταξιδιών είναι αισθητά λιγότερο συνηθισμένες από ότι για παράδειγμα η αγορά ενός βιβλίου ή η παρακολούθηση μιας ταινίας και τα ίδια τα αντικείμενα μπορεί να έχουν

πολύ πιο περίπλοκη δομή, κάνοντας έτσι δύσκολο να καθοριστεί προφίλ χρηστών. Επιπλέον, μία ρύθμιση ταξιδιού μπορεί να αποτελείται από πολλές, ανεξάρτητα διαμορφώσιμες υπηρεσίες. Συνήθως, μόνο προκαθορισμένα πακέτα είναι διαθέσιμα στο διαδίκτυο, για παράδειγμα "all-inclusive" ή "πτήση και ξενοδοχείο". Υπάρχει επίσης η έννοια της επίγνωσης πλαισίου [22], ή σε ποια κατάσταση βρίσκεται ο χρήστης όταν ζητάει συστάσεις. Οι χρήστες αλληλεπιδρούν με συστήματα σε ένα συγκεκριμένο πλαίσιο και οι αξιολογήσεις για ένα στοιχείο σε ένα περιβάλλον ενδέχεται να είναι εντελώς διαφορετικές σε ένα άλλο. Για παράδειγμα ένας άνδρα που πηγαίνει σε ένα σπορ μπαρ με μια ομάδα φίλων συγκριτικά με μια την σχέση του. Είναι πιθανό ότι ο χρήστης θα έδινε καλύτερη βαθμολογία στην πρώτη περίπτωση.



Σχήμα 1.2: Υπερφόρτωση πληροφοριών [2]

Το Harvard Business Review [23] έκανε μια ισχυρή δήλωση περιγράφοντας τα συστήματα συστάσεων ως την πιο σημαντική αλγοριθμική διάκριση μεταξύ «γεννημένων ψηφιακών» επιχειρήσεων και εταιρειών παλαιού τύπου. Η HBR περιέγραψε επίσης τον ενάρτετο επιχειρηματικό κύκλο που μπορούν να δημιουργήσουν: όσο περισσότερα άτομα χρησιμοποιούν το σύστημα συστάσεων μιας εταιρείας, τόσο πιο πολύτιμα γίνονται και τόσο περισσότερα άτομα τα χρησιμοποιούν.

Μια άλλη οπτική είναι να εξεταστούν τα συστήματα σύστασης, όχι ως απλά τρόπος πώλησης περισσότερων διαδικτυακών προϊόντων, αλλά ως ανανεώσιμο πόρο για τη συνεχή βελτίωση των πληροφοριών των πελατών και των ιδεών της επιχείρησης επίσης. Είναι γνωστό ότι πολλές εταιρείες παλαιού τύπου έχουν επίσης τόνους χρηστών και επομένως τόνους δεδομένων. Ο λόγος που ο ενάρτετος κύκλος τους δεν έχει εξελιχθεί τόσο πολύ όσο οι Amazon, Netflix και Spotify οφείλεται στην έλλειψη γνώσεων σχετικά με το πώς να μετατρέψουν τα δεδομένα χρήστη τους σε ενεργές πληροφορίες, οι οποίες μπορούν στη συνέχεια να χρησιμοποιηθούν για τη βελτίωση του προϊόντος ή των υπηρεσιών τους. Για παράδειγμα, η Google διαθέτει τεράστια δεδομένα, οπότε μπορεί να δημιουργήσει μια ακριβή μηχανή αναζήτησης ως προϊόν (A). Αυτό το προϊόν επιτρέπει στην Google να αποκτήσει περισσότερους χρήστες (B), οι οποίοι στη συνέχεια δημιουργούν περισσότερα δεδομένα στην Google (C). Επίσης κοιτάζοντας το Netflix, δείχνει πόσο σημαντικό είναι αυτό, καθώς το 80% αυτού που παρακολουθούν οι χρήστες του προέρχονται από κάποιες προτάσεις. Οι διαφημίσεις μετατοπίζονται όχι μόνο στην ψηφιοποίηση, αλλά και στην εξατομίκευση, που τροφοδοτούνται από ολοένα και πιο εξελιγμένους αλγόριθμους και προγνωστικά μοντέλα που αναλύουν δεδομένα συναλλαγών και τάσεις ψηφιακών μέσων (για παράδειγμα, ποια θέματα είναι δημοφιλή στα κοινωνικά δίκτυα) Το 2015 σε ένα από τα άρθρα τους [24] αναφέρθηκε ότι ο συνδυασμός εξατομίκευσης και συστάσεων έχει ως αποτέλεσμα να εξοικονομούν περισσότερο από 1 δισεκατομμύριο δολάρια το χρόνο. Σε παρόμοια κατάσταση βρίσκεται και η Amazon αφού το 35% των προϊόντων που αγοράζουν οι χρήστες προέρχεται από συστάσεις. Ομοίως στο Airbnb η κατάταξη αναζήτησης και οι παρόμοιες καταχωρίσεις ορίζουν το 99% όλων των αλλαγών σε κρατήσεις.

1.2 Δομή της εργασίας

Παρακάτω παρουσιάζεται μια συνοπτική περιγραφή σχετικά με την οργάνωση της παρούσας διπλωματικής εργασίας. Καταρχήν, εννοιολογικά η εργασία χωρίζεται σε δύο μέρη, ένα θεωρητικό και

ένα πρακτικό. Στο θεωρητικό μέρος παρουσιάζονται και αναλύονται διάφορες τεχνικές συστάσεων ενώ στο πρακτικό μέρος γίνεται ανάλυση σχετικά με την υλοποίηση και πως αυτή επιτεύχθηκε. Πιο συγκεκριμένα:

- Κεφάλαιο 1: Εισαγωγή και συνοπτική περιγραφή του σκοπού της εργασίας. Πιο συγκεκριμένα στο κεφάλαιο αυτό δίνεται μια πρώτη εικόνα σχετικά με τα συστήματα σύστασης και πως αυτά μπορούν να επηρεάσουν την καθημερινότητα των ανθρώπων. Παρουσιάζονται συνοπτικά τα προβλήματα που καλούνται να λύσουν τέτοιου είδους συστήματα, καθώς και τα ερεθίσματα που οδήγησαν στην συγκεκριμένη διπλωματική εργασία. Επίσης αναλύεται πως τέτοιου είδους συστήματος συμβάλλουν στη βελτίωση των ίδιων των επιχειρήσεων και τη μεγάλη συνεισφορά τους στην αύξηση των κερδών.
- Κεφάλαιο 2: Αναλυτική περιγραφή όλων των θεωρητικών ζητημάτων και τεχνικών που αφορούν τα συστήματα συστάσεων. Σε αυτό το κεφάλαιο παρέχονται όλοι οι σχετικοί ορισμοί καθώς και οι υπάρχον κατηγοριοποιήσεις τους αλγοριθμικά αλλά και μέσω παραδειγμάτων. Επίσης παρέχονται λεπτομέρειες σχετικά με την ειδική υποκατηγορία που αφορά αυτήν την εργασία, το πλαίσιο επίγνωσης πλαισίου. Τέλος εξηγείται αναλυτικά ο τρόπος που παρέχονται οι αξιολογήσεις των χρηστών και πως αυτές αλληλεπιδρούν με το σύστημα.
- Κεφάλαιο 3: Αυτό το κεφάλαιο αφορά αποκλειστικά την επίγνωση πλαισίου. Προσφέρονται πληροφορίες σχετικά με τις διαστάσεις και τις κατηγορίες που την διέπουν. Κάθε κατηγορία αναλύεται ξεχωριστά, τονίζοντας τι πληροφορίες μπορούν να αντληθούν καθώς και πως αυτές μπορούν να χρησιμοποιηθούν από το σύστημα ώστε οι προτάσεις να γίνουν ακόμα πιο εξατομικευμένες και να οδηγήσουν τελικά σε καλύτερα αποτελέσματα, αντιμετωπίζοντας τον κάθε χρήστη ξεχωριστά.
- Κεφάλαιο 4: Στο κεφάλαιο αυτό παρουσιάζεται πως η συγκεκριμένη εργασία καλείται να αντιμετωπίσει όλες τις προκλήσεις που έχουν αναφερθεί στα προηγούμενα κεφάλαια. Γίνεται αναφορά στις τεχνικές που θα χρησιμοποιηθούν και στο βασικό εργαλείο της εργασίας, το *lightfm*. Επιπροσθέτως επεξηγούνται βασικές τεχνικές των συστημάτων συστάσεων όπως την *παραγοντοποίηση πίνακα* (Matrix factorization - MF) και πως όλες αυτές συνδυάζονται ώστε να οδηγήσουν στο τελικό αποτέλεσμα.
- Κεφάλαιο 5: Στο κεφάλαιο 5 εξ ολοκλήρου αφορά την υλοποίηση ολόκληρου του συστήματος. Ξεκινάει περιγράφοντας την διαδικασία συλλογής των δεδομένων που χρησιμοποιήθηκαν, αναφέροντας ταυτόχρονα τα προβλήματα που προέκυψαν σε αυτή τη διαδικασία. Στη συνέχεια ακολουθεί η περιγραφή όλων των υπηρεσιών που παρέχονται, σε αλγοριθμικό επίπεδο αλλά και ως πλατφόρμα (περιβάλλον αλληλεπίδρασης χρήστη). Τέλος επεξηγούνται όλα τα *use cases* που καλύπτει η συγκεκριμένη εργασία, και τα αποτελέσματα που προέκυψαν σε κάθε ένα από αυτά.
- Κεφάλαιο 6: Τέλος, ακολουθεί το κεφάλαιο με τα τελικά συμπεράσματα και τις μελλοντικές επεκτάσεις που μπορεί να έχει αυτή η εργασία. Γίνεται ένας επίλογος συγκριτικά με την έρευνα που προηγήθηκε συγκρίνοντας την με τα αποτελέσματα που προέκυψαν. Όσο αφορά τις μελλοντικές επεκτάσεις, αναφέρεται τι θα άξιζε να μελετηθεί περαιτέρω ενώ δεν συμπεριλήφθηκε στην εν λόγω εργασία στα πλαίσια του μεταπτυχιακού προγράμματος.

Κεφάλαιο 2

Συστήματα συστάσεων

Τα συστήματα προτάσεων είναι ένα ουσιαστικό εργαλείο για τους χρήστες, καθώς μπορούν να βοηθήσουν στην εύρεση των πιο σχετικών στοιχείων και να προτείνουν τα πιο ενδιαφέροντα αντικείμενα που ταιριάζουν περισσότερο σε ένα χρήστη. Σήμερα, οι περισσότερες εφαρμογές συστημάτων συστάσεων βρίσκονται σε τομείς ψυχαγωγίας όπως μουσική, βίντεο και ταινίες [25]. Υπήρξε σημαντική εξέλιξη σε αυτόν τον τομέα, εστιάζοντας σε μεγάλο βαθμό στις δύο πιο δημοφιλείς τεχνικές τέτοιου είδους συστημάτων: συνεργατικό φιλτράρισμα [26] και φιλτράρισμα βάσει περιεχομένου [27]. Στην τεχνική CF, τα συστήματα σύστασης βασίζονται στη «σοφία του πλήθους» και στην υπόθεση ότι παρόμοιοι χρήστες θα πρέπει να έχουν τις ίδιες προτιμήσεις, και ως εκ τούτου δίνουν την ίδια βαθμολογία σε παρόμοια στοιχεία. Η CBF, από την άλλη πλευρά, βασίζεται σε περιγραφές στοιχείων σε σχέση με τους χρήστες, καθώς και στις προτιμήσεις τους, καθώς σχετίζονται με τεχνικές ανάκτησης πληροφοριών και φιλτραρίσματος, π.χ. *term frequency-inverse document frequency* - TF-IDF, το οποίο παρουσιάζει μια βαθμολογία βάρους όρων που εμφανίζονται στο έγγραφο και την αναπαράστασή τους (διανυσματικός χώρος) [28]. Για παράδειγμα, [29] εφάρμοσε το TF-IDF για να δημιουργήσει ένα εξατομικευμένο προφίλ χρήστη στον τομέα της ηλεκτρονικής μάθησης.

Αυτό το κεφάλαιο εισάγει το θεωρητικό υπόβαθρο στα συστήματα σύστασης, σκιαγραφεί τους ορισμούς που σχετίζονται με την κύρια συμβολή αυτής της έρευνας. Οι προκλήσεις που αντιμετωπίζουν τα συστήματα σύστασης περιγράφονται λεπτομερώς. Στη συνέχεια, το κεφάλαιο παρέχει το ιστορικό του πίσω από αυτά τα συστήματα. Οι βασικές μέθοδοι συστημάτων συστάσεων δίνονται στη συνέχεια με περισσότερες λεπτομέρειες και συγκρίνονται ως προς την ακρίβεια και την απόδοση, προκειμένου να δικαιολογηθεί η χρήση του αλγόριθμου που χρησιμοποιείται σε αυτήν την εργασία, και για να απεικονιστεί κάθε μέθοδος και να γίνει κατανοητού πόσο διαφορετικές είναι μεταξύ τους.

2.1 Ορισμοί συστημάτων συστάσεων

Τα συστήματα συστάσεων μπορούν να οριστούν ως προσαρμόσιμα εργαλεία που μπορούν να βοηθήσουν τους χρήστες στην αναζήτηση, φιλτράρισμα, ταξινόμηση πληροφοριών καθώς και στο να βρουν σχετικά αντικείμενα που πιθανώς να τους ενδιαφέρουν [20]. Μπορούν να βοηθήσουν τους χρήστες προτείνοντας προϊόντα παρόμοια με αυτά βλέπουν. Περιλαμβάνουν τεχνικές που παρέχουν προτάσεις για αντικείμενα που είναι πιο πιθανό να ενδιαφέρουν έναν συγκεκριμένο χρήστη [4]. Τέτοια συστήματα είναι πράκτορες λογισμικού που προκαλούν τα ενδιαφέροντα και τις προτιμήσεις των μεμονωμένων καταναλωτών και παρέχουν συστάσεις ανάλογα. Έχουν τη δυνατότητα να υποστηρίξουν και να βελτιώσουν την ποιότητα των αποφάσεων που λαμβάνουν οι καταναλωτές κατά την αναζήτηση και την επιλογή προϊόντων στο Διαδίκτυο [30].

Τα συστήματα συστάσεων ερευνήθηκαν για πρώτη φορά στα μέσα της δεκαετίας του 1990, βασισμένοι στην ιδέα ότι οι χρήστες μοιράζονται παρόμοια στοιχεία ή απόψεις, βοηθώντας έτσι να κάνουν προτάσεις σε άλλους [31]. Η πιο συνηθισμένη τεχνική ήταν ο υπολογισμός των αξιολογήσεων για στοιχεία που δεν είχαν ακόμη βαθμολογηθεί από κάποιο συγκεκριμένο χρήστη. Ο Malone [32] παρέχει μια ιδέα για το πώς η τεχνητή νοημοσύνη βοηθά τους ανθρώπους να μοιράζονται σχετικές πληροφορίες, ενώ στο [20] οι συγγραφείς επεσήμαναν ότι στην πραγματικότητα οι άνθρωποι συχνά ζητούν από άλλους να προτείνουν κάτι με βάση τις δικές τους εμπειρίες. Στο [33] περιγράφεται το

EntreeC ως σύστημα που εξατομικεύει προτάσεις βάσει των ενδιαφερόντων του χρήστη. Το σύστημα μπορεί έτσι να οδηγήσει αποτελεσματικά τους χρήστες σε ένα ενδιαφέρον προϊόν. Για να αναπτύξουν ένα τέτοιο σύστημα σύστασης, οι προγραμματιστές έχουν εφαρμόσει πολλά συστήματα με διαφορετικούς τομείς για να βελτιώσουν τις προτάσεις. Ωστόσο, παραμένουν ορισμένοι περιορισμοί που πρέπει να εξαιρεθούν προκειμένου να βελτιωθεί η ποιότητα των παρεχόμενων συστάσεων.

Τα συστήματα προτάσεων προσπαθούν να προβλέψουν ποια είναι τα πιο κατάλληλα προϊόντα ή υπηρεσίες, με βάση τις προτιμήσεις και τους περιορισμούς του χρήστη. Προκειμένου να ολοκληρωθεί μια τέτοια υπολογιστική εργασία, τα συστήματα σύστασης συλλέγουν πληροφορίες από τους χρήστες σχετικά με τις προτιμήσεις τους, οι οποίες είτε εκφράζονται ρητά (π.χ. ως βαθμολογίες για αντικείμενα) ή συνάγεται από την ερμηνεία των ενεργειών του χρήστη. Για παράδειγμα, το σύστημα σύστασης μπορεί να θεωρήσει την πλοήγηση σε μια συγκεκριμένη σελίδα προϊόντος ως σιωπηρή ένδειξη προτίμησης για τα στοιχεία που εμφανίζονται σε αυτήν τη σελίδα [4].

Ο στόχος ενός συστήματος σύστασης είναι να παράγει προτάσεις σε μια συλλογή χρηστών για αντικείμενα ή προϊόντα που μπορεί να τους ενδιαφέρουν[34]. Προτάσεις βιβλίων στο amazon, ή ταινίες στο netflix είναι δύο από τα πολλά παραδείγματα που συναντά κάποιος στην καθημερινότητα του. Κάθε τέτοιο σύστημα διαφέρει αρκετά στην υλοποίηση αφού εξαρτάται σε πολύ μεγάλο βαθμό από τα διαθέσιμα δεδομένα για τους χρήστες αλλά και τα προϊόντα. Για παράδειγμα στην περίπτωση του netflix η συλλογή δεδομένων επιτυγχάνεται από τα likes και dislikes του χρήστη για τις ταινίες που παρακολούθησαν δηλώνοντας έτσι αν η αλληλεπίδρασή ήταν θετική ή αρνητική. Το ίδιο συμβαίνει και στην περίπτωση ενός συστήματος που προτείνει εστιατόρια. Η βαθμολογία του χρήστη για το εκάστοτε εστιατόριο που επισκέφθηκε εκφράζει αν η εμπειρία του ήταν ευχάριστη ή δυσάρεστη. Άλλες πηγές πληροφοριών περιλαμβάνουν δημογραφικές βάσεις δεδομένων, κοινωνικά δίκτυα και δεδομένα από συσκευές *Internet of Things* (IoT), με το τελευταίο να είναι μια νέα τάση στην ανάπτυξη συστημάτων συστάσεων [35]. Η χρήση συσκευών IoT επιτρέπει την απόκτηση και τη χρήση πληροφοριών επίγνωσης πλαισίου με διακριτικό για τους χρήστες τρόπο, κάτι που είναι επιθυμητό χαρακτηριστικό στα συστήματα συστημάτων επόμενης γενιάς [2].

Ένα σύστημα σύστασης έχει δύο κύρια χαρακτηριστικά που χρησιμοποιούνται ως είσοδος στο σύστημα: χρήστης και αντικείμενο. Οι χρήστες δίνουν τις αξιολογήσεις / τους αρέσει σε ένα συγκεκριμένο αντικείμενο / προϊόν. Οι χρήστες υποδεικνύονται με $U = \{u_1, u_2, \dots, u_n\}$, όπου n είναι ο αριθμός των χρηστών που χρησιμοποιούν το σύστημα συστάσεων. Τα αντικείμενα επισημαίνονται με $I = \{i_1, i_2, \dots, i_m\}$, όπου m αντιπροσωπεύει τον αριθμό των αντικειμένων που αξιολογούνται. Συνήθως, κάθε χρήστης έχει ένα προφίλ που περιέχει μια λίστα αντικειμένων. Αυτή η λίστα αντιπροσωπεύει τα στοιχεία που έχει αξιολογήσει ο χρήστης στο παρελθόν. Το αντικείμενο έχει επίσης κάποιο περιεχόμενο. για παράδειγμα, όσον αφορά τις ταινίες, αυτό θα μπορούσε να είναι είδος, σκηνοθέτης, ηθοποιοί και ούτω καθεξής.

Τέλος, τα στοιχεία που έχουν βαθμολογηθεί από δύο χρήστες u και v , για παράδειγμα, I_u και I_v , είναι μια σημαντική έννοια στην παρουσιαζόμενη μέθοδο, και το $I_{u,v}$ έχει χρησιμοποιηθεί για να υποδηλώσει αυτή η έννοια. Με παρόμοιο τρόπο, το $U_{i,j}$ χρησιμοποιείται για να δηλώσει το σύνολο των χρηστών που έχουν βαθμολογήσει και τα δύο στοιχεία i και j .

Δύο από τα πιο σημαντικά προβλήματα που σχετίζονται με τα συστήματα σύστασης είναι το πρόβλημα πρόβλεψης, το οποίο αφορά την εκτίμηση των αξιολογήσεων στοιχείων για έναν συγκεκριμένο χρήστη και το πρόβλημα κορυφαίων- N προτάσεων, το οποίο σχετίζεται με την κατάταξη προτάσεων. Το πρώτο πρόβλημα συνίσταται στην εύρεση, για έναν συγκεκριμένο χρήστη, του νέου αντικειμένου $i \in I$ το οποίο είναι πιο πιθανό να τον ενδιαφέρει. Όταν οι αξιολογήσεις είναι διαθέσιμες, αυτή η εργασία ορίζεται συχνότερα ως πρόβλημα ταξινόμησης παλινδρόμησης (ή πολλαπλών τάξεων), όπου ο στόχος είναι μια συνάρτηση που προβλέπει την αξιολόγηση του χρήστη u για ένα νέο στοιχείο i . Αυτή η συνάρτηση χρησιμοποιείται στη συνέχεια για να προτείνει στον ενεργό χρήστη ένα στοιχείο i για το οποίο η εκτιμώμενη βαθμολογία έχει την υψηλότερη τιμή.

Τα συστήματα προτάσεων μπορούν να αποθηκεύσουν και να ενημερώσουν το ιστορικό κάθε χρήστη βάσει της προηγούμενης αλληλεπίδρασης και συμπεριφοράς εντός του συστήματος. Στη συνέχεια, προτείνονται άγνωστα στοιχεία που ταιριάζουν με τον χρήστη-στόχο. Ωστόσο, για να δη-

μιουργήσουν εξατομικευμένες προτάσεις, οι χρήστες πρέπει να αξιολογήσουν ορισμένα στοιχεία στην αρχή. Όπως, για παράδειγμα, στο σύστημα σύστασης MovieLens, ένας νέος χρήστης πρέπει να αξιολογήσει ορισμένα στοιχεία για να λάβει σχετικές προτάσεις από το σύστημα.

2.2 Είδη και τεχνικές

Κάθε σύστημα συνήθως εξυπηρετεί διαφορετικούς σκοπούς και απαιτήσεις, συνεπώς παρατηρούνται αρκετές διαφοροποιήσεις μεταξύ των συστημάτων αυτών. Για παράδειγμα δεν ισχύει πάντα ότι η έκβαση ότι θέλουμε να προτείνουμε αντικείμενα στους χρήστες. Πιο συγκεκριμένα σε πολλές περιπτώσεις το ρόλο του προϊόντος τον έχει ο χρήστης. Ένα τέτοιο παράδειγμα μπορεί να θεωρηθούν οι χορηγούμενες διαφημίσεις. Για κάθε αντικείμενο που πρέπει να προταθεί, ο αλγόριθμος αναζητεί παρόμοιους χρήστες που θα προταθεί το συγκεκριμένο προϊόν. Είναι αναμενόμενο ότι αυτό το σύστημα θα λειτουργεί διαφορετικά αφού οι ρόλοι έχουν αντιστραφεί και το σύστημα θέλουμε να συμπεριφερθεί με διαφορετικό τρόπο. Υπάρχουν διαφορετικά είδη και τεχνικές σε ότι αφορά τα συστήματα συστάσεων. Οι πιο γνωστές τεχνικές που διέπουν τα συστήματα συστάσεων συνοψίζονται στις ακόλουθες υποενότητες [36, 37]

2.2.1 Συνεργατική διήθηση

Η τεχνική της συνεργατικής διήθησης ήταν η πρώτη που χρησιμοποιήθηκε στα συστήματα σύστασης και θεωρείται η πιο δημοφιλής και ευρέως εφαρμοσμένη [4]. Παραδείγματα γνωστών ιστοσελίδων που κάνουν χρήση αυτής της τεχνικής είναι οι Amazon, TiVo και Netflix [38, 26].

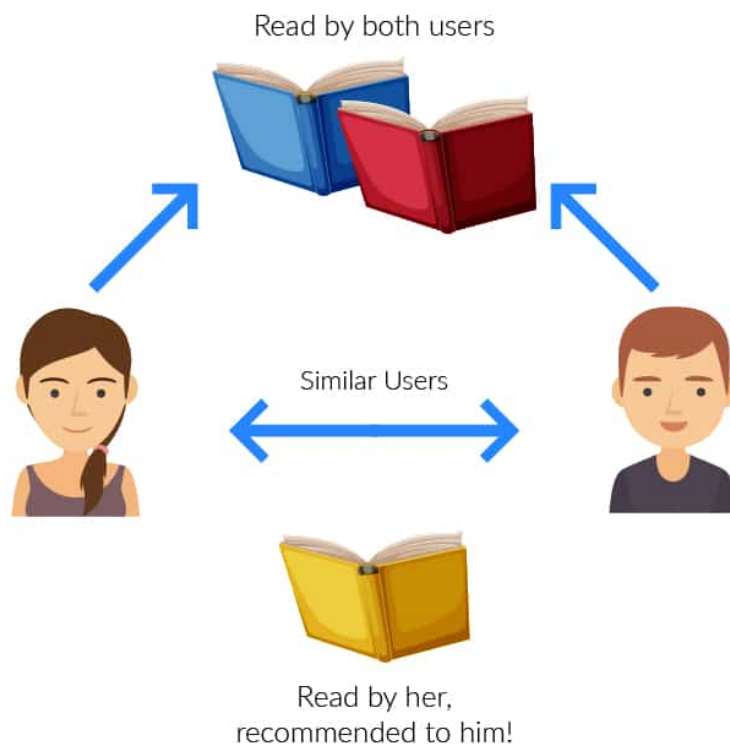
Η CF παρέχει προτάσεις βάσει των απόψεων άλλων ατόμων που έχουν τα ίδια ενδιαφέροντα με τον χρήστη που επιθυμεί να λάβει συστάσεις [39]. Αυτά τα ενδιαφέροντα αντιπροσωπεύονται συχνά ως πίνακες βαθμολογίας R [40]. Αυτοί οι πίνακες έχουν διαστάσεις $m \times n$ περιέχοντας m χρήστες και n αντικείμενα. Ως εκ τούτου, η βαθμολογία του χρήστη i για το στοιχείο j δίνεται από το r_{ij} . Επειδή σε οποιοδήποτε σύστημα σύστασης, ο αριθμός των αξιολογήσεων που λαμβάνονται είναι συνήθως πολύ μικρός σε σύγκριση με τον αριθμό των συστάσεων που πρέπει να γίνουν, ο πίνακας R είναι συχνά αρκετά αραιός [2]. Αυτή την ιδιότητα συνήθως την συναντάμε σε υπηρεσίες προβολής βίντεο. Οι περισσότεροι χρήστες έχουν παρακολουθήσει μόνο ένα μικρό υποσύνολο των διαθέσιμων βίντεο. Επομένως, ακόμα και αν οι χρήστες βαθμολογήσουν όλα τα βίντεο που παρακολούθησαν, πολλά στοιχεία r_{ij} στον πίνακα R θα παραμείνουν κενά.

Η προσέγγιση που υιοθετήθηκε στις μεθόδους CF είναι ότι αυτές οι βαθμολογίες που λείπουν μπορούν να προβλεφθούν, επειδή οι παρατηρούμενες αξιολογήσεις συσχετίζονται συχνά σε διάφορους χρήστες και αντικείμενα [41]. Για παράδειγμα, εάν δύο χρήστες μοιράζονται παρόμοιες αξιολογήσεις μεταξύ αντικειμένων, ο αλγόριθμος CF θα αναγνωρίσει αυτοί οι χρήστες έχουν παρόμοια ενδιαφέροντα. Επομένως, αυτή η ομοιότητα μπορεί να χρησιμοποιηθεί με σκοπό να δημιουργηθούν προβλέψεις σχετικά με αυτές τις αξιολογήσεις που είναι κενές. Τα περισσότερα μοντέλα CF επικεντρώνονται στην αξιοποίηση συσχετισμών μεταξύ αντικειμένων ή συσχετίσεις μεταξύ χρηστών [41]. Αυτός είναι ο λόγος για τον οποίο το CF συχνά αναφέρεται ως «συσχέτιση μεταξύ ανθρώπων» [42]. Το Σχήμα 2.1 απεικονίζει αυτό το συμπέρασμα αξιολόγησης με βάση τις ομοιότητες μεταξύ των χρηστών. Σε αυτήν την εικόνα, είναι εμφανές ότι επειδή και οι 2 χρήστες βρήκαν ενδιαφέρον το ίδιο βιβλίο, είναι πολύ πιθανόν οι απόψεις τους να ταιριάζουν και σε άλλα βιβλία.

Όπως αναφέρθηκε παραπάνω αυτή η μέθοδος έχει προβλήματα όπως τους αραιούς πίνακες και προβλήματα κάλυψης, ενώ ακόμα και σήμερα γίνονται σχετικές έρευνες. Οι τεχνικές μείωσης δεδομένων φαίνεται να αποτελούν πολλά υποσχόμενες ερευνητικές κατευθύνσεις για την επίλυση αυτού του προβλήματος [2]. Η χρήση του CF έχει επίσης το πλεονέκτημα της χρήσης μιας προσέγγισης που του επιτρέπει να αποκτηθούν ουσιαστικές γνώσεις μεταξύ των σχέσεων χρηστών και αντικειμένων που δεν είναι άμεσα συνδεδεμένα [43].

Σύμφωνα με τους Adamavicius και Tuzhilin [2], οι αλγόριθμοι CF μπορούν να ομαδοποιηθούν σε δύο κατηγορίες: βάσει μνήμης και μοντέλου.

COLLABORATIVE FILTERING



Σχήμα 2.1: Συνεργατική διήθηση [3]

1. Βάσει μνήμης: Οι αλγόριθμοι με βάση τη μνήμη, που αναφέρονται επίσης ως αλγόριθμοι με βάση τη γειτονιά, ήταν μεταξύ των πρώτων αλγορίθμων [40]. Αυτοί οι αλγόριθμοι προβλέπουν τις αξιολογήσεις των χρηστών για ένα αντικείμενο με βάση την γειτονιά τους [41]. Αυτές οι γειτονιές μπορούν να οριστούν με έναν από τους δύο τρόπους:
 - Βασισμένο στο χρήστη CF: Αυτοί οι αλγόριθμοι παρέχουν προτάσεις για αντικείμενα που άρεσαν σε παρόμοιους χρήστες [2]. Επομένως, στόχος είναι να προταθούν αξιολογήσεις για αντικείμενα που δεν έχουν αξιολογηθεί ακόμη από το χρήστη. Αυτό μπορεί να επιτευχθεί με τον υπολογισμό των σταθμισμένων μέσων όρων των αξιολογήσεων που παρέχονται από χρήστες που έχουν τα ίδια ενδιαφέροντα με τον χρήστη u [41]. Για παράδειγμα, όταν η Αλίκη και ο Μπομπ έχουν παρόμοιες βαθμολογίες μεταξύ ταινιών, ένας αλγόριθμος βασισμένος στο χρήστη CF μπορεί να χρησιμοποιήσει την βαθμολογία που έδωσε η Alice στην ταινία Rambo για να προβλέψει την βαθμολογία που θα έδινε ο Μπομπ αν έπρεπε να την αξιολογήσει ο ίδιος.
 - Βασισμένος στο αντικείμενο CF: Αυτοί οι αλγόριθμοι παρέχουν προτάσεις για αντικείμενα παρόμοια με αυτά που άρεσαν στο χρήστη στο παρελθόν [2]. Κατά συνέπεια, για να προβλεφθεί η βαθμολογία για ένα στοιχείο i που δίνεται από έναν χρήστη u , το πρώτο βήμα περιλαμβάνει τον καθορισμό του συνόλου S των στοιχείων που είναι περισσότερο παρόμοια με το στοιχείο i . Οι βαθμολογίες στο σύνολο S χρησιμοποιούνται στη συνέχεια για να προβλέψουν εάν ο χρήστης θα θέλει το στοιχείο i . Ως εκ τούτου, αν ο Μπομπ έδωσε θετική βαθμολογία σε κλασικές ταινίες όπως το *Gone with the Wind*, αυτές οι βαθμολογίες μπορούν να χρησιμοποιηθούν για να προβλέψουν τις αξιολογήσεις του για άλλες κλασικές ταινίες, όπως το *Citizen Kane*. Η απόφαση σχετικά με το ποια προσέγγιση ταιριάζει καλύτερα, βασίζεται συνήθως στην

αναλογία του αριθμού των χρηστών σε σχέση με τον αριθμό των αντικειμένων. Στις περιπτώσεις όπου ο αριθμός των χρηστών είναι μεγαλύτερος από το αριθμό αντικειμένων, προσεγγίσεις βασισμένες στο αντικείμενο είναι καταλληλότερες επειδή παρέχουν περισσότερο ακριβείς συστάσεις ενώ είναι υπολογιστικά γρηγορότεροι [43]. Από την άλλη βασισμένες στο χρήστη προσεγγίσεις συνήθως παρέχουν πιο πρωτότυπες συστάσεις [44]. Αυτές οι μέθοδοι που βασίζονται στη γειτονιά μπορούν να θεωρηθούν ως γενικεύσεις των ταξινομητών k-πλησιέστερων γειτόνων. Επομένως, αυτές οι μέθοδοι θεωρούνται μαθησιακές μέθοδοι που βασίζονται σε στιγμιότυπα, οι οποίες είναι συγκεκριμένες για την υπόθεση που προβλέπεται [45]. Μεταξύ των πλεονεκτημάτων αυτών των αλγορίθμων είναι η απλότητα εφαρμογής τους και η ευκολία εξήγησης των συστάσεων τους. Παρόλα αυτά, αυτοί οι αλγόριθμοι δεν λειτουργούν πολύ καλά με αραιούς πίνακες αξιολόγησης [41].

2. Βάσει μοντέλου: Οι αλγόριθμοι που βασίζονται στο μοντέλο χρησιμοποιούν μεθόδους μηχανικής μάθησης και εξόρυξης δεδομένων για τη δημιουργία μοντέλων πρόβλεψης που θα χρησιμοποιηθούν για την εύρεση πιθανών τιμών που λείπουν στον πίνακα βαθμολογίας R [45]. Όπως συμβαίνει και με άλλες επιβλεπόμενες ή μη επιβλεπόμενες τεχνικές μηχανικής μάθησης, τα μοντέλα πρόβλεψης δημιουργούνται (εκπαιδεύονται) πριν από τη φάση πρόβλεψης. Παραδείγματα των παραδοσιακών μεθόδων μηχανικής μάθησης που μπορούν να γενικευτούν στο CF αυτής της περίπτωσης είναι: τα δέντρα αποφάσεων (decision trees), μεθόδους βάσει κανόνων (rule-based methods), ταξινομητές Bayes (Bayes classifiers), μοντέλα παλινδρόμησης (regression models), μηχανές διανυσμάτων υποστήριξης (support vector machines - SVM) και νευρωνικά δίκτυα (and neural networks) [46].

Σύμφωνα με τον Aggarwal [45], τα CF συστήματα συστάσεων βάσει μοντέλων Model based collaborative filtering recommendation systems (MB-CFRS) έχουν συχνά πολλά οφέλη έναντι των μεθόδων που βασίζονται στη μνήμη:

- Απόδοση χώρου: συνήθως, το μοντέλο που δημιουργείται είναι πολύ μικρότερο από τον αρχικό πίνακα βαθμολογίας
- Ταχύτητα εκπαίδευσης και πρόβλεψης: ενώ οι μέθοδοι που βασίζονται στη μνήμη έχουν τετραγωνικό στάδιο προεπεξεργασίας, τα συστήματα που βασίζονται σε μοντέλα είναι συνήθως πολύ ταχύτερα στη φάση προεπεξεργασίας, ενώ στις περισσότερες περιπτώσεις μπορούν επίσης να κάνουν προβλέψεις με ευκολία.
- Αποφυγή υπερμοντελοποίησης (overfitting): αν και η υπερμοντελοποίηση είναι ένα σοβαρό πρόβλημα σε αλγορίθμους μηχανικής μάθησης, η προσέγγιση της συνοψίωσης που χρησιμοποιείται στα συγκεκριμένα μοντέλα μπορεί συχνά βοηθήσει στην αποφυγή του παραπάνω φαινομένου.

Παρόλο που τα συστήματα που βασίζονται στη μνήμη προσφέρουν το πλεονέκτημα της απλής εφαρμογής [41], συχνά υστερούν σε ακρίβεια. Σε γενικές γραμμές, οι τεχνικές που βασίζονται σε μοντέλα είναι οι πιο ακριβείς, ειδικά όταν χρησιμοποιούνται μοντέλα λανθάνουσών παραγόντων (latent factor models) [45].

Ανεξάρτητα από τη μέθοδο που χρησιμοποιείται, η CF παρέχει συστάσεις με βάση τις απόψεις άλλων που έχουν τα ίδια ενδιαφέροντα με τον χρήστη [39]. Αυτές οι απόψεις συλλαμβάνονται με τη μορφή αξιολογήσεων και συχνά καθορίζονται σε κλίμακα που δείχνει πόσο ικανοποιημένος είναι ένας χρήστης με ένα στοιχείο. Αν και υπάρχει πιθανότητα η κλίμακα βαθμολογίας να αντιπροσωπεύεται ως σύνολο συνεχών τιμών, αυτό είναι σχετικά σπάνιο. Συνήθως η κλίμακα βαθμολογίας αντιπροσωπεύεται ως διακριτό διάστημα που ορίζει πόσο ικανοποιημένος είναι ο χρήστης είναι με ένα συγκεκριμένο αντικείμενο [41]. Επιπλέον, αντί της χρήσης αριθμητικών τιμών, η κλίμακα βαθμολογίας μπορεί να αναπαρασταθεί ως εικονίδια. Ένα κοινό παράδειγμα μιας τέτοιας αναπαράστασης απεικονίζεται στο Σχήμα 2.2, όπου απεικονίζεται μια κλίμακα πέντε σημείων που αντιπροσωπεύει το σύνολο -2, -1, 0, 1, 2 ως αστέρι.

Star Representation	Numerical Representation	Textual Representation
★★★★★	2	Excellent
★★★★☆	1	Good
★★★☆☆	0	Fair
★★☆☆☆	-1	Poor
★☆☆☆☆	-2	Terrible

Σχήμα 2.2: Παράδειγμα βαθμολόγησης με εικονίδια

Επιπλέον, η ερμηνεία της κλίμακας βαθμολογίας μπορεί να διαφέρει από τη μια εταιρία στην άλλη. Παρόλο που τόσο η Amazon όσο και το Netflix χρησιμοποιούν την ίδια κλίμακα βαθμολογίας πέντε σημείων, η ερμηνεία διαφέρει. Ενώ η Amazon χρησιμοποιεί γραμμική κλίμακα, το Netflix επέλεξε να χρησιμοποιήσει ένα σύστημα αξιολόγησης πέντε αστερών στο οποίο το επίπεδο σημείου τεσσάρων αστερών αντιστοιχεί στο «πραγματικά του άρεσε» και το κεντρικό επίπεδο των τριών αστερών αντιστοιχεί στο «άρεσε». Επομένως, υπάρχουν τρεις ευνοϊκές βαθμολογίες και δύο δυσμενείς βαθμολογίες [41].

Μια ειδική περίπτωση κλίμακας βαθμολογίας είναι η unary κλίμακα[41]]. Σε αυτή τη περίπτωση, οι χρήστες έχουν την δυνατότητα να δώσουν τη γνώμη τους "like" για ένα αντικείμενο, αλλά δεν υπάρχει εναλλακτική λύση για να προσδιορίσουν μια αντιπαράθεση. Μια unary κλίμακα βαθμολογίας σχετίζεται συνήθως με συστήματα που χρησιμοποιούν έμμεσα σχόλια [47], στα οποία οι προτιμήσεις των χρηστών λαμβάνονται βάσει των δραστηριοτήτων τους αντί να δίδονται ρητά από τους χρήστες. Ο αντίκτυπος της χρήσης μιας τέτοιας κλίμακας πρέπει να ληφθεί υπόψη από τον αλγόριθμο σύστασης, επειδή δεν υπάρχουν πληροφορίες σχετικά με τις αντιπάθειες των χρηστών.

2.2.2 Φιλτράρισμα βάσει περιεχομένου

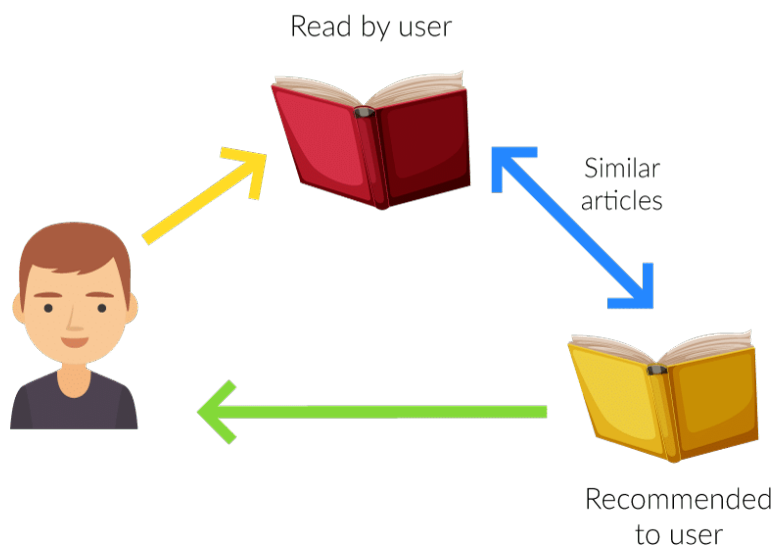
Στο φιλτράρισμα βάσει περιεχομένου (CBF) τόσο το στοιχείο όσο και το προφίλ χρήστη περιγράφονται από ένα σύνολο χαρακτηριστικών και τα προτεινόμενα στοιχεία ταξινομούνται με βάση το πόσο παρόμοια είναι με τα προφίλ χαρακτηριστικών χρήστη [4]. Αυτός ο τύπος RS είναι πολύ δημοφιλής λόγω της απλότητας του και μπορεί να χρησιμοποιηθεί για να προτείνει αντικείμενα όπως ιστότοποι, μουσική, ταινίες, βιβλία, εστιατόρια και ξενοδοχεία.

Τα βήματα που πρέπει να λάβει κάποιος υπόψη κατά τη δημιουργία ενός συστήματος βάσει περιεχομένου είναι τα ακόλουθα (υποθέτοντας ότι ο σχεδιαστής έχει πρόσβαση στα απαιτούμενα δεδομένα):

1. Δημιουργία του προφίλ χαρακτηριστικών που περιγράφουν τα αντικείμενα που θα προταθούν.
2. Δημιουργία προφίλ χρήστη χρησιμοποιώντας το ίδιο σύνολο χαρακτηριστικών.
3. Επινόηση ενός μέτρου για τον υπολογισμό της ομοιότητας μεταξύ χρήστη και αντικειμένου.
4. Κατάταξη των στοιχείων σε σχέση με την ομοιότητα και σύσταση αυτών στον χρήστη.
5. Βελτίωση του προφίλ χρήστη με βάση τα σχόλια.

Λόγω του ότι αυτή η μέθοδος βασίζεται αποκλειστικά στο περιεχόμενο, έχει αρκετές επιθυμητές ιδιότητες, όπως καλή επεκτασιμότητα, δηλαδή το σύστημα λειτουργεί καλά ανεξάρτητα από τον αριθμό των χρηστών που υπάρχουν στο σύστημα. Επιπλέον, λειτουργεί καλά για την πρόβλεψη στοιχείων που είναι νέα στο σύστημα, καθώς το μόνο που χρειάζεται είναι ο υπολογισμός της ομοιότητας μεταξύ του στοιχείου και του προφίλ χρήστη. Αυτό προϋποθέτει ότι το προφίλ στοιχείων έχει ήδη

CONTENT-BASED FILTERING



Σχήμα 2.3: Φιλτράρισμα βάσει περιεχομένου [3]

δημιουργηθεί. Η CBF είναι κατανοητή και ο χρήστης μπορεί εύκολα να κατανοήσει γιατί συνιστάται ένα συγκεκριμένο αντικείμενο, κάνοντας το σύστημα πιο αξιόπιστο για τον χρήστη.

Το Σχήμα 2.3 απεικονίζει τον τρόπο λειτουργίας της συγκεκριμένης τεχνικής. Οι προβλέψεις βασίζονται στα χαρακτηριστικά του αντικειμένου σε συνδυασμό με τα χαρακτηριστικά που αρέσουν στον χρήστη.

Από την άλλη πλευρά, υπάρχουν ορισμένα μειονεκτήματα. Ως συνέπεια των συστάσεων που βασίζονται αποκλειστικά στο προφίλ του χρήστη, ενδέχεται να γίνουν υπερβολικά εξειδικευμένες. Για παράδειγμα, εάν σε έναν χρήστη άρεσαν πολλές ταινίες της ίδιας σειράς ταινιών, το σύστημα θα αμβλύνει για να προτείνει τις υπόλοιπες ταινίες αυτής της σειράς και τίποτα άλλο. Αυτό είναι ένα παράδειγμα όπου υπάρχει πολύ μικρή διακύμανση στην έξοδο του συστήματος και οι συστάσεις γίνονται βαρετές. Ένα άλλο μειονέκτημα είναι η δυσκολία δημιουργίας του προφίλ των στοιχείων. Για να εξαχθούν χαρακτηριστικά ενός αντικειμένου και ληφθούν υπόψη όλες οι διαφορετικές πτυχές μπορεί να είναι μια απαιτητική εργασία. Για παράδειγμα, όταν δημιουργείται ένα προφίλ αντικειμένου για έναν ιστότοπο, είναι δύσκολο να αναπαρασταθεί η εμπειρία του χρήστη ή οι αισθητικές λεπτομέρειες, τα οποία είναι βασικά χαρακτηριστικά που πρέπει να ληφθούν υπόψη όταν αξιολογούνται ιστοσελίδες.

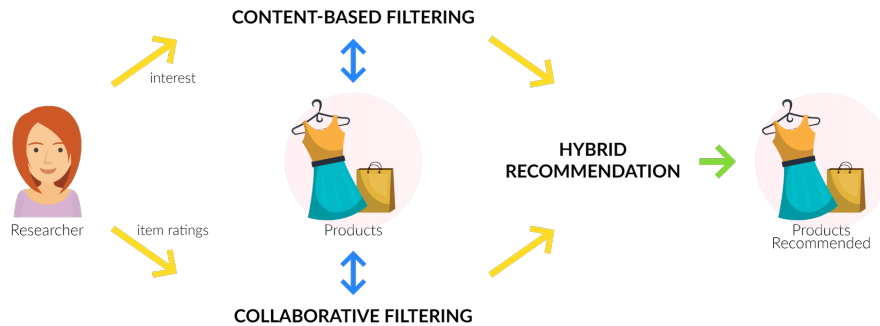
2.2.3 Υβριδικά συστήματα συστάσεων

Ένα υβριδικό σύστημα σύστασης χρησιμοποιείται για να αποφευχθούν ορισμένοι περιορισμοί των μεθόδων φιλτραρίσματος βάσει περιεχομένου ή συνεργατικής. Οι συνεργατικές μέθοδοι έχουν μια σχετικά μεγάλη αδυναμία γύρω από το πρόβλημα της ψυχρής εκκίνησης (cold start), καθώς ο χρήστης πρέπει να έχει αξιολογήσει ορισμένα στοιχεία για να λάβει προτάσεις. Στην καθημερινότητα τα προβλήματα συστάσεων είναι πολύ πιο περίπλοκα από ένα σύστημα σύστασης ταινιών [2]. Μια πραγματική επιχειρηματική περίπτωση θα έχει σημαντικές ποσότητες παραμέτρων, όπου ένα υβριδικό μοντέλο θα βοηθήσει στην εκμάθηση ενός πιο περίπλοκου μοντέλου του χρήστη, δίνοντας έτσι ακριβέστερες προτάσεις. Τα υβριδικά μοντέλα είναι ένα μείγμα των μεθόδων που περιγράφηκαν προηγουμένως. Δεδομένου ότι μπορούν να συνδυαστούν διαφορετικές μέθοδοι, μπορούν να δημιουργηθούν πολλαπλές υβριδικές λύσεις. Ο Burke [33] καλύπτει επτά διαφορετικές προσεγγίσεις για το συνδυασμό συστάσεων:

- Τα σταθμισμένα υβριδικά συστήματα σύστασης περνούν από μια φάση εκπαίδευσης όπου τα βάρη στο σύστημα συστάσεων μαθαίνονται από ένα σύνολο δεδομένων εκπαίδευσης κατά τη διάρκεια μιας περιόδου εκπαίδευσης. Το αποτέλεσμα αυτής της φάσης είναι τα βάρη κάθε στοιχείου. Οι υποψηφιοί από κάθε στοιχείο σταθμίζονται σε σχέση με αυτές τις αριθμητικές τιμές κάθε φορά που γίνεται μια πρόταση. Το αποτέλεσμα είναι είτε ένωση είτε διασταύρωση αυτών των λιστών. Αυτή η προσέγγιση δεν κάνει διάκριση μεταξύ χρηστών, υπάρχει η υπόθεση ότι κάθε στοιχείο θα έχει σταθερή απόδοση για όλους τους χρήστες
- Τα μεικτά υβριδικά μοντέλα παρουσιάζουν μια σύνθετη λίστα αντικειμένων δίπλα-δίπλα από κάθε ένα από τα στοιχεία. Δεν υπάρχει πραγματικός συνδυασμός των διαφορετικών αλγορίθμων σύστασης. Η πρόκληση στο μεικτό υβριδικό είναι η σειρά και η εμπιστοσύνη σε καθένα από τα στοιχεία που συντίθενται μαζί. Μια τεχνική συχνά περιλαμβάνει τη συγχώνευση των λιστών σε μια κανονικοποίηση της προβλεπόμενης βαθμολογίας.
- Ένα εναλλασσόμενο υβριδικό σύστημα συστάσεων έχει εμπιστοσύνη σε κάθε στοιχείο και επιλέγει τη σύσταση που δίνει τα καλύτερα αποτελέσματα τη δεδομένη στιγμή. Αυτή η μέθοδος, παρόμοια με τη μεικτή προσέγγιση, δεν συνδυάζει τους αλγόριθμους σε έναν νέο. Ένα σημαντικό μέρος του μοντέλου είναι τα κριτήρια επιλογής. Τόσο η στατιστική εμπιστοσύνη όσο και τα εξωτερικά κριτήρια μπορούν να χρησιμοποιηθούν για την επιλογή του καλύτερου συστήματος σε κάθε περίπτωση χρήστη.
- Τα διαδοχικά υβριδικά μοντέλα προσπαθούν να δημιουργήσουν μια αυστηρή ιεραρχία στοιχείων συστάσεων. Τα διάφορα στοιχεία κατατάσσονται και τα δευτερεύοντα συμπληρώνουν τον κύριο συνιστώμενο όταν υπάρχει ισοπαλία.
- Τα υβριδικά συστήματα μετα-επιπέδων περνούν πρώτα από μια φάση εκπαίδευσης Αυτή η φάση εκπαίδευσης είναι να εκπαιδεύσει ένα σύστημα συστάσεων να έχει μια ιδέα ενός μαθημένου μοντέλου. Το πραγματικό σύστημα στη συνέχεια χρησιμοποιεί αυτό το μοντέλο για να παράγει προτάσεις για τον τελικό χρήστη. Δεν είναι όλα τα ζεύγη διαφορετικών συστατικών σύστασης κατάλληλα για αυτό το υβριδικό.
- Ο συνδυασμός χαρακτηριστικών είναι ένα υβριδικό μοντέλο όπου ένα από τα στοιχεία χρησιμοποιεί την έξοδο του ως είσοδο στο πραγματικό σύστημα σύστασης. Για παράδειγμα, η έξοδος από έναν συνεργατικό αλγόριθμο φιλτραρίσματος μπορεί να χρησιμοποιηθεί ως είσοδος σε έναν αλγόριθμο βάσει περιεχομένου. Αυτή η μέθοδος επιτυγχάνει ότι το πρώτο στοιχείο είναι ένα σύστημα που δίνει μια τιμή κατεύθυνσης στον αλγόριθμο που δίνει τη τελική λίστα, κάθε φορά.
- Η αύξηση δυνατοτήτων είναι μια στρατηγική όπου ένα στοιχείο χρησιμοποιείται για την παραγωγή ενός νέου επαυξημένου χαρακτηριστικού στο επόμενο στοιχείο. Αυτή η προσέγγιση χρειάζεται επίσης μια φάση εκπαίδευσης που μαθαίνει το επαυξημένο προφίλ. Η αύξηση δυνατοτήτων μπορεί να αντλήσει νέες δυνατότητες περιεχομένου σε αυτό το στάδιο που μπορεί να είναι πολύτιμες κατά τη δημιουργία υποψηφίων στοιχείων.

Τα περισσότερα συστήματα συστάσεων χρησιμοποιούν πλέον μια υβριδική προσέγγιση, συνδυάζοντας CF, CBF και άλλες προσεγγίσεις. Δεν υπάρχει λόγος για τον οποίο δεν μπορούσαν να συνδυαστούν πολλές διαφορετικές τεχνικές του ίδιου τύπου. Οι υβριδικές προσεγγίσεις μπορούν να εφαρμοστούν με διάφορους τρόπους όπως περιγράφηκαν παραπάνω. Αρκετές μελέτες που συγκρίνουν εμπειρικά την απόδοση του υβριδικού μοντέλου με τις άλλες δύο τεχνικές, έδειξαν ότι οι υβριδικές μέθοδοι μπορούν να παρέχουν πιο ακριβείς συστάσεις. Το netflix είναι ένα παράδειγμα χρήσης υβριδικού μοντέλου. Πιο συγκεκριμένα συνδυάζει τις αναζητήσεις του χρήστη (CBF) ενώ ταυτόχρονα χρησιμοποιεί την ομοιότητα των χρηστών (CF).

Το Σχήμα 2.4 απεικονίζει τον τρόπο λειτουργίας της συγκεκριμένης τεχνικής. Οι προτάσεις όπως βασίζονται στο συνδυασμό των δυο προηγούμενων τεχνικών που είχαν αναλυθεί παραπάνω.



Σχήμα 2.4: Υβριδικό σύστημα συστάσεων [3]

2.2.4 Συστήματα συστάσεων επίγνωσης πλαισίου

Τα περισσότερα από τα συστήματα σύστασης δεν λαμβάνουν υπόψη το πλαίσιο επίγνωσης του χρήστη κατά την εύρεση των στοιχείων που προτείνουν [4]. Ωστόσο, μπορεί συχνά να είναι χρήσιμο να ληφθεί υπόψη το πλαίσιο στο οποίο βρίσκεται ο χρήστης όταν δίνονται προτάσεις. Για ορισμένους τομείς, διάφορες πληροφορίες όπως η ημέρα της εβδομάδας και το που βρίσκεται ο χρήστης μπορούν να έχουν σημαντικό αντίκτυπο σε αυτό που θα ήθελε ο χρήστης εκείνη τη στιγμή.

Τι είναι το πλαίσιο επίγνωσης

Πριν εξηγηθούν περαιτέρω τα *συστήματα σύστασης επίγνωσης πλαισίου* Context aware recommendation systems Συστήματα συστάσεων επίγνωσης πλαισίου (CARS), είναι σημαντικό να προσδιορισθεί τι είναι ένα πλαίσιο επίγνωσης. Όπως γράφουν οι G. Adomavicius και A. Tuzhilin [4], ο όρος πλαίσιο χρησιμοποιείται σε τόσους πολλούς διαφορετικούς κλάδους και ο ορισμός ποικίλλει συχνά αυτών (επιστήμη υπολογιστών, γλωσσολογία, φιλοσοφία, ψυχολογία και οργανωτικές επιστήμες). Επικεντρώνονται στους ορισμούς στα πεδία που σχετίζονται περισσότερο με τα συστήματα σύστασης και θα συνοψιστούν εδώ.

- Στην εξόρυξη δεδομένων, το πλαίσιο μερικές φορές ορίζεται ως εκείνα τα γεγονότα που χαρακτηρίζουν τα στάδια ζωής ενός πελάτη και μπορεί να καθορίσει μια αλλαγή στις προτιμήσεις, την κατάσταση και την αξία του για μια εταιρεία [48]. Αυτές οι πληροφορίες μπορεί να είναι γεγονότα όπως μια νέα εργασία, η γέννηση ενός παιδιού, ο γάμος, το διαζύγιο ή η συνταξιοδότηση.
- Τα συστήματα εξατομίκευσης του ηλεκτρονικού εμπορίου μερικές φορές εξετάζουν την πρόθεση ενός πελάτη για την πραγματοποίηση μιας αγοράς ως πληροφορία επίγνωσης πλαισίου [49]. Για παράδειγμα, εάν κάποιος αγοράσει κάτι ως δώρο, δεν χρειάζεται απαραίτητα προτάσεις βάσει αυτού όταν επισκέπτεται ξανά το ίδιο κατάστημα αγοράζοντας κάτι για τον εαυτό του.
- Σε πανταχού παρόντα συστήματα που κάνουν χρήση επίγνωσης πλαισίου, το πλαίσιο ορίστηκε αρχικά ως η τοποθεσία του χρήστη, η ταυτότητα των ατόμων κοντά στον χρήστη, τα αντικείμενα γύρω, και οι αλλαγές σε αυτά τα στοιχεία [50]. Ημερομηνία, σεζόν, θερμοκρασία, φυσικές και εννοιολογικές καταστάσεις ενδιαφέροντος για έναν χρήστη, συναισθηματικές καταστάσεις προστέθηκαν αργότερα [51, 52, 53].
- Στην έρευνα μάρκετινγκ και διαχείρισης έχει μελετηθεί η διαφορετική συμπεριφορά και λήψη αποφάσεων, που ο πελάτης μπορεί να υιοθετήσει και να προτιμήσει διαφορετικά προϊόντα με βάση το πλαίσιο [54, 55]. Ο C. K. Prahalad [56] ορίζει το πλαίσιο ως «την ακριβή φυσική τοποθεσία ενός πελάτη ανά πάσα στιγμή, το ακριβές λεπτό που χρειάζεται η υπηρεσία και το είδος της τεχνολογικής κινητής συσκευής πάνω στην οποία θα ληφθεί αυτή η εμπειρία».

Διακρίνει μεταξύ τριών διαστάσεων: χρονική (πότε να προσφέρει εμπειρίες πελατών), χωρική (πού να παραδίδεται) και τεχνολογική (πώς να παραδίδεται).

Μοντέλο

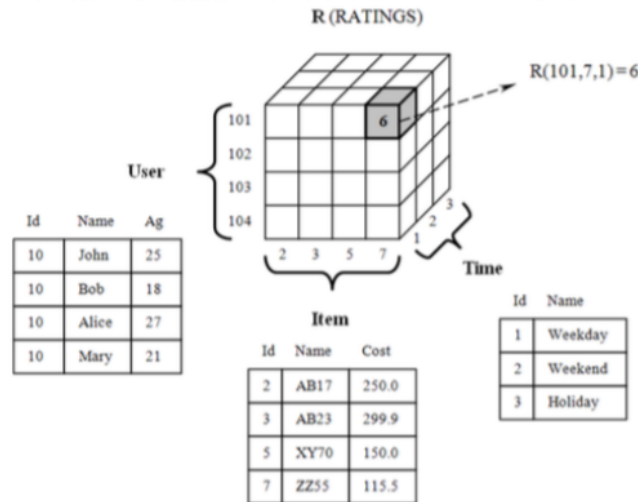
Ένα τυπικό σύστημα σύστασης αποτελείται από χρήστες και αντικείμενα και προσπαθεί να μάθει μια συνάρτηση για την εκτίμηση της βαθμολογίας για έναν χρήστη σε ένα δεδομένο αντικείμενο.

$$Rating = User \times Item \quad (2.1)$$

Όταν λαμβάνεται υπόψη και το πλαίσιο του χρήστη, τότε η συνάρτηση καταλήγει με την ακόλουθη μορφή:

$$Rating = User \times Item \times Context \quad (2.2)$$

Εάν χρησιμοποιηθεί το παράδειγμα των G. Adomavicius και A. Tuzhilin [4] όπου θεωρούν τον χρόνο



Σχήμα 2.5: Πολυδιάστατο μοντέλο για $User \times Item \times Time$ recommendation space [4]

ως πλαίσιο, τότε αυτό μπορεί να μοντελοποιηθεί σε τρεις διαστάσεις όπως φαίνεται στο Σχήμα 2.5. Οι G. Adomavicius και A. Tuzhilin [4] περιγράφουν δύο διαφορετικές προσεγγίσεις για τη χρήση πληροφοριών επίγνωσης πλαισίου σε μια διαδικασία σύστασης. (1) σύσταση μέσω ερωτημάτων και αναζήτησης πλαισίου και (2) σύσταση μέσω διερεύνησης και εκτίμησης προτιμήσεων με βάση το πλαίσιο επίγνωσης. Τα κινητά και τουριστικά συστήματα [57, 58, 59] χρησιμοποιούν την πρώτη προσέγγιση, χρησιμοποιώντας τις πληροφορίες επίγνωσης πλαισίου κατά την αναζήτηση για ορισμένες πληροφορίες, όπως την αναζήτηση για εστιατόρια κοντά στο χρήστη.

Η επόμενη προσέγγιση, η σύσταση μέσω διερεύνησης και εκτίμησης προτιμήσεων με χρήση επίγνωσης πλαισίου, αντιπροσωπεύει μια πιο πρόσφατη τάση για συστήματα συστάσεων με γνώμονα το περιβάλλον [60, 61, 62, 63]. Αυτή η προσέγγιση προσπαθεί να μοντελοποιήσει και να μάθει τις προτιμήσεις του χρήστη αποκτώντας πληροφορίες ενώ ο χρήστης χρησιμοποιεί την εφαρμογή και παρακολουθώντας τον τρόπο με τον οποίο αλληλεπιδρά με προηγούμενες συστάσεις. Οι G. Adomavicius και ο A. Tuzhilin [4] περιγράφουν τρεις τρόπους με τους οποίους μπορεί να εφαρμοστεί αυτή η διαδικασία σύστασης:

- **Contextual pre-filtering:** Οι πληροφορίες επίγνωσης πλαισίου χρησιμοποιούνται ως μηχανισμός φιλτραρίσματος που εφαρμόζεται στα δεδομένα, πριν από την εφαρμογή του αλγορίθμου σύστασης.

- **Contextual post-filtering:** Οι πληροφορίες περιβάλλοντος αγνοούνται αρχικά και οι προτάσεις υπολογίζονται με την εφαρμογή παραδοσιακών αλγορίθμων σύστασης σε ολόκληρα τα δεδομένα. Το προκύπτον σύνολο προτάσεων στη συνέχεια φιλτράρεται σύμφωνα με τις πληροφορίες περιβάλλοντος που σχετίζονται με τον χρήστη.
- **Contextual modeling:** είναι μια τεχνική όπου οι πληροφορίες επίγνωσης πλαισίου χρησιμοποιούνται απευθείας στη συνάρτηση σύστασης. Αντί για χρήση μια παραδοσιακής 2D προσέγγιση είτε πριν είτε μετά, όπως οι δύο πρώτες προσεγγίσεις, αυτό θα χρησιμοποιήσει ένα πολυδιάστατο μοντέλο. Έχουν αναπτυχθεί διαφορετικοί αλγόριθμοι, αλλά δεν θα αναφερθούν εδώ σε περισσότερες λεπτομέρειες.

Όπως έχει αναφερθεί και σε προηγούμενες, συνδυάζοντας διαφορετικές τεχνικές συστάσεων, είναι δυνατό να γίνει το ίδιο εντός του CARS, συνδυάζοντας πολλές από τις στρατηγικές που περιγράφονται παραπάνω.

Απόκτηση πληροφοριών επίγνωσης πλαισίου

Οι G. Adomavicius και ο A. Tuzhilin [4] περιγράφουν τρεις τρόπους για την απόκτηση αυτών των πληροφοριών:

- Ζητώντας ρητά από τον χρήστη ερωτήσεις που βοηθούν στον καθορισμό αυτών των πληροφοριών.
- Σιωπηρή λήψη πληροφοριών από τα δεδομένα ή το περιβάλλον, π.χ. χρησιμοποιώντας το GPS σε κινητό τηλέφωνο για τον προσδιορισμό της θέσης του χρήστη.
- Λήψη συμπερασμάτων για το πλαίσιο χρησιμοποιώντας μεθόδους στατιστικής ή εξόρυξης δεδομένων. Για παράδειγμα, ενδέχεται να μπορεί να προσδιοριστεί ποιος παρακολουθεί τηλεόραση σε μια οικογένεια με βάση αυτό που παρακολουθείται τώρα.

Εκτός από τον τρόπο συλλογής των πληροφοριών, είναι εξίσου σημαντικό να αποφασίζεται ποιες πληροφορίες χρειάζεται να συλλεχθούν, προσφέροντας πραγματικά βοήθεια στην βελτίωση του συστήματος. Ο G. Adomavicius [60] προτείνει να χρησιμοποιηθεί ένα ευρύ φάσμα χαρακτηριστικών, τα οποία επιλέγονται από ειδικούς που πιστεύουν ότι οι συγκεκριμένες πληροφορίες θα είναι χρήσιμες. Μετά τη συλλογή των δεδομένων, διάφορες στατιστικές δοκιμές μπορούν να βρουν τα πιο σημαντικά χαρακτηριστικά.

Προβλήματα

Τέλος, υπάρχουν και ορισμένα προβλήματα στην συγκεκριμένη περίπτωση. Πρέπει να γίνει περισσότερη έρευνα σχετικά με τον τρόπο συνδυασμού των διαφορετικών προσεγγίσεων και των αντισταθμίσεων για καθεμία από αυτές. Η προσθήκη του πλαισίου στα συστήματα σύστασης αυξάνει επίσης την πολυπλοκότητα και μέχρι στιγμής το μεγαλύτερο μέρος της εργασίας ήταν εννοιολογική. Για να χρησιμοποιηθεί σε εφαρμογές πραγματικού κόσμου με τόνους δεδομένων, χρειάζεται να γίνει περισσότερη έρευνα για προβλήματα όπως δομές δεδομένων και μεθόδους αποθήκευσης.

Υπάρχει επίσης πρόβλημα στις περιπτώσεις όπου το πλαίσιο λαμβάνεται ρωτώντας τον χρήστη. Αυτές οι ερωτήσεις θα απαιτήσουν περισσότερη ηλεκτρονική ταξινόμηση από τον χρήστη, και τα περισσότερα συστήματα συστάσεων θέλουν να ανακτήσουν όσο το δυνατόν περισσότερες σχετικές πληροφορίες χωρίς να δώσουν μεγαλύτερη πίεση στον χρήστη [64].

2.2.5 Συστήματα συστάσεων πολλαπλών κριτηρίων

Τα *συστήματα συστάσεων πολλαπλών κριτηρίων* Multi-criteria recommender systems (MCRS) μπορούν να οριστούν ως συστήματα συστάσεων που ενσωματώνουν πληροφορίες προτίμησης σε πολλαπλά κριτήρια. Αντί να αναπτύσσουν τεχνικές προτάσεων που βασίζονται σε μία τιμή κριτηρίου,

τη συνολική προτίμηση του χρήστη u για το στοιχείο i , αυτά τα συστήματα προσπαθούν να προβλέψουν μια βαθμολογία για ανεξερεύνητα στοιχεία του u εκμεταλλευόμενοι πληροφορίες προτιμήσεων σε πολλά κριτήρια που επηρεάζουν αυτήν τη συνολική τιμή προτίμησης. Αρκετοί ερευνητές προσεγγίζουν το MCRS ως *πρόβλημα λήψης αποφάσεων πολλαπλών κριτηρίων* (Multi-criteria decision making -MCDM) και εφαρμόζουν μεθόδους και τεχνικές MCDM για την εφαρμογή συστημάτων MCRS.

2.2.6 Συστήματα συστάσεων επίγνωσης ρίσκου

Η πλειονότητα των υπαρχουσών προσεγγίσεων στα συστήματα σύστασης προτείνει τα πιο σχετικά αντικείμενα στους χρήστες χρησιμοποιώντας πληροφορίες με βάση τα συμφραζόμενα, αλλά δεν λαμβάνουν υπόψη τον κίνδυνο ενοχλήσεως του χρήστη με ανεπιθύμητες ειδοποιήσεις. Είναι σημαντικό να λαμβάνεται υπόψη ο κίνδυνος αναστάτωσης του χρήστη προωθώντας προτάσεις σε ορισμένες περιπτώσεις. Για παράδειγμα, κατά τη διάρκεια επαγγελματικής συνάντησης, χωρίς το πρωί ή αργά το βράδυ. Επομένως, η απόδοση του συστήματος σύστασης εξαρτάται εν μέρει από τον βαθμό στον οποίο έχει ενσωματώσει τον κίνδυνο στη διαδικασία σύστασης. Μία επιλογή για τη διαχείριση αυτού του ζητήματος είναι το DRARS, ένα σύστημα επίγνωσης πλαισίου που διαμορφώνεται ως multi-armed bandit problem. Αυτό το σύστημα συνδυάζει CBF και έναν αλγόριθμο επίγνωσης πλαισίου.

2.2.7 Συστήματα συστάσεων κινητών συσκευών

Τα συστήματα συστάσεων κινητών συσκευών κάνουν χρήση του διαδικτύου μέσω των έξυπνων συσκευών ώστε να προσφέρουν εξατομικευμένες, ευαίσθητες στο πλαίσιο επίγνωσης προτάσεις. Πρόκειται για έναν ιδιαίτερα δύσκολο τομέα έρευνας, καθώς τα δεδομένα κινητής τηλεφωνίας είναι πιο περίπλοκα από τα δεδομένα που συχνά πρέπει να διαχειριστούν τα συστήματα συστάσεων. Είναι ετερογενής, θορυβώδης, απαιτείται χωρική και χρονική αυτόματη συσχέτιση και έχει προβλήματα επικύρωσης και γενικότητας. Ένα παράδειγμα ενός συστήματος συστάσεων για κινητές συσκευές είναι οι προσεγγίσεις που έχουν υιοθετήσει εταιρείες όπως η Uber και η Lyft για τη δημιουργία διαδρομών οδήγησης για οδηγούς ταξί σε μια πόλη. Αυτό το σύστημα χρησιμοποιεί δεδομένα GPS των διαδρομών που παίρνουν οι οδηγοί ταξί κατά τη διάρκεια της εργασίας τους, τα οποία περιλαμβάνουν τοποθεσία (γεωγραφικό πλάτος και μήκος), χρονικές σφραγίδες και 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 πληρότητας και των κερδών.

2.2.8 Συστήματα συστάσεων βασισμένα στη συνεδρία

Αυτά τα συστήματα σύστασης χρησιμοποιούν τις αλληλεπιδράσεις ενός χρήστη εντός μιας περιόδου σύνδεσης. Τα συστήματα συστάσεων βασισμένα στη συνεδρία χρησιμοποιούνται συνήθως στο Youtube και στο Amazon. Είναι ιδιαίτερα χρήσιμα όταν το ιστορικό (όπως προηγούμενα κλικ, αγορές) ενός χρήστη δεν είναι διαθέσιμο ή δεν είναι σχετικό στην τρέχουσα περίοδο σύνδεσης. Στις περισσότερες περιπτώσεις τέτοιου είδους συστημάτων σύστασης, το σύστημα βασίζεται στην ακολουθία των πρόσφατων αλληλεπιδράσεων σε μια περίοδο λειτουργίας χωρίς να απαιτούνται πρόσθετες λεπτομέρειες (ιστορικό, δημογραφικό) του χρήστη. Για την υλοποίηση τους χρησιμοποιούνται κυρίως γενετικά διαδοχικά μοντέλα όπως Recurrent Neural Networks, Transformers και άλλες προσεγγίσεις με βάση τη βαθιά μάθηση.

2.2.9 Συστήματα συστάσεων ενισχυτικής μάθησης

Το πρόβλημα των συστάσεων μπορεί να θεωρηθεί ως μια ειδική περίπτωση ενός προβλήματος Ενισχυτικής Μάθησης όπου ο χρήστης είναι το περιβάλλον στο οποίο ενεργεί ο πράκτορας (agent), το σύστημα προτάσεων προκειμένου να λάβει μια ανταμοιβή, για παράδειγμα, ένα κλικ ή μια αλληλεπίδραση από τον χρήστη. Μια πτυχή της Ενισχυτικής Μάθησης που έχει ιδιαίτερη χρήση στον

τομέα των συστημάτων σύστασης είναι το γεγονός ότι τα μοντέλα ή οι πολιτικές μπορούν να μάθουν παρέχοντας μια ανταμοιβή στον παράγοντα σύστασης. Αυτό έρχεται σε αντίθεση με τις παραδοσιακές τεχνικές μάθησης που βασίζονται στην Επιβλεπόμενη Μάθηση και είναι λιγότερο ευέλικτες. Οι τεχνικές Ενισχυτικής Μάθησης επιτρέπουν να εκπαιδεύσουν δυνητικά μοντέλα που μπορούν να βελτιστοποιηθούν άμεσα σε μετρήσεις αφοσίωσης και τις προτιμήσεις του χρήστη.

2.3 Τρόπος αξιολόγησης

Υπάρχει ποικιλία μεθόδων αξιολόγησης που έχουν χρησιμοποιηθεί στον τομέα των συστημάτων σύστασης. Μόνο με τη μελέτη συστημάτων σύστασης μπορούν να επινοηθούν νέες και βελτιωμένες μέθοδοι [65]. Ανάλογα με τον στόχο του συστήματος, μπορεί να γίνει και η αντίστοιχη αξιολόγηση. Για παράδειγμα, υπάρχουν δύο βασικές πτυχές: πρώτον, η πρόβλεψη του αντικείμενου και ο τρόπος με τον οποίο ανταποκρίνεται στα ενδιαφέροντα του χρήστη, η οποία ονομάζεται «πρόβλεψη αξιολόγησης». Αυτό λαμβάνει υπόψη μόνο τις παρατηρούμενες βαθμολογίες και μετράται από το μέσο απόλυτο σφάλμα *mean absolute error* (MAE) και το αρχικό τετραγωνικό σφάλμα *Root Mean Square Error* (RMSE). Δεύτερον, δυνατότητα παροχής μιας λίστας αντικειμένων που συνιστώνται σε έναν συγκεκριμένο χρήστη. Αυτό ονομάζεται «κατάταξη» και συχνά ποσοτικοποιείται από την άποψη της ακρίβειας (precision) και της ανάκλησης (recall). Σε κάθε περίπτωση, υπάρχουν κοινοί πίνακες που εφαρμόζονται σε πολλά ερευνητικά άρθρα. Ωστόσο, υπάρχουν διαφορετικές διαδικασίες αξιολόγησης. Υπάρχουν τρεις κύριες μέθοδοι, και αυτές εξηγούνται λεπτομερέστερα στις επόμενες υπό-ενότητες. Η επιλογή της μεθόδου εξαρτάται από τα πειράματα που διεξάγονται από τους ερευνητές και από το εάν χρησιμοποιούν πραγματικά συστήματα ή δημόσια σύνολα δεδομένων.

2.3.1 Offline αξιολόγηση

Λόγω της διαθεσιμότητας του συνόλου δεδομένων, αυτός ο τύπος αξιολόγησης είναι ο ευκολότερος στην εκτέλεση σε πειράματα χρησιμοποιώντας ένα υπάρχον σύνολο δεδομένων που παρέχει πληροφορίες χρήστη με αξιολογήσεις για ορισμένα στοιχεία [4]. Η πλειοψηφία των υπαρχόντων εργασιών για την αξιολόγηση συστημάτων σύστασης επικεντρώνεται κυρίως στην προγνωστική ακρίβεια χρησιμοποιώντας offline ανάλυση [66]. Μια τέτοια αξιολόγηση έχει το πλεονέκτημα ότι δεν χρειάζεται να ζητηθεί από τους πραγματικούς χρήστες να συμμετάσχουν, γεγονός που δίνει στους προγραμματιστές περισσότερες επιλογές για την εκτέλεση του πειράματος με περισσότερους από έναν αλγόριθμους και διαφορετικά σύνολα δεδομένων. Χρησιμοποιείται για τη διεξαγωγή τόσο πρόβλεψης όσο και N κορυφαίων προτάσεων. Το μεγαλύτερο μέρος της έρευνας σε συστήματα συστάσεων εστιάζεται στην ανάλυση offline της προβλεπτικής ακρίβειας και στην αξιολόγηση της αποτελεσματικότητας των συστάσεων. Ο Ahn [67] εφαρμόζει το MAE για τη μέτρηση της αποτελεσματικότητας μιας νέας μεθόδου επίλυσης του προβλήματος σύστασης ψυχρής εκκίνησης (cold start). Ο Hannon [68] χρησιμοποίησε την offline αξιολόγηση για να προτείνει ακόλουθους στο Twitter με βάση το προφίλ κάθε χρήστη.

Η αξιολόγηση εκτός σύνδεσης παρέχει έναν τυπικό τρόπο μέτρησης της απόδοσης των αλγορίθμων προτάσεων. Παρέχει επίσης τη δυνατότητα διεξαγωγής μιας πιο αποδοτικής σύγκρισης μεταξύ των προτεινόμενων μεθόδων σύστασης [69]. Ο Shenghui [70] εφάρμοσε το MAE σε μια αξιολόγηση της ακρίβειας της σημασιολογικής βελτιστοποιημένης πρότασης περιεχομένου. Πιο πρόσφατα, Οι Wasid και Ali [71] αξιολόγησαν τη χρήση πολλαπλών κριτηρίων έναντι των παραδοσιακών τεχνικών CF χρησιμοποιώντας προγνωστική ακρίβεια MAE.

Η αξιολόγηση εκτός σύνδεσης μπορεί να χωριστεί σε δύο κύριες αξιολογήσεις προτάσεων: μια λίστα με τις N κορυφαίες προτάσεις και την πρόβλεψη αξιολόγησης. Σε κάθε μέθοδο αξιολόγησης, λαμβάνονται υπόψη δύο παράγοντες: αποδοτικότητα και αποτελεσματικότητα. Όσον αφορά την αποτελεσματικότητα, η πρόβλεψη βαθμολογίας χρησιμοποιείται ευρέως, η οποία υπολογίζει την προβλεπόμενη βαθμολογία για τις βαθμολογίες που λείπουν και αξιολογεί τους αλγόριθμους που προβλέπουν αυτές τις τιμές. Με άλλα λόγια, η μέτρηση σφάλματος πρόβλεψης ακρίβειας εφαρμόζεται ευρέως

στην αξιολόγηση εκτός σύνδεσης, η οποία περιλαμβάνει MAE, MSE και RMSE. Μετρά τις διαφορές μεταξύ της πραγματικής βαθμολογίας και της προβλεπόμενης βαθμολογίας

Η πτυχή της αποδοτικότητας, από την άλλη πλευρά, μετράται από το χρόνο που αφιερώνεται για την επεξεργασία του αλγορίθμου ή τον χρόνο απόκρισης όσον αφορά τη δημιουργία της πρότασης. Ωστόσο, η αξιολόγηση της αποτελεσματικότητας παίζει πιο σημαντικό ρόλο από την αξιολόγηση της αποδοτικότητας για προσεγγίσεις συστάσεων.

Παρά το γεγονός ότι η ακρίβεια των προτάσεων αποτελεί σημαντικό παράγοντα για τη δημιουργία ενός αποτελεσματικού αλγορίθμου προτάσεων, η ικανοποίηση των χρηστών πρέπει επίσης να θεωρηθεί εξίσου σημαντική στην αξιολόγηση. Πιο συγκεκριμένα, με την ανάπτυξη ενός νέου συστήματος απαιτείται μια συγκεκριμένη παρατήρηση αλληλεπίδρασης για κάθε χρήστη ξεχωριστά, για τη μέτρηση της αποτελεσματικότητας. Ορισμένοι χρήστες μπορεί να θεωρήσουν ότι η πρόταση είναι πολύ σχετική και άλλοι μπορεί να μην τη βρίσκουν χρήσιμη.

2.3.2 Online αξιολόγηση

Σε ένα online σύστημα σύστασης, οι χρήστες επηρεάζονται από το σύστημα και αλληλεπιδρούν άμεσα με αυτό, συνεπώς ο σχεδιαστής στοχεύει να μετρήσει τη συμπεριφορά των χρηστών κατά τη χρήση του συστήματος. Επομένως, μπορεί να αντικατοπτρίζει πόσο καλά το σύστημα προτείνει σχετικά / ενδιαφέροντα προϊόντα [72].

Από τις διαδικτυακές διαφημίσεις εμφανίστηκαν για πρώτη φορά οι διαδικτυακές αξιολογήσεις και έκτοτε χρησιμοποιούνται για να εκτιμήσουν την αποδοχή των συστάσεων που προσφέρονται από τα συστήματα σύστασης. Οι αναλογίες κλικ προς αριθμό εμφανίσεων Clickthrough rate (CTR) χρησιμοποιούνται ως διακομιστής μεσολάβησης για τα ποσοστά αποδοχής. Για παράδειγμα, εάν γίνουν 1.000 προτάσεις και γίνει κλικ σε οκτώ, η αναλογία κλικ προς αριθμό εμφανίσεων είναι 0,8%. Εναλλακτικά μέτρα αποδοχής περιλαμβάνουν την αναλογία αντικειμένων που αγοράστηκαν ή την αναλογία λήψεων. Η αποδοχή μπορεί να υπονοείται επειδή εάν ένα άτομο έχει κάνει κλικ, κατεβάσει ή αγοράσει ένα στοιχείο που έχει προταθεί, τότε πιθανώς βρήκε ότι η πρόταση ήταν χρήσιμη. Ωστόσο, αυτό δεν ισχύει απαραίτητα, επειδή είναι πιθανό ένας χρήστης να αγοράσει ένα προϊόν και αργότερα να συνειδητοποιήσει ότι δεν είναι κατάλληλο για τις ανάγκες του. Εάν ο σκοπός του συστήματος σύστασης είναι απλώς η μεγιστοποίηση των εσόδων, η απόδοση μπορεί να μετρηθεί χρησιμοποιώντας μετρήσεις όπως το CTR [65].

Πιο έμμεσα, πολλές επιχειρήσεις και οργανισμοί εκτελούν ελεγχόμενα πειράματα στα συστήματά τους με τη μορφή δοκιμών A / B ή εναλλακτικών τεχνικών [73]). Συνήθως, αυτά τα πειράματα ανακατευθύνουν ένα μέρος της επισκεψιμότητας μιας πλατφόρμας προς το σύστημα που αξιολογείται και μετρά την απόδοση του συστήματος μέσω μετρήσεων αφοσίωσης χρηστών όπως προβολές σελίδας, CTR [74] ή, πιο άμεσα, το οικονομικό όφελος του συστήματος [75]. Αυτό το είδος αξιολόγησης παρέχει τα ισχυρότερα στοιχεία, επειδή πραγματοποιείται σε πραγματικούς χώρους με πραγματικούς χρήστες. Ωστόσο, τα αποτελέσματα αυτών των ειδών πειράματος πρέπει να αναλυθούν προσεκτικά προκειμένου να εξαχθούν αξιόπιστα συμπεράσματα και να απορριφθούν διαφορές που μπορεί να προκληθούν από εξωτερικούς παράγοντες ή πιθανότητες. Επιπλέον, υπάρχει κίνδυνος για την πραγματοποίηση αξιολογήσεων σε πραγματικά συστήματα, επειδή η δοκιμή συστημάτων με χαμηλή απόδοση ενδέχεται να επηρεάσει αρνητικά τις εμπειρίες των πραγματικών πελατών.

Συχνά τα πειράματα όπως αυτά είναι δαπανηρά, επειδή η επινόηση συστημάτων για διαδικτυακές δοκιμές είναι επίπονη. Επιπλέον, είναι απαραίτητο να δοκιμάζονται τους αλγόριθμους πριν από την παρουσίαση των αποτελεσμάτων στους χρήστες, καθώς αυτό αποτρέπει τους χρήστες να υποστούν αρνητικές εμπειρίες. Για παράδειγμα, εάν οι χρήστες δέχονται λάθος συστάσεις κατά τη διάρκεια της δοκιμής, είναι απίθανο να επιστρέψουν για να χρησιμοποιήσουν το τελικό σύστημα στο μέλλον. Επιπλέον, εάν τα συστήματα προτάσεων ενσωματώνονται σε εφαρμογές πριν από την ανάπτυξη, δεν υπάρχει πιθανότητα διεξαγωγής δοκιμών. Επομένως, είναι επιτακτική ανάγκη οι δοκιμές απόδοσης αλγορίθμου να διεξάγονται εκτός σύνδεσης ενώ μιμούνται στενά την online συμπεριφορά [72]

Κεφάλαιο 3

Διαστάσεις επίγνωσης πλαισίου

3.1 Περιγραφή βασικών διαστάσεων

Έχει σημειωθεί μεγάλη πρόοδος στα συστήματα συστάσεων από τότε που αναγνωρίστηκε η μεγάλη αξία της επίγνωσης πλαισίου στην παροχή προτάσεων [4]. Οι επιλογές των χρηστών δεν καθοδηγούνται αποκλειστικά από ένα σταθερό σύνολο δεδομένων που σχετίζονται με τα χαρακτηριστικά των αντικειμένων, ενώ επίσης δεν μπορούν να καλυφθούν πλήρως από προσεγγίσεις CBF και CF. Αντιθέτως υπάρχει εξάρτηση σε μεγάλο βαθμό από το πλαίσιο, τις τρέχουσες ανάγκες τους και την κατάσταση στην οποία βρίσκονται. Υπάρχουν διάφοροι τύποι πληροφοριών επίγνωσης πλαισίου που θα μπορούσαν να βοηθήσουν στην παροχή καλύτερων συστάσεων όταν ληφθούν υπόψη, τα κύρια παραδείγματα είναι:

- **Πληροφορίες Τοποθεσίας:** Οι πληροφορίες τοποθεσίας μπορεί να είναι χρήσιμες σε μια ποικιλία σεναρίων προτάσεων [76]. Η γεωγραφική τοποθεσία του χρήστη μπορεί να επηρεάσει σημαντικά τις προτιμήσεις όσον αφορά τη γεύση, τον πολιτισμό και τις συνήθειες. Μπορεί επίσης να υποδεικνύει τον τύπο της τρέχουσας δραστηριότητας του χρήστη, π.χ. την εργασία στο γραφείο ή τον ελεύθερο χρόνο σε ένα διασκεδαστικό μέρος, το οποίο με τη σειρά του επηρεάζει το σύνολο των σχετικών προτιμήσεων. Σε ορισμένες συγκεκριμένες εφαρμογές, π.χ. σύσταση εστιατορίου, η τοποθεσία του χρήστη είναι απαραίτητη για το φιλτράρισμα των απρόσιτων επιλογών.
- **Χρονικές πληροφορίες:** Ανάλογα με την ώρα της ημέρας, την ημέρα της εβδομάδας ή την εποχή, οι χρήστες δεν εμφανίζουν τις ίδιες προτιμήσεις σε διάφορους τομείς, π.χ. ρούχα και τουρισμός. Επομένως, οι χρονικές πληροφορίες έχουν αξιοποιηθεί για τη βελτίωση της ποιότητας της σύστασης [77].
- **Κοινωνικές πληροφορίες:** Η συμπεριφορά του χρήστη ποικίλλει ανάλογα με τους κοινωνικούς παράγοντες. Για παράδειγμα, ένας χρήστης μπορεί να κάνει διαφορετικές επιλογές ανάλογα με το αν βρίσκεται με παρέα των φίλων ή της οικογένειας [4]. Επιπλέον, και ειδικά στο πλαίσιο των διαδικτυακών κοινωνικών δικτύων, οι συμπεριφορές των στενών χρηστών αναμένεται να επηρεάσουν τον χρήστη και θα πρέπει να λαμβάνονται υπόψη κατά την παροχή προτάσεων [78].

Όπως αναφέρεται στο Κεφάλαιο 2, η συντριπτική πλειοψηφία των συστημάτων συστάσεων επίγνωσης πλαισίου υιοθετεί την αντιπροσωπευτική άποψη του πλαισίου [Dourish, 2004]. Αυτή η άποψη κάνει τις ακόλουθες παραδοχές: (i) το πλαίσιο είναι μια μορφή πληροφοριών: Είναι κάτι που μπορεί να γίνει γνωστό. (ii) το πλαίσιο είναι οριοθετημένο: Είναι δυνατόν να καθοριστεί τι θεωρείται ως το πλαίσιο δραστηριοτήτων για κάποια σειρά εφαρμογών. (iii) το πλαίσιο είναι σταθερό: Τα στοιχεία του πλαισίου δεν διαφέρουν μεταξύ των περιπτώσεων μιας δραστηριότητας στην ίδια εφαρμογή. (iv) το πλαίσιο και οι δραστηριότητες διαχωρίζονται: Μια δραστηριότητα συμβαίνει μέσα σε ένα πλαίσιο. Σε σύγκριση, η αλληλεπιδραστική άποψη υποστηρίζει ότι: (i) το πλαίσιο είναι μια σχεσιακή ιδιότητα μεταξύ αντικειμένων ή δραστηριοτήτων: Μια διάσταση μπορεί να είναι σχετική ή όχι με το περιεχόμενο σε μια συγκεκριμένη δραστηριότητα · (ii) το πεδίο των χαρακτηριστικών επίγνωσης πλαισίου

ορίζεται δυναμικά. (iii) το πλαίσιο είναι μια περιστασιακή ιδιοκτησία που σχετίζεται με συγκεκριμένα γεγονότα (iv) το πλαίσιο προκύπτει από τη δραστηριότητα. Ενώ η πρώτη άποψη θεωρεί ότι το πλαίσιο είναι ένα σταθερό χαρακτηριστικό του περιβάλλοντος που είναι ανεξάρτητο από τις ενέργειες των ατόμων, η δεύτερη άποψη αναλαμβάνει αμφίδρομη σχέση μεταξύ πλαισίου και δραστηριοτήτων.

3.2 Παραδείγματα χρήσης επίγνωσης πλαισίου

3.2.1 Τεχνολογία πληροφορίας και τουρισμός

Ενώ ο τουρισμός έχει σημαντικό οικονομικό αντίκτυπο, η βιομηχανία έχει δει δραματικές αλλαγές τα τελευταία χρόνια [79]. Η τουριστική συμπεριφορά αλλάζει με την εξέλιξη της ταξιδιωτικής εμπειρίας και με την αλλαγή προσδοκιών και κινήτρων. Παρουσιάζει σήμερα ποικίλες προσωπικότητες που είναι δύσκολο να αποκωδικοποιηθούν. Επιπλέον, η ταχεία ανάπτυξη της πληροφορικής έχει συμβάλει σε αυτόν τον ριζικό μετασχηματισμό [80], ειδικά λαμβάνοντας υπόψη τον τρόπο πρόσβασης σε πληροφορίες και αγορά προϊόντων. Από την άλλη πλευρά, το πεδίο της πληροφορικής ανοίγει επίσης νέες επιχειρηματικές ευκαιρίες με αποτέλεσμα τη δημιουργία πολλών αξιόλογων νέων στρατηγικών και υπόσχεται να βελτιώσει την εμπειρία του τουρισμού [81]. Οι προκλήσεις που αντιμετωπίζουν αυτές οι εφαρμογές απαιτούν τον συνδυασμό έρευνας και ανάπτυξης στη διασταύρωση διαφόρων κλάδων, όπως η επιστήμη των υπολογιστών, οι γνωστικές τεχνολογίες και η τουριστική έρευνα. Τα σχετικά ερευνητικά θέματα μπορούν να χωριστούν σε πέντε επίπεδα [80], καλύπτοντας το τουριστικό οικοσύστημα, τα οποία αναφέρονται στα ακόλουθα: (i) ατομικό, (ii) ομαδικό, (iii) εταιρικό, (iv) βιομηχανία, και (v) κυβέρνηση.

Το κάθε επίπεδο επικεντρώνεται στην αλληλεπίδραση μεταξύ οποιουδήποτε χρήστη από την πλευρά της ζήτησης ή της προσφοράς και οποιασδήποτε υπηρεσίας πληροφορικής. Δεδομένης της ετερογένειας των χρηστών, υπάρχει ανάγκη για εξατομικευμένες εφαρμογές και προσαρμοστικά μοντέλα, προκειμένου να βελτιωθεί η συνολική εμπειρία και να μειωθεί το γνωστικό φορτίο των χρηστών. Οι ομάδες σχηματίζονται από δύο ή περισσότερους χρήστες και μπορούν να αναγνωριστούν ότι μοιράζονται παρόμοιες προτιμήσεις ή ότι παρευρίσκονται μαζί στην ίδια κατάσταση. Μπορούν να αναπτυχθούν συγκεκριμένες υπηρεσίες για την υποστήριξη σχηματισμών ομάδων, εμπειριών ομάδας ή λήψης αποφάσεων σε ομάδες. Το εταιρικό επίπεδο αφορά οποιονδήποτε οργανισμό στον τομέα του τουρισμού και καλύπτει θέματα όπως η διαχείριση γνώσεων, η κατανόηση των πελατών και η διαδικτυακή φήμη. Το επίπεδο της βιομηχανίας συλλαμβάνει τη συνολική δομή του τομέα και τις αλληλεπιδράσεις μεταξύ οργανισμών. Υπογραμμίζεται ότι το πιο σημαντικό είναι η άνοδος των διαδικτυακών ταξιδιωτικών γραφείων. Τέλος, το κυβερνητικό επίπεδο ασχολείται με κανόνες και κανονισμούς της πληροφορικής στον τομέα του τουρισμού.

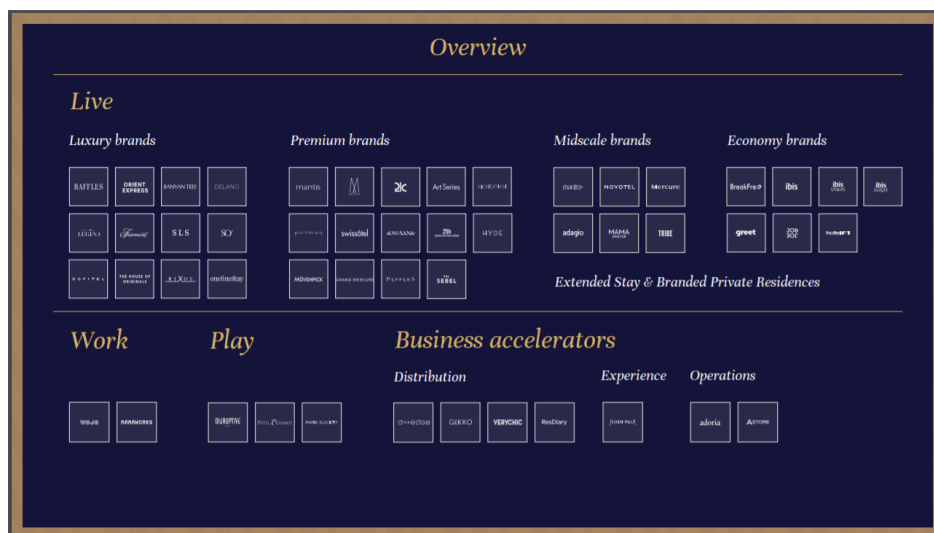
RS στον τουρισμό

Τα συστήματα συστάσεων παρεμβαίνουν στα ατομικά και ομαδικά επίπεδα και καθίστανται απαραίτητα στον τομέα του τουρισμού. Οι ταξιδιώτες αναζητούν ενεργά πληροφορίες για να συνθέσουν τα πακέτα διακοπών τους και αντιμετωπίζουν το ευρέως γνωστό πρόβλημα της υπερφόρτωσης πληροφοριών [82]. Τα συστήματα συστάσεων είναι σε θέση να υποστηρίξουν ταξιδιώτες με διάφορους τρόπους και να τους συνοδεύουν καθ' όλη τη διάρκεια της εμπειρίας. Μπορούν να προτείνουν προορισμούς που ταιριάζουν στις προτιμήσεις των χρηστών κατά την επιλογή προορισμού πριν από το ταξίδι, και στη συνέχεια καταλύματα, αξιοθέατα, εστιατόρια, εκδηλώσεις και σημεία ενδιαφέροντος, εκτός από εξατομικευμένες διαδρομές που καθοδηγούν τους ταξιδιώτες σε πολλά αξιοθέατα. Από επιχειρηματική άποψη, συμμετέχουν στην αύξηση της πιστότητας των χρηστών βελτιώνοντας τις εμπειρίες τους. Ως εκ τούτου, η εξατομίκευση γίνεται προτεραιότητα για τους οργανισμούς, ειδικά λόγω του σκληρού ανταγωνισμού στην αγορά.

RS στον ξενοδοχειακό τομέα

Κατά την εκπόνηση ταξιδιωτικών σχεδίων, η διαμονή αποτελεί ουσιαστικό μέρος της διαδικασίας και η επιλογή της καταλληλότερης είναι μια επίπονη και χρονοβόρα εργασία για τους περισσότερους ανθρώπους. Όπως σε οποιονδήποτε βιομηχανικό τομέα όπου αναπτύσσεται τα συστήματα συστάσεων, έτσι και στον ξενοδοχειακό κλάδο διευκολύνει τον προγραμματισμό ταξιδιών και βοηθά στην αύξηση της κερδοφορίας των ξενοδοχειακών εταιρειών και την πίστη των χρηστών. Το ιδιαίτερο πρόβλημα της παροχής ξενοδοχειακών προτάσεων είναι ένα προκλητικό πρόβλημα που είχε σχετική σχετικά μικρή έρευνα. Θέτει συγκεκριμένα ζητήματα που δεν μπορούν να αντιμετωπιστούν χρησιμοποιώντας παραδοσιακές προσεγγίσεις για σύσταση. Λαμβάνοντας υπόψη τη συμπεριφορά του παρελθόντος, ο στόχος είναι να μαθαίνονται προτιμήσεις για ξενοδοχεία και να χρησιμοποιούνται για την εκτέλεση προτάσεων ξενοδοχείων. Το υπόλοιπο αυτής της ενότητας επικεντρώνεται στο πρόβλημα της σύστασης του ξενοδοχείου και τα κύρια χαρακτηριστικά του μέσω μιας υλοποίησης.

Το AccorHotels προσφέρει μια τεράστια επιλογή καταλυμάτων σε περισσότερες από 90 χώρες σε όλες τις ηπείρους. Λειτουργεί με πάνω από 4.500 ξενοδοχεία και συγκεντρώνει πάνω από 35 μάρκες ξενοδοχείων που αντιπροσωπεύονται στο Σχήμα 3.1. Αυτές οι μάρκες κυμαίνονται από οικονομικές έως πολυτελές και προσπαθούν να καλύψουν τις ανάγκες των ταξιδιωτών που μετακινούνται σε όλο τον κόσμο για διαφορετικούς λόγους. Σε μια προσπάθεια να διατηρήσουν τους ταξιδιώτες, η ομάδα παρουσίασε το Le Club AccorHotels πρόγραμμα πίστης. Οι εγγεγραμμένοι χρήστες επωφελούνται από προνομιακές τιμές, έγκαιρη πρόσβαση σε ιδιωτικές πωλήσεις, πλεονεκτήματα από συνεργάτες και επιπλέον ανταμοιβές από το δίκτυο των ιδιοκτητών.



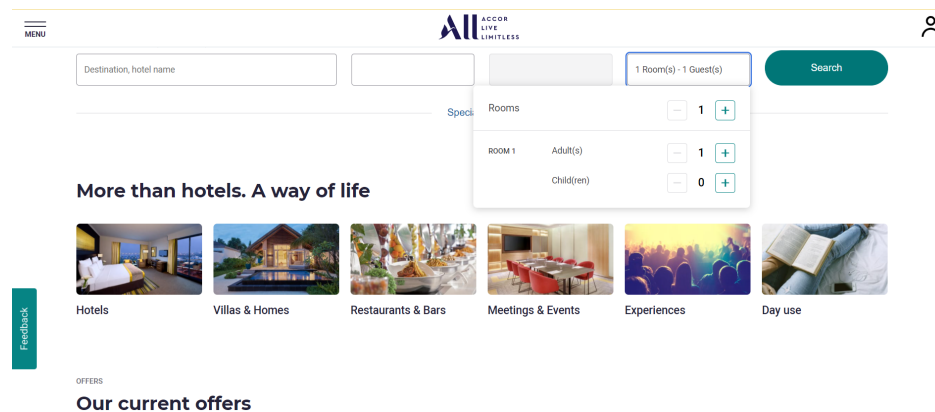
Σχήμα 3.1: Επωνυμίες που ανήκουν στην AccorHotels[5]

Προκειμένου να προσφέρει την καλύτερη εμπειρία στους πελάτες της, η AccorHotels ανακοίνωσε στα τέλη του 2014 την έναρξη του ψηφιακού της μετασχηματισμού με ένα πενταετές επενδυτικό σχέδιο. Ένας από τους κύριους πυλώνες αυτού του σχεδίου αφορά την εξατομίκευση και προσαρμογή των υπηρεσιών για την αύξηση της πιστότητας των πελατών. Σε έναν κόσμο όπου τα ταξίδια έχουν γίνει συνήθεια για εκατομμύρια ανθρώπους, οι προσδοκίες αλλάζουν και οι τουριστικοί φορείς πρέπει να προσπαθήσουν να εγγυηθούν την ικανοποίηση των πελατών. Οι υπάρχοντες και νέοι διαδικτυακοί φορείς, π.χ. τα Expedia και Airbnb, αλλάζουν επίσης τη δομή της αγοράς και καθιστούν πιο δύσκολη την προσέλκυση πελατών. Ενώ η εξατομίκευση έχει γίνει απαραίτητη, αναμένεται επίσης να συνοδεύει τους ταξιδιώτες σε κάθε βήμα του ταξιδιού τους. Πριν από το ταξίδι, χρειάζονται βοήθεια για να επιλέξουν μια κατάλληλη επιλογή για διαμονή, να προετοιμάσουν τη διαμονή τους και να χειριστούν τα προβλήματα μεταφοράς για παράδειγμα. Κατά τη διάρκεια της διαμονής, πολλές υπηρεσίες μπορούν να προσφερθούν όπως υπηρεσία δωματίου, πρόσβαση σε ξενοδοχειακές εγκαταστάσεις, οργάνωση εκδρομών και κράτηση ταξί. Συλλέγεται αξιολόγηση από τους χρήστες μετά τη λήξη της διαμονής.

Χώρα	Φύλο	Πόλη	Ημ/Γέννησης	Εθνικότητα	Επάγγελμα	Προτιμήσεις
100%	93.6%	75.6%	16.7%	3.2%	2.3%	0.7%

Πίνακας 3.1: Ποσοστό διαθέσιμων δεδομένων ανά λειτουργία χρήστη για ένα σύνολο εγγεγραμμένων πελατών

Μέσω της ιστοσελίδας κρατήσεων όπως φαίνεται στο Σχήμα 3.2 συλλέγονται όλες οι κατάλληλες πληροφορίες. Όπως αναφέρθηκε και παραπάνω οι δυνατότητα αυτή παρέχεται και σε εγγεγραμμένους χρήστες και όχι, με τους πρώτους να επωφελούνται από συγκεκριμένες προσφορές καθώς και πιο εξατομικευμένες προτάσεις αφού υπάρχει το ιστορικό τους. Το σύστημα αντλεί πληροφορίες από το χρήστη όπως προορισμό (πληροφορίες τοποθεσίας), ημερομηνίες μετάβασης και αποχώρησης(χρονικές πληροφορίες) καθώς και με ποιους έχει σκοπό να ταξιδέψει (κοινωνικές πληροφορίες).



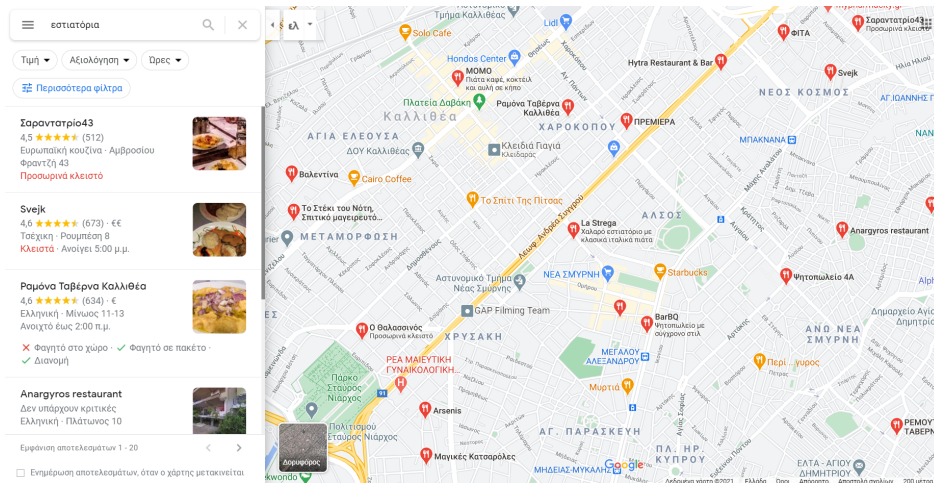
Σχήμα 3.2: Περιβαλλον κρατήσεων της Accor[5]

RS στον τομέα της εστίασης

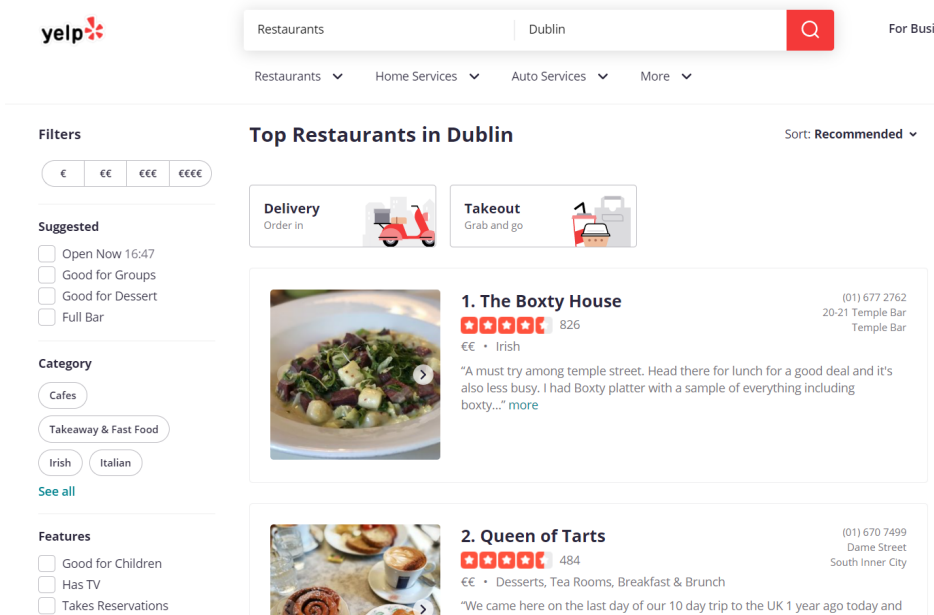
Υπάρχουν πολλές εφαρμογές προτάσεων εστιατορίων αυτή τη στιγμή διαθέσιμες για δημόσια χρήση και χρησιμοποιούνται από χρήστες σε όλο τον κόσμο. Ένα παράδειγμα θα ήταν οι Χάρτες Google, που χρησιμοποιούν το ερώτημα αναζήτησης ενός χρήστη και την τρέχουσα τοποθεσία για να προταθούν μέρη για φαγητό κοντά, όπως φαίνεται στο Σχήμα 3.3. Ο τρόπος λειτουργίας των Χαρτών Google είναι η χρήση της τρέχουσας τοποθεσίας του χρήστη ως το μοναδικό χαρακτηριστικό περιβάλλοντος και το ερώτημα αναζήτησης για την κατάταξη των κοντινών σημείων ενδιαφέροντος και στη συνέχεια παρουσίασή τους στον χρήστη. Αυτή η προσέγγιση είναι βολική και λειτουργεί όπως αναμένεται. Ωστόσο, δυστυχώς, η υπάρχουσα λύση δεν είναι προσαρμόσιμη σε συγκεκριμένους χρήστες, δεδομένου ότι δεν παρέχει τρόπο στους τελικούς χρήστες να προσαρμόσουν τις συστάσεις στις προτιμήσεις τους για τη δημιουργία εξατομικευμένων συστάσεων με βάση τις γαστρονομικές προτιμήσεις τους και άλλους σχετικούς παράγοντες.

Μια άλλη δημοφιλής λύση για σύσταση εστιατορίου είναι η εφαρμογή Yelp. Παρέχει στους χρήστες πολλές επιλογές, όπως επιλογή του εύρους τιμών, ταξινόμηση εστιατορίων από απόσταση και πολλές άλλες εξελιγμένες επιλογές, όπως φαίνεται στο σχήμα 3.4.

Αυτή η προσέγγιση λειτουργεί όπως αναμενόταν για τους περισσότερους χρήστες. Ωστόσο, η ίδια η εφαρμογή δεν προσαρμόζεται στις σιωπηρές προτιμήσεις των χρηστών καθώς τη χρησιμοποιούν όλο και περισσότερο και παρέχει μια στατική λίστα εστιατορίων τις περισσότερες φορές, η οποία είναι και πάλι αποδεκτή για τους περισσότερους ανθρώπους. Ωστόσο, για κάποιον που είναι λάτρης του φαγητού και θέλει το σύστημα να του προτείνει νέα μέρη για φαγητό σε κάθε νέα αναζήτηση βάσει της γαστρονομικής του προτίμησης και άλλων σχετικών παραγόντων, το Yelp δεν ανταποκρίνεται τόσο καλά.



Σχήμα 3.3: Google Maps [6]



Σχήμα 3.4: Αναζήτηση και φίλτρα [7]

Αυτό το ζήτημα έχει ληφθεί σοβαρά υπόψη στη λύση μας, καθώς ο κύριος στόχος αυτής της εργασίας ήταν να βοηθηθεί ένας χρήστης να βρει μέρη για φαγητό σύμφωνα με την προτίμησή του, άλλους σχετικούς παράγοντες επίγνωσης πλαισίου, καθώς και να τον βοηθήσει να ανακαλύψει νέα μέρη που ίσως δεν είχε επισκεφτεί στο παρελθόν και δεν θα είχε ανακαλύψει ποτέ αν δεν είχε χρησιμοποιήσει αυτή την εργασία.

Η τεράστια χρήση των κινητών συσκευών με υπηρεσίες διαδικτύου και τοποθεσίας καθιστά την τοποθεσία των χρηστών, τα σχετικά δεδομένα επίγνωσης πλαισίου διαθέσιμα για διάφορα είδη λύσεων, συμπεριλαμβανομένων συστάσεων. Επομένως, συστήματα συστάσεων με βάση την τοποθεσία, αλλά και γενικά συστήματα που κάνουν χρήση επίγνωσης πλαισίου σε ένα μεγάλο αριθμό εφαρμογών όπως προτάσεις ραντεβού, κοινωνικών δικτύων, φαγητού, ταξιδιωτικές και τουριστικές προτάσεις, προτάσεις ταινιών κ.λπ. Όπως αναμενόταν, έχουν ήδη γίνει αρκετές ερευνητικές εργασίες υψηλής ποιότητας στους σχετικούς τομείς.

Για παράδειγμα, υπήρξε παρόμοια δουλειά σχετικά με τις προτάσεις εστιατορίων που βασίζονται στην εκμετάλλευση της προτίμησης του φαγητού ενός χρήστη από τους Fuzheng Zhang, Nicholas Jing Yuan και άλλους [83]. Χρησιμοποιούν ένα πλαίσιο που μοιάζει πολύ με τη δουλειά μας χρη-

σιμοποιώντας τα σιωπηρά σχόλια των χρηστών (επισκέψεις στο εστιατόριο μέσω check-in), ρητά σχόλια (κριτικές εστιατορίων) και άλλα μετά-δεδομένα για να προτείνουν σημεία ενδιαφέροντος για το επόμενο γεύμα.

Επίσης, υπήρξε παρόμοια δουλειά στον τομέα προτάσεων επίγνωσης πλαισίου με τη μορφή Τουριστικής Σύστασης από τους Christopher Laß και Daniel Herzog [84] σχετικά με το πώς να σχεδιαστεί ένα σύστημα συστάσεων με χρήση επίγνωσης πλαισίου για τον τομέα ταξιδιών και τουρισμού. Υπάρχει επίσης έρευνα για το πώς οι κριτικές και η φήμη επηρεάζει τα έσοδα μιας αλυσίδας εστιατορίων από τον Michael Luca [85]. Εξηγεί με μεγάλη λεπτομέρεια τον τρόπο με τον οποίο οι διαδικτυακές κριτικές καταναλωτών βελτιώνουν τις διαθέσιμες πληροφορίες ποιότητα προϊόντος, η οποία μπορεί να είναι καθοριστικός παράγοντας για τους χρήστες που αποφασίζουν πού να φάνε. Επιπλέον, τελευταίο αλλά όχι λιγότερο σημαντικό, υπήρξε επίσης συναρπαστική έρευνα από τους Christoph Trattner, Alexander Oberegger, και άλλους για την κατανόηση του αντίκτυπου του καιρού για τις συστάσεις αντικειμένων ενδιαφέροντος [86]. Η παραπάνω εργασία παρέχει μια μεγάλη εικόνα για το πώς διαφορετικές καιρικές συνθήκες μπορούν να επηρεάσουν την αλληλεπίδραση των χρηστών και την επιλογή τους κατά τη χρήση του συστήματος σύστασης.

Στην επόμενη ενότητα και μετά αρχίζει η παρουσίαση του αλγορίθμου και γενικά ολόκληρου του συστήματος που δημιουργήθηκε σε αυτή την εργασία. Η λύση μας προσπαθεί να λύσει πολλά ζητήματα στο ερευνητικό έργο που αναφέρθηκε προηγουμένως, να βελτιώσει την αξιοπιστία και την εμπειρία των χρηστών, παρέχοντας ισχυρές προτάσεις στον τελικό χρήστη με βάση διάφορους παράγοντες, συμπεριλαμβανομένων των χαρακτηριστικών επίγνωσης πλαισίου.

Κεφάλαιο 4

Περιγραφή θέματος

Σε αυτό το κεφάλαιο γίνεται μια βασική ανάλυση του θέματος καθώς και της βασικής ιδέας που πραγματεύεται η διπλωματική εργασία. Αρχικά αναφέρονται οι λόγοι οι οποίοι οδήγησαν σε αυτή την ιδέα. Πιο συγκεκριμένα αναφέρεται το "γιατί" και το "πώς" αυτή η εργασία θα μπορούσε να επιλύσει και να διευκολύνει καθημερινές καταστάσεις που βρίσκεται μια πληθώρα χρηστών, εκμεταλλευόμενοι την ραγδαία εξέλιξη της τεχνολογίας αλλά και τα σχετικά διαθέσιμα προγραμματιστικά εργαλεία.

4.1 Η κεντρική ιδέα

Για να ορίσουμε την κεντρική ιδέα της παρούσας διπλωματικής αρκεί να δοθεί μια μικρή εξήγηση για τις ανάγκες που καλύπτονται μέσω αυτής. Ένα μεγάλο ποσοστό ανθρώπων στον ελεύθερο τους χρόνο ή ακόμα και στην επαγγελματική τους καθημερινότητα επισκέπτεται διάφορα καταστήματα εστίασης όπως καφετέριες, εστιατόρια και μπαρ. Είναι πολλές οι περιπτώσεις που ενώ κάποιος έχει τη διάθεση και τον χρόνο να βγει δεν γνωρίζει εκ των προτέρων που θέλει να πάει. Υπάρχει πληθώρα επιλογών και αυτό συνήθως δυσκολεύει τον χρήστη να αποφασίσει, είτε γιατί δεν έχει τις κατάλληλες πληροφορίες συγκεντρωμένες είτε γιατί υπάρχουν μέρη σε διάφορες περιοχές που δεν γνώριζε καν. Αυτή ακριβώς την δυσκολία προσπαθεί να επιλύσει η συγκεκριμένη εργασία.

Έχουν συγκεντρωθεί περισσότερα από 2000 καταστήματα στην Αθήνα, παρέχοντας έτσι μεγάλη ποικιλία και ταυτόχρονα ανταπόκριση σε πραγματικά σενάρια. Το σύστημα αξιοποιώντας όλες τις διαθέσιμες πληροφορίες που έχουν συλλεχθεί από τα καταστήματα, σε συνδυασμό με τις πληροφορίες που παρέχει ο χρήστης είναι σε θέση να τον βγάλει από τη δύσκολη κατάσταση της επιλογής. Επίσης οι άνθρωποι γενικά δρουν δυναμικά. Οι προτιμήσεις τους δεν είναι σταθερές αλλά αντιθέτως μεταβάλλονται με τον χρόνο. Συνεπώς δεν είναι αρκετό να γίνονται προτάσεις, αλλά αυτές θα πρέπει να συμβαδίζουν με τις προτιμήσεις των ατόμων που τις χρησιμοποιούν. Εδώ εντάσσεται η μηχανική μάθηση που εξ ορισμού κάνει ακριβώς αυτό το πράγμα. Κάθε φορά που λαμβάνει νέα δεδομένα προσαρμόζεται σε αυτά με αποτελέσματα την παροχή εξατομικευμένων προτάσεων. Εδώ πρέπει να τονισθεί ότι δεν δημιουργείται ένα σύστημα συστάσεων κοινό για όλους, αλλά σε αντίθεση ένα σύστημα που συμπεριφέρεται τελείως διαφορετικά με κάθε χρήστη.

Σε αυτό το σημείο αξίζει να σημειωθεί ότι δεν υπάρχει διαθέσιμο κάτι αντίστοιχο που να αφορά καταστήματα στην Ελλάδα. Ενώ έχει δοθεί μεγάλη έμφαση σε αντίστοιχες εφαρμογές που σχετίζονται με τη διανομή τέτοιους είδους αγαθών, καμία από αυτές δεν παρέχει προτάσεις για επίσκεψη τέτοιων καταστημάτων. Αυτομάτως αυτό δημιουργεί μια μεγάλη πρόκληση καθώς και αρκετό ενδιαφέρον με σκοπό την κάλυψη αυτού του μεγάλου κενού που υπάρχει στην αγορά. Συνεπώς το αρχικό πρόβλημα παραμένει. Ταυτόχρονα πίσω από το συγκεκριμένο πρόβλημα συνήθως κρύβονται και άλλες δυσκολίες που δεν είναι εμφανές με τη πρώτη ματιά. Ένα απλό παράδειγμα που περιγράφει την συγκεκριμένη κατάσταση θα μπορούσε να αφορά ένα άτομο που βρίσκεται σε μια περιοχή για πρώτη φορά (είτε για διασκέδαση, είτε για επαγγελματικούς λόγους). Στην περίπτωση που θέλει να εξερευνήσει αυτό το μέρος και ταυτόχρονα να γεμίσει τον χρόνο του με επισκέψεις σε ενδιαφέρον καταστήματα, αλλά και καταστήματα που ανταποκρίνονται στις δικές του προτιμήσεις, μπορεί απλά με λιγότερο από 2-3 κλικ να πάρει τις συστάσεις που θέλει. Επίσης μπορεί μέσα από την ίδια πλατφόρμα να δει

όλα τα προτεινόμενα καταστήματα στο χάρτη, και να λάβει οδηγίες για τη μετάβαση του σε αυτά. Επιπροσθέτως παρέχονται επιπλέον πληροφορίες που αφορούν τις κριτικές του εκάστοτε καταστήματος, αλλά και φωτογραφίες για να έχουν οι χρήστες μια πρώτη εικόνα για το μέρος που σκοπεύουν να επισκεφθούν.

Φυσικά όλες οι παραπάνω πληροφορίες παρέχονται στο διαδίκτυο. Όμως, όπως έχει ξανά τονισθεί μεγάλη σημασία έχει η ευκολία με την οποία οι χρήστες μπορούν να έχουν πρόσβαση σε αυτές. Είναι τελείως διαφορετικό κάποιος να αναζητά χειροκίνητα, χωρίς ταυτόχρονα να είναι σίγουρος ότι θα βρει αυτό που ψάχνει, συγκριτικά με το να παρέχονται όλες οι πληροφορίες, και να εκτελούνται όλες οι ενέργειες μέσα από το ίδιο περιβάλλον. Εξετάζοντας συγκεκριμένα τον τρόπο που παρέχει η Google πληροφορίες, κάποιος μπορεί πολύ απλά να κάνει αναζήτηση : "Εστιατόρια κοντά μου". Παρόλο που θα επιστραφούν αρκετά αποτελέσματα, το μόνο που λαμβάνεται υπόψιν είναι η τοποθεσία του χρήστη, χωρίς να δίνεται βαρύτητα στις προτιμήσεις που είχε ο χρήστης στο παρελθόν. Συνεπώς οι προτάσεις που προκύπτουν είναι κατά κάποιο τρόπο κοινές για όλους που βρίσκονται σε εκείνη την περιοχή, και δεν εξατομικεύονται στον συγκεκριμένο χρήστη.

Επίσης, είναι πολύ σημαντικό να παρέχεται ευελιξία στον χρήστη. Εμβαθύνοντας σε αυτό η συγκεκριμένη πλατφόρμα θα είναι διαθέσιμη σε όλους. Ενώ παρέχεται η δυνατότητα δημιουργίας λογαριασμού και εγγραφής για πιο εξατομικευμένα αποτελέσματα, όλες σχεδόν οι λειτουργίες παρέχονται και σε μη συνδεδεμένους χρήστες. Έτσι, το μόνο που στην ουσία χρειάζεται είναι η πρόσβαση στο διαδίκτυο, που πλέον είναι κάτι σχεδόν δεδομένο για όλους. Ταυτόχρονα, εκτός από την παροχή υπηρεσιών σε χρήστες, όλο αυτό μπορεί να συμβάλλει ακόμα και στην καλύτερη λειτουργία των ίδιων των επιχειρήσεων. Ενώ υπάρχουν πολύ καλά και προσεγμένα καταστήματα, πολλές φορές δεν οδηγούνται στην επιτυχία επειδή αρκετός κόσμος δεν γνώριζε για την ύπαρξη τους. Το παραπάνω πρόβλημα επιλύεται με την συγκεκριμένη εργασία, αφού στις προτάσεις δεν υπάρχουν μόνο γνωστά και περιζήτητα καταστήματα, αλλά και μικρότερες επιχειρήσεις που δεν είναι τόσο δημοφιλείς στον κόσμο. Συνεπώς, υπάρχει ικανοποίηση και από τις δύο πλευρές (καταναλωτή και επιχείρησης) αφού ένας ευχαριστημένος χρήστης συνεπάγεται με την αντίστοιχη επιτυχία της επιχείρησης.

4.2 LightFM

4.2.1 Γενική περιγραφή

Το LightFM [87] είναι ένα υβριδικό μοντέλο παραγοντοποίησης πινάκα που αντιπροσωπεύει τους χρήστες και τα αντικείμενα ως γραμμικούς συνδυασμούς των λανθάνων παραγόντων των χαρακτηριστικών τους. Το μοντέλο υπερτερεί τόσο των συνεργατικών όσο και των βασισμένα στο περιεχόμενο μοντέλων σε σενάρια ψυχρής εκκίνησης ή αραιών πινάκων αλληλεπίδρασης (χρησιμοποιώντας μεταδεδομένα χρηστών και αντικειμένων) και αποδίδει τουλάχιστον όσο και τα κλασσικά συνεργατικά μοντέλα παραγοντοποίησης μήτρας όπου τα δεδομένα αλληλεπίδρασης είναι άφθονα. Επιπλέον, οι ενσωματώσεις χαρακτηριστικών που παράγονται από το μοντέλο κωδικοποιούν σημασιολογικές πληροφορίες κατά τρόπο που θυμίζει προσεγγίσεις ενσωμάτωσης λέξεων, καθιστώντας τις χρήσιμες για μια σειρά σχετικών εργασιών, όπως προτάσεις ετικετών.

Η δημιουργία συστημάτων σύστασης που έχουν καλή απόδοση σε σενάρια ψυχρής εκκίνησης (όπου υπάρχουν λίγα δεδομένα για νέους χρήστες και αντικείμενα) παραμένει μια πρόκληση. Το πρότυπο MF δεν έχει καλή απόδοση σε αυτήν τη περίπτωση: είναι δύσκολο να εκτιμηθεί αποτελεσματικά οι λανθάνοντες παράγοντες χρηστών και στοιχείων όταν τα δεδομένα συνεργατικής αλληλεπίδρασης είναι αραιά.

Οι μέθοδοι βάσει περιεχομένου το αντιμετωπίζουν αναπαριστώντας στοιχεία μέσω των μεταδεδομένων τους [69]. Καθώς αυτά είναι γνωστά εκ των προτέρων, οι προτάσεις μπορούν να υπολογιστούν ακόμη και για νέα αντικείμενα για τα οποία δεν έχουν συλλεχθεί δεδομένα συνεργασίας. Δυστυχώς, δεν υπάρχει εκμάθηση μεταφοράς σε μοντέλα CB: τα μοντέλα για κάθε χρήστη υπολογίζονται μεμονωμένα και δεν επωφελούνται από δεδομένα άλλων χρηστών. Κατά συνέπεια, τα μοντέλα CB αποδίδουν χειρότερα από τα μοντέλα MF όπου υπάρχουν διαθέσιμες συνεργατικές πληροφορίες και

απαιτούν μεγάλο όγκο δεδομένων για κάθε χρήστη, καθιστώντας τα ακατάλληλα για εκκίνηση ψυχρής εκκίνησης [2].

Στην συγκεκριμένη εργασία, είναι απαραίτητη η επίλυση αυτών των προβλημάτων. Βασικός στόχος της εργασίας είναι να παρέχει στους χρήστες έναν βολικό και ελκυστικό τρόπο περιήγησης και προτάσεων καταστημάτων εστίασης στο διαδίκτυο. Για το σκοπό αυτό διατηρείται ένας πολύ μεγάλος κατάλογο καταστημάτων: τη στιγμή της σύνταξης, συγκεντρώνονται πάνω από 2000 καταστήματα στην Αττική, με στόχο την μελλοντική επέκταση σε όλες τις περιοχές της Ελλάδας.

Τρεις παράγοντες συνωμοτούν για να κάνουν τις συστάσεις προκλητικές στη συγκεκριμένη περίπτωση. Πρώτον, το σύστημά περιέχει πολύ μεγάλο αριθμό αντικειμένων. Αυτό καθιστά τα δεδομένα πολύ αραιά. Δεύτερον, στον χώρο της εστίασης: συχνά, ορισμένα από τα πιο σχετικά αντικείμενα έχουν προστεθεί πρόσφατα, επιτρέποντάς μόνο ένα σύντομο παράθυρο για τη συλλογή δεδομένων και την παροχή αποτελεσματικών συστάσεων. Τέλος, ένα μεγάλο ποσοστό των χρηστών είναι επισκέπτες για πρώτη φορά: θα ήταν ιδανικό να τους παρουσιάζονται συναρπαστικές προτάσεις ακόμη και με λίγα δεδομένα. Αυτός ο συνδυασμός χρηστών και αντικειμένων με ψυχρή εκκίνηση καθιστά τις καθαρές μεθόδους συνεργασίας και βάσει περιεχομένου ακατάλληλες για αυτό το σκοπό.

Για την επίλυση αυτού του προβλήματος, χρησιμοποιήθηκε ένα υβριδικό μοντέλο συνεργασίας περιεχομένου, που ονομάζεται LightFM. Στο LightFM, όπως σε ένα συνεργατικό μοντέλο φιλτραρίσματος, οι χρήστες και τα αντικείμενα αντιπροσωπεύονται ως λανθάνοντα διανύσματα (ενσωματώσεις). Ωστόσο, όπως και σε ένα μοντέλο CB, αυτά ορίζονται εξ ολοκλήρου από συναρτήσεις (σε αυτήν την περίπτωση, γραμμικούς συνδυασμούς) ενσωματώσεων των χαρακτηριστικών περιεχομένου που περιγράφουν κάθε προϊόν ή χρήστη. Για παράδειγμα, εάν η ταινία «Wizard of Oz» περιγράφεται από τα ακόλουθα χαρακτηριστικά: «μουσική φαντασία», «Judy Garland» και «Wizard of Oz», τότε η λανθάνουσα αναπαράστασή της θα δοθεί από το άθροισμα αυτών των χαρακτηριστικών «λανθάνων» παραστάσεων.

Αυτό έχει πολλά οφέλη για συστήματα σύστασης πραγματικού κόσμου. Επειδή το LightFM λειτουργεί καλά τόσο σε πυκνά όσο και σε αραιά δεδομένα, καταργεί την ανάγκη δημιουργίας και συντήρησης πολλαπλών εξειδικευμένων μοντέλων μηχανικής μάθησης για κάθε ρύθμιση. Επιπλέον, καθώς μπορεί να χρησιμοποιήσει μεταδεδομένα χρήστη και στοιχείου, έχει την ιδιότητα να εφαρμόζεται σε σενάρια ψυχρής εκκίνησης τόσο αντικειμένων όσο και χρηστών.

4.2.2 Μοντέλο

Για να περιγραφεί το μοντέλο, ας οριστεί ως U το σύνολο των χρηστών και I το σύνολο των αντικειμένων, F^U το σύνολο των χαρακτηριστικών χρήστη και F^I το σύνολο των χαρακτηριστικών των αντικειμένων. Κάθε χρήστης αλληλεπιδρά με έναν αριθμό αντικειμένων, είτε με ευνοϊκό τρόπο (θετική αλληλεπίδραση), ή με δυσμενή τρόπο (αρνητική αλληλεπίδραση). Το σύνολο όλων των ζευγών αλληλεπίδρασης χρήστη-αντικειμένου $(u, i) \in U \times I$ είναι η ένωση και των θετικών S^+ και αρνητικών αλληλεπιδράσεων S^- .

Οι χρήστες και τα αντικείμενα περιγράφονται πλήρως από τα χαρακτηριστικά τους. Κάθε χρήστης u περιγράφεται από ένα σύνολο χαρακτηριστικών $f_u \subset F^U$. Το ίδιο ισχύει και για κάθε αντικείμενο i του οποίου τα χαρακτηριστικά δίνονται από τη σχέση $f_i \subset F^I$. Τα χαρακτηριστικά είναι γνωστά εκ των προτέρων και αντιπροσωπεύουν μεταδεδομένα χρήστη και αντικειμένου.

Το μοντέλο παραμετροποιείται σε όρους n -διάστατων ενσωματώσεων χαρακτηριστικών χρήστη και αντικειμένων e_f^u και e_f^i για κάθε χαρακτηριστικό f . Κάθε χαρακτηριστικό περιγράφεται επίσης με έναν όρο βαθμιαίας προκατάληψης (b_f^U για τον χρήστη και b_f^I για το αντικείμενο). Η λανθάνουσα αναπαράσταση του χρήστη u δίνεται από το άθροισμα των λανθάνων διανυσμάτων των χαρακτηριστικών του:

$$q_u = \sum_{j \in f_u} e_j^U \quad (4.1)$$

Το ίδιο ισχύει για το αντικείμενο i :

$$p_i = \sum_{j \in f_i} e_j^I \quad (4.2)$$

Ο όρος βαθμιαίας προκατάληψης για τον χρήστη u υπολογίζεται από το άθροισμα των όρων βαθμιαίας προκατάληψης για κάθε χαρακτηριστικό:

$$b_u = \sum_{j \in f_u} b_j^U \quad (4.3)$$

Αντίστοιχα για το αντικείμενο i :

$$b_i = \sum_{j \in f_i} b_j^I \quad (4.4)$$

Η πρόβλεψη του μοντέλου για το χρήστη u και το στοιχείο i δίνεται στη συνέχεια από το εσωτερικό γινόμενο των αναπαραστάσεων χρήστη και αντικειμένου, προσαρμοσμένο από τις προκαταλήψεις χαρακτηριστικών χρήστη και αντικειμένου:

$$\hat{r}_{ui} = f(q_u \cdot p_i + b_u + b_i) \quad (4.5)$$

Υπάρχουν αρκετές συναρτήσεις που θα μπορούσαν να χρησιμοποιηθούν στη θέση της f . Παρέχεται η δυνατότητα επιλογής μεταξύ 4 συναρτήσεων:

- **logistic**: χρήσιμο όταν υπάρχουν τόσο θετικές (1) όσο και αρνητικές (-1) αλληλεπιδράσεις.
- **BPR: Εξατομικευμένη κατάταξη** (Bayesian Personalized Ranking - BPR) [88] απώλεια κατά ζεύγη. Μεγιστοποιεί τη διαφορά πρόβλεψης μεταξύ ενός θετικού παραδείγματος και ενός τυχαία επιλεγμένου αρνητικού παραδείγματος. Χρήσιμο όταν υπάρχουν μόνο θετικές αλληλεπιδράσεις και επιθυμείται βελτιστοποίηση του *Χαρακτηριστικό Λειτουργίας Δέκτη* (Receiver Operating Characteristic - ROC) *Περιοχή Κάτω από την Καμπύλη* (Area Under the Curve - AUC).
- **WARP: Ζυγισμένη κατά προσέγγιση κατάταξη κατά ζεύγη** Weighted Approximate-Rank Pairwise cite12. Μεγιστοποιεί την κατάταξη των θετικών παραδειγμάτων με επανειλημμένη δειγματοληψία αρνητικών παραδειγμάτων έως ότου βρεθεί η κατάταξη παραβίασης ενός. Χρήσιμο όταν υπάρχουν μόνο θετικές αλληλεπιδράσεις και είναι επιθυμητή η βελτιστοποίηση της κορυφής της λίστας προτάσεων (ακριβής @ k).
- **k-OS WARP**: απώλεια στατιστικής παραγγελίας k -th [89]. Μια τροποποίηση του WARP που χρησιμοποιεί το θετικό παράδειγμα k -th για κάθε δεδομένο χρήστη ως βάση για ενημερώσεις κατά ζεύγη.

Σχέση με άλλα μοντέλα

Η σχέση μεταξύ LightFM και του συνεργατικού μοντέλου MF διέπεται από τη δομή των συνόλων χαρακτηριστικών χρήστη και στοιχείων. Εάν τα σύνολα χαρακτηριστικών αποτελούνται αποκλειστικά από μεταβλητές δείκτη για κάθε χρήστη και στοιχείο, το LightFM μειώνεται στο τυπικό μοντέλο MF. Εάν τα σύνολα χαρακτηριστικών περιέχουν επίσης λειτουργίες μεταδεδομένων που μοιράζονται περισσότερο από ένα στοιχεία ή χρήστες, το LightFM επεκτείνει το μοντέλο MF επιτρέποντας στους παράγοντες λανθάνουσας λειτουργίας να εξηγούν μέρος της δομής των αλληλεπιδράσεων χρηστών.

Αυτό είναι σημαντικό για τρεις λόγους.

1. Στις περισσότερες εφαρμογές θα υπάρχουν λιγότερα χαρακτηριστικά μεταδεδομένων από ότι χρήστες ή αντικείμενα, είτε επειδή χρησιμοποιείται μια οντολογία με δομή σταθερού τύπου / κατηγορίας, είτε επειδή ένα λεξικό σταθερού μεγέθους με τους πιο συνηθισμένους όρους διατηρείται όταν χρησιμοποιείτε ακατέργαστα χαρακτηριστικά κειμένου. Αυτό σημαίνει ότι λιγότερες παράμετροι πρέπει να εκτιμηθούν από περιορισμένα δεδομένα εκπαίδευσης, μειώνοντας τον κίνδυνο υπερεκπαίδευσης και βελτιώνοντας την απόδοση της γενίκευσης.

2. Τα λανθάνοντα διανύσματα για μεταβλητές δεικτών δεν μπορούν να εκτιμηθούν για νέους χρήστες ή αντικείμενα που ξεκινούν εν ψυχρώ. Η εκπροσώπησή τους ως συνδυασμοί χαρακτηριστικών μεταδεδομένων που μπορούν να εκτιμηθούν από το εκπαιδευτικό σύνολο καθιστά δυνατή την πραγματοποίηση προβλέψεων ψυχρής εκκίνησης.
3. Εάν υπάρχουν μόνο δυνατότητες ένδειξης, το LightFM θα πρέπει να λειτουργεί στο ίδιο επίπεδο με το τυπικό μοντέλο MF.

Όταν υπάρχουν μόνο χαρακτηριστικά μεταδεδομένων και δεν υπάρχουν μεταβλητές δείκτη, το μοντέλο γενικά δεν μειώνεται σε ένα καθαρό σύστημα βασισμένο σε περιεχόμενο. Οι εκτιμήσεις του LightFM ενσωματώνουν χαρακτηριστικά ενσωματώνοντας τον πίνακα συνεργατικής αλληλεπίδρασης. Αυτό είναι διαφορετικό από τα συστήματα που βασίζονται σε περιεχόμενο τα οποία (όταν χρησιμοποιείται η μείωση των διαστάσεων) συντελούν σε πίνακες συνύπαρξης καθαρού περιεχομένου.

Μία ειδική περίπτωση όπου το LightFM μειώνεται σε ένα καθαρό μοντέλο CB είναι όπου κάθε χρήστης περιγράφεται από μια μεταβλητή δείκτη και έχει αλληλεπιδράσει μόνο με ένα στοιχείο. Σε αυτήν τη ρύθμιση, το διάνυσμα χρήστη είναι ισοδύναμο με ένα διάνυσμα εγγράφων στη διαμόρφωση και μόνο οι δυνατότητες που εμφανίζονται μαζί στις περιγραφές προϊόντων θα έχουν παρόμοιες ενσωματώσεις

Το γεγονός ότι το LightFM περιέχει τόσο το καθαρό μοντέλο CB στο αραιό άκρο δεδομένων του φάσματος όσο και το μοντέλο MF στο πυκνό άκρο υποδηλώνει ότι θα πρέπει να προσαρμόζεται καλά σε σύνολα δεδομένων με διαφορετικό sparsity. Στην πραγματικότητα, τα εμπειρικά αποτελέσματα δείχνουν ότι αποδίδει τουλάχιστον όσο και το κατάλληλο εξειδικευμένο μοντέλο σε κάθε σενάριο.

4.3 Παραγοντοποίηση Πίνακα

Το πρόβλημα συνεργατικού φιλτραρίσματος μπορεί να αναπαρασταθεί με έναν πίνακα βαθμολογιών στοιχείων χρήστη. Λόγω της υψηλής διάστασης των δεδομένων, αυτός ο αρχικός πίνακας συνήθως παραγοντοποιείται σε δύο ή περισσότερους «μικρότερους» πίνακες. Αυτοί οι πίνακες έχουν το πλεονέκτημα των μικρότερων διαστάσεων, με αποτέλεσμα μειωμένο απαιτούμενο χώρο αποθήκευσης και λιγότερο απαιτούμενο χρόνο για την επεξεργασία τους. Επομένως, μπορούν να υποστούν επεξεργασία πιο αποτελεσματικά με αλγόριθμους από τον αρχικό πίνακα. [8]

Μια ειδική περίπτωση μείωσης διαστάσεων είναι η παραγοντοποίηση πίνακα όπου ένας πίνακας δεδομένων αποσυντίθεται στο εσωτερικό γινόμενο δύο πινάκων χαμηλής τάξης. Οι πιο συχνά χρησιμοποιούμενες μέθοδοι αποσύνθεσης μήτρας είναι η *Αποσύνθεση Μοναδικής Αξίας* (Singular Value Decomposition - SVD), η *Παραγοντοποίηση QR*, η *Ανάλυση Κύριων Συνιστωσών* (Factor Analysis - FA) και η *Ανάλυση Κύριων Εξαρτημάτων* (Principal Component Analysis - PCA).

Δεδομένου ενός ελλειπούς πίνακα $R \in \mathbb{R}^{m \times n}$, το MF (στην πιο απλή του εκδοχή) βρίσκει δύο πίνακες $U \in \mathbb{R}^{m \times k}$ και $V \in \mathbb{R}^{n \times k}$ τέτοιους ώστε $r_{i,j} \approx u_i u_j^T$, $\forall i, j \in \Omega$, όπου Ω υποδηλώνει τους δείκτες των υπάρχοντων στοιχείων στο R , $r_{i,j}$ είναι το στοιχείο στη σειρά i και τη στήλη j στο R , $u_i \in \mathbb{R}^k$ είναι η i σειρά του U , $u_j \in \mathbb{R}^k$, είναι η j σειρά του V και $k \leq \text{rank}$, είναι ο προκαθορισμένος αριθμός λανθανουσών χαρακτηριστικών. Εάν ο πίνακας R δεν είναι τετραγωνικός, η κατάταξη βρίσκεται από $R^T R$ ή $R R^T$. Συγκεκριμένα, εάν $m < n$, τότε η κατάταξη προέρχεται από το $R R^T$, καθώς θέλουμε να βρούμε την κατάταξη σειράς. Στην περίπτωση $m > n$, τότε βρίσκουμε την κατάταξη της στήλης από το $R^T R$. Η κύρια ιδέα είναι ότι κάθε πίνακας $m \times n$ R της τάξης $k \ll \min(m, n)$ μπορεί πάντα να εκφράζεται στην ακόλουθη μορφή γινομένου της κατάταξης- παραγόντων k :

$$R \approx UV^T \quad (4.6)$$

ή

$$\begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1k} \\ u_{21} & u_{22} & \dots & u_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mk} \end{bmatrix} \times \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k1} & u_{k2} & \dots & u_{kn} \end{bmatrix}^T \quad (4.7)$$

Προκειμένου να προβλεφθεί πώς ένας χρήστης θα αξιολογούσε ένα στοιχείο, το εσωτερικό γινόμενο ενός διανύσματος χρήστη i και $\text{item } j$ θα έχει ως αποτέλεσμα έναν μόνο αριθμό ως:

$$r_{ij}^{\hat{}} = u_i u_j^T = \sum_{k=1}^K u_{ik} u_{kj} \quad (4.8)$$

Το επόμενο βήμα της μεθόδου είναι να βρεθεί ένας τρόπος υπολογισμού των στοιχείων U και V . Ένας τρόπος για να λυθεί αυτό το πρόβλημα είναι να αρχικοποιηθούν οι δύο πίνακες με κάποιες τυχαίες τιμές και να υπολογιστεί η αριθμητική διαφορά του εσωτερικού γινομένου τους σε σύγκριση με το R . Στη συνέχεια στόχος είναι να ελαχιστοποιηθεί αυτή η διαφορά, επαναλαμβανόμενα με το Gradient Descent.

Αυτή η εργασία επιτυγχάνεται με την επίλυση του ακόλουθου κυρτού προβλήματος:

$$\min J = \frac{1}{2} \left(\sum_{(i,j) \in R} (r_{i,j} - u_i u_j^T)^2 + (\|u_i\|^2 + \|u_j\|^2) \right) \quad (4.9)$$

όπου λ είναι παράμετρος κανονικοποίησης. Να σημειωθεί ότι η διαδικασία επίλυσης για U και V αναφέρεται ως διαδικασία εκπαίδευσης. Για να αξιολογηθεί η ποιότητα της λύσης, μπορούν να αντιμετωπιστούν ορισμένα γνωστά στοιχεία ως άγνωστα στη διαδικασία εκπαίδευσης και να τα συλλεχθούν ως σύνολο δοκιμών. Μόλις βρεθούν U και V , το RMSE, το MAE και πολλά άλλα κριτήρια αξιολόγησης μπορούν να μετρηθούν στο σύνολο δοκιμών για την αξιολόγηση της ακρίβειας της πρόβλεψης.

4.4 Συνδυασμός εργαλείων

Για την κατασκευή του συστήματος συστάσεων αυτής της εργασίας χρησιμοποιήθηκε το LightFM σε συνδυασμό με άλλες τεχνικές για την επέκτασή του. Πιο συγκεκριμένα η παροχή προτάσεων χωρίζεται σε 3 φάσεις, μια εκ των οποίων είναι το LightFM. Χρησιμοποιείται ταυτόχρονα prefiltering και postfiltering. Ανάλογα την περίπτωση, και τι ζητάει ο χρήστης σε πρώτο στάδιο φιλτράρεται η είσοδος προτού δοθεί στο LightFM. Επίσης, αφού παραχθούν τα αποτελέσματα από τον αλγόριθμο του LightFM χρησιμοποιείται ξανά φιλτράρισμα ώστε να γίνουν οι προτάσεις όσο το δυνατόν πιο εξατομικευμένες γίνεται.

Για την παροχή των συστάσεων λαμβάνονται υπόψιν μια πληθώρα παραγόντων και μεταβλητών. Σε αυτά περιλαμβάνονται τα χαρακτηριστικά των χρηστών και αντικειμένων, ενώ επίσης συμπεριλαμβάνονται πληροφορίες επίγνωσης πλαισίου όπως η χρονική στιγμή της ημέρας και η τοποθεσία του χρήστη. Τυπικά το LightFM παρέχει συστάσεις σε χρήστες που έχουν συμπεριληφθεί στο στάδιο της εκπαίδευσης, επειδή όμως το σύστημα αυτό απευθύνεται ακόμα και σε χρήστες που δεν έχουν κάνει εγγραφή, έχει προσαρμοστεί με τρόπο τέτοιο ώστε να είναι δυνατή η παροχή συστάσεων και σε αυτή τη κατηγορία χρηστών.

Επίσης στα συστήματα συστάσεων έχει μεγάλη σημασία το περιβάλλον που χρησιμοποιείται και αλληλεπιδρά ο χρήστης για να δεχτεί προτάσεις. Συνεπώς έχει δοθεί μεγάλη βαρύτητα στην δημιουργία ενός περιβάλλοντος τέτοιου ώστε να είναι φιλικό στο χρήστη και να του παρέχει όσο το δυνατόν περισσότερες λειτουργίες γίνεται ώστε να προσαρμόζεται στις ανάγκες του εκάστοτε χρήστη. Ένα τέτοιο παράδειγμα είναι ο τρόπος με τον οποίο ταξινομούνται τα αποτελέσματα. Αρχικά παρέχεται ένας προκαθορισμένος τρόπος που με την χρήση βαρών (σχετικά με απόσταση, βαθμολογία, προηγούμενες αλληλεπιδράσεις του χρήστη) παρέχει τις προτάσεις. Στη συνέχεια όμως παρέχεται η δυνατότητα αυτή η ταξινόμηση να αλλάξει και να προσαρμοστεί σε αυτό που ζητάει ο χρήστης τη συγκεκριμένη χρονική στιγμή. Έτσι τα αποτελέσματα μπορούν να ταξινομηθούν είτε με τον προκαθορισμένο τρόπο, είτε με βάση την βαθμολογία, τις κριτικές και την απόσταση. Επιπροσθέτως υπάρχει η δυνατότητα φιλτράρισματος των αποτελεσμάτων με βάση το είδος του εκάστοτε καταστήματος (café, bar, εστιατόρια και night club).

Στα περισσότερα συστήματα συστάσεων συνήθως δεν δικαιολογείται το πως προέκυψαν οι συστάσεις και αυτό έχει ως αποτέλεσμα ο χρήστης να μη γνωρίζει το "γιατί" το σύστημα του πρότεινε τα συγκεκριμένα καταστήματα. Για το λόγο αυτό μια ακόμη επέκταση που έχει προστεθεί είναι η αιτιολόγηση των συστάσεων. Δίπλα από κάθε πρόταση παρέχεται μια αιτιολόγηση αναφορικά με την στιγμή της ημέρας και με βάση τα καταστήματα που έχει επισκεφθεί ο χρήστης στο παρελθόν. Η δεύτερη δυνατότητα παρέχεται μόνο σε εγγεγραμμένους χρήστες, αφού σε άλλη περίπτωση δεν υπάρχουν διαθέσιμα τέτοια δεδομένα ώστε να αιτιολογηθεί η σύσταση.

Επίσης για ακόμη μεγαλύτερη διευκόλυνση των χρηστών, έχει προστεθεί η δυνατότητα λήψης οδηγιών μέσα από το περιβάλλον εκμεταλλευόμενοι την τοποθεσία του χρήστη. Με αυτόν τον τρόπο οι χρήστες μπορούν να έχουν μια γενική εικόνα για την τοποθεσία όλων των καταστημάτων που τους προτείνονται βλέποντας τα στον χάρτη. Επίσης με βάση το μεταφορικό μέσο που έχει προταθεί (αυτοκίνητο ή πόδια) μπορούν να δουν την προτεινόμενη διαδρομή στον χάρτη. Για την πρόταση του μεταφορικού μέσου έχει οριστεί μια συγκεκριμένη ακτίνα γύρω από τον χρήστη ώστε όσα σημεία ενδιαφέροντος βρίσκονται κοντά του να του προτείνεται η μετάβαση με τα πόδια, ενώ όσα θα τον ενδιέφεραν αλλά δεν είναι τόσο κοντά του, προτείνεται η μετάβαση με αυτοκίνητο.

Επίσης για να είναι το σύστημα σε θέση να βελτιώνεται και να προσαρμόζεται στον κάθε χρήστη ξεχωριστά έχει προστεθεί η δυνατότητα αξιολόγησης για εγγεγραμμένους χρήστες. Κάθε χρήστης μπορεί να βαθμολογήσει όποιο κατάστημα επιθυμεί, παρέχοντας μια βαθμολογία μεταξύ 1 και 5. Με αυτόν τον τρόπο το σύστημα ανανεώνεται και επανεκπαιδεύεται ανά τακτά χρονικά διαστήματα ανάλογα με το πλήθος των καινούριων κριτικών. Αυτό βοηθάει με 2 τρόπους. Από την μια τα αποτελέσματα γίνονται όλο και πιο εξατομικευμένα, ενώ ταυτόχρονα εντοπίζονται χρήστες με παρόμοια ενδιαφέροντα και έτσι γίνεται καλύτερη εκμετάλλευση τη CF τεχνικής.

Όλα τα παραπάνω εξηγούνται με πολύ περισσότερες λεπτομέρειες στο επόμενο κεφάλαιο όπου αναλύονται τα 3 στάδια που αναφέρθηκαν παραπάνω, ενώ επίσης εξηγούνται τα δεδομένα που αποκτήθηκαν για τη δημιουργία αυτού του συστήματος. Χρησιμοποιήθηκε αρκετή προ-επεξεργασία για την εξαγωγή μοτίβων και την καλύτερη κατανόηση των δεδομένων. Επίσης η συγκεκριμένη εργασία έχει δημιουργηθεί με τέτοιο τρόπο ώστε να μπορεί να καλύψει αρκετές περιπτώσεις και να προσαρμόζεται το σύστημα δυναμικά. Πιο συγκεκριμένα το σύστημα που έχει δημιουργηθεί προσφέρει κάποιες δυνατότητες διαθέσιμες για όλους τους χρήστες, ενώ επεκτείνεται με κάποια ακόμα χαρακτηριστικά που αφορούν μόνο τους συνδεδεμένους χρήστες.

Κεφάλαιο 5

Υλοποίηση

Η αρχιτεκτονική της υλοποίησής εν συντομία σε στυλ διαγράμματος παρουσιάζεται στο Σχήμα 5.1. Όπως φαίνεται και στο σχήμα πριν δοθούν τα δεδομένα στο LightFM που έχει τον ρόλο του υβριδικού μοντέλου, γίνεται ένα φιλτράρισμα με βάση τις προτιμήσεις του χρήστη όπως τις έχει δηλώσει κατά την εγγραφή του. Στην συνέχεια στα αποτελέσματα που δίνει το LightFM, χρησιμοποιούνται οι μεταβλητές επίγνωσης πλαισίου ώστε να προσαρμοστούν τα αποτελέσματα στο περιβάλλον, και να προκύψει ένα νέο σκορ προτιμήσεις μέσα από βάρη που έχουν οριστεί για αυτά τα χαρακτηριστικά. Στην συνέχεια λαμβάνοντας υπόψιν όλα τα παραπάνω γίνεται ένα τελικό φιλτράρισμα και ταξινόμηση ώστε να δοθούν τα τελικά αποτελέσματα στο χρήστη.

5.1 Το Σύνολο Δεδομένων

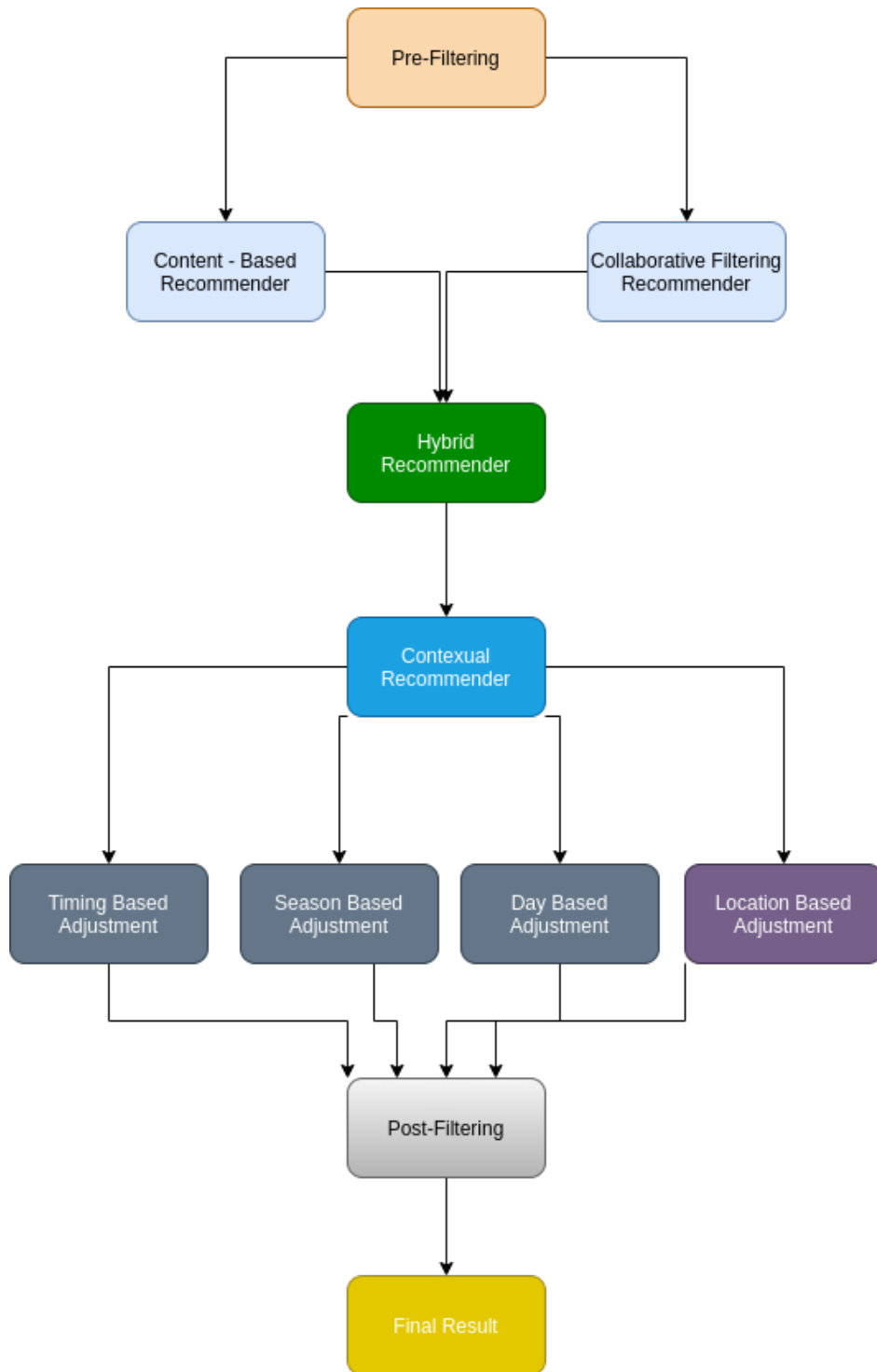
Όλα τα δεδομένα που έχουν χρησιμοποιηθεί σε αυτή την εργασία προέρχονται από το Google maps. Ορισμένα από αυτά συλλέχθηκαν μέσω του API που παρέχει η Google, ενώ τα υπόλοιπα επειδή δεν ήταν διαθέσιμα με κάποιο άλλο τρόπο, ο μόνος τρόπος συλλογής τους ήταν μέθοδοι crawling.

5.1.1 Google API

Αρχικά σε πρώτο στάδιο έπρεπε να οριστούν οι περιοχές ενδιαφέροντος. Για να πραγματοποιηθεί μια αναζήτηση δεν μπορεί να είναι γενική και αόριστη, αλλά πρέπει να έχει οριστεί ένα κέντρο και μια ακτίνα. Πιο συγκεκριμένα χρησιμοποιήθηκε το Nearby Search API. Για να χρησιμοποιηθεί η συγκεκριμένη υπηρεσία προϋποθέτει ότι διατίθεται API key μέσω εγγραφής στην Google platform.

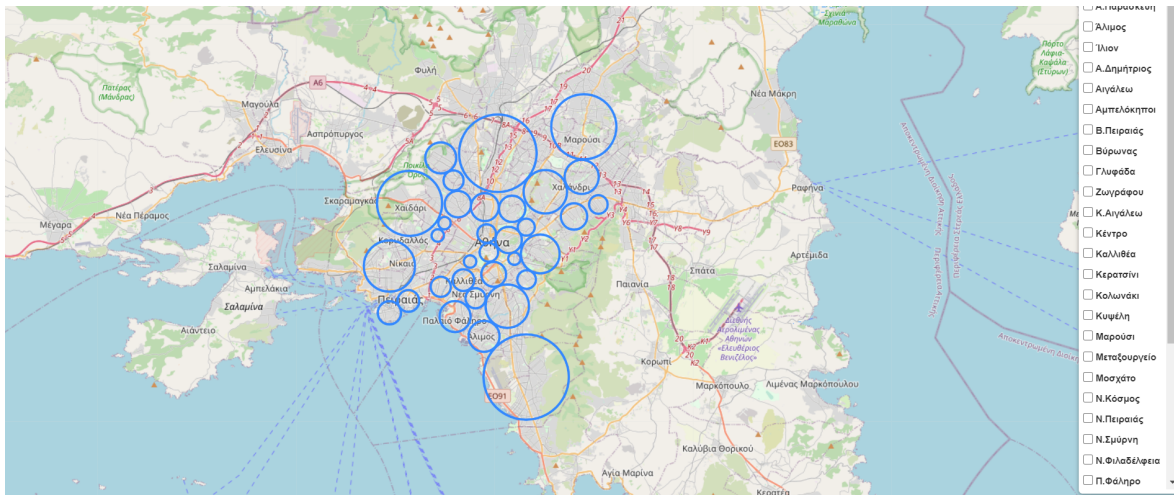
Το γεωγραφικό μήκος και πλάτος μαζί με την ακτίνα αναζήτησης αποτελούν τα απαραίτητα χαρακτηριστικά που χρειάζονται για να πραγματοποιηθεί η αναζήτηση. Λόγω ότι το συγκεκριμένο API επιστρέφει max 60 αποτελέσματα, για να μεγιστοποιηθούν τα διαθέσιμα καταστήματα, αλλά και για λόγους πρακτικότητας οι αναζητήσεις πραγματοποιήθηκαν με βάση την περιοχή και τον τύπο των καταστημάτων. Συνεπώς για κάθε μια από τις διαθέσιμες περιοχές πραγματοποιήθηκε ξεχωριστά αναζήτηση για κάθε μια από τις διαθέσιμες κατηγορίες. Όπως φαίνεται και στο Σχήμα 5.2 ο τρόπος που έχει χωριστεί η Αθήνα αποτελείται από 33 διαφορετικές περιοχές. Κάθε μια από αυτές ανάλογα την έκταση της και την τοποθεσία της έχει διαφορετική ακτίνα. Για μεγάλες περιοχές όπως τον Πειραιά θεωρήθηκε προτιμότερο ο χωρισμός σε ακόμη μικρότερα κομμάτια ώστε να αντληθούν όσο το δυνατόν καταστήματα γίνεται. Πιο συγκεκριμένα αναλυτικά οι περιοχές είναι:

- Αγία Παρασκευή
- Βόρειος Πειραιάς
- Καλλιθέα
- Άγιος Δημήτριος
- Βύρωνας
- Κερατσίνι
- Αλιμος
- Γλυφάδα
- Κολωνάκι
- Ίλιον
- Ζωγράφου
- Κυψέλη
- Αιγάλεω
- Αιγάλεω κέντρο
- Μαρούσι
- Αμπελόκηποι
- Κέντρο
- Μεταξουργείο



Σχήμα 5.1: Αρχιτεκτονική υλοποίησης

- Μοσχάτος
- Νότιος Πειραιάς
- Νέος Κόσμος
- Νέα Σμύρνη
- Νέα Φιλαδέλφεια
- Παλαιό Φάληρο
- Παγκράτι
- Περιστέρι
- Πετράλωνα
- Πετρούπολη
- Σεπόλια
- Χαλάνδρι
- Χαϊδάρι
- Χολαργός
- Ψυχικό



Σχήμα 5.2: Χωρισμός Αττικής σε περιοχές

Αφού πραγματοποιήθηκε ο χωρισμός σε περιοχές, για κάθε περιοχή επιστράφηκαν οι ακόλουθοι τύποι δεδομένων. Πιο αναλυτικά κάθε json που επιστράφηκε από τα ερωτήματα στο API περιλάμβανε:

- **Business Status:** Δηλώνει την τωρινή κατάσταση της επιχείρησης, με τις πιθανές τιμές να είναι είτε λειτουργική είτε κλειστή προσωρινά. Δεδομένου των συνθηκών της πανδημίας ήταν αναμενόμενο ότι αρκετές επιχειρήσεις εστίασης θα παρουσίαζαν την δεύτερη κατάσταση. Επειδή αυτή η κατάσταση όμως κάποια στιγμή θα τελειώσει θα ήταν λάθος και άσκοπο να αφαιρεθούν αυτές οι επιχειρήσεις σκεπτόμενοι μελλοντικά.
- **Εικόνα:** Για κάθε επιχείρηση υπάρχουν διαθέσιμες φωτογραφίες που μπορούν να χρησιμοποιηθούν ώστε να δώσουν μια καλύτερη αίσθηση στο χρήστη για το τι πρόκειται να επισκεφθεί. Παρόλο αυτά, αυτό θα πρέπει να γίνει σε δεύτερο στάδιο για κάθε επιχείρηση ξεχωριστά, διότι το link που δίνεται χρησιμοποιείται για να ληφθεί η εικόνα από άλλο API χωρίς να την καθιστά διαθέσιμη από το nearby search.
- **Διαστάσεις εικόνας:** Παρέχονται επίσης οι διαστάσεις των φωτογραφιών ώστε να είναι ευκολότερος ο σχεδιασμός της πλατφόρμας παρέχοντας ομοιομορφία μεταξύ των επιχειρήσεων.
- **Place id:** Κάθε επιχείρηση έχει ένα μοναδικό id το οποίο είναι πολύ χρήσιμο και στην εκπαίδευση του μοντέλου του LightFM καθώς και για τη χρήση του ως index στη βάση που χρησιμοποιείται για την αποθήκευση όλων των δεδομένων.
- **Όνομα:** Για κάθε επιχείρηση παρέχεται ένα όνομα με το οποίο αυτή είναι γνωστή.
- **Εύρος Τιμών:** Για αρκετές επιχειρήσεις παρέχεται πληροφορία σχετικά με το επίπεδο των τιμών. Πιο συγκεκριμένα χωρίζονται σε φθηνές, μέτριες, ακριβές και πολύ ακριβές. Το συγκεκριμένο μπορεί να βοηθήσει αρκετά τον χρήστη στο φιλτράρισμα των αποτελεσμάτων που επιθυμεί να βλέπει.
- **Βαθμολογία:** Η μέση βαθμολογία της επιχείρησης που προκύπτει από το μέσο όρο όλων των βαθμολογιών που έχει λάβει από τους χρήστες.
- **Τύπος επιχείρησης:** Σε κάθε επιχείρηση αντιστοιχεί τουλάχιστον ένας τύπος. Αρκετές από αυτές έχουν περισσότερους από έναν χαρακτηρισμούς. Συνολικά χωρίζονται σε: cafe, bar, restaurant και night club. Το συγκεκριμένο χαρακτηριστικό χρησιμοποιήθηκε για να φιλτραριστούν τα διαθέσιμα καταστήματα, αφού πολλά από αυτά που είχαν επιστραφεί δεν ανήκαν στις κατηγορίες που στοχεύει αυτή η εργασία. Επίσης και αυτό το χαρακτηριστικό συμβάλλει σε μεγάλο

βαθμό στις συστάσεις και από πλευρά κατηγοριοποίησης αλλά και φιλτραρίσματος αποτελεσμάτων.

- Πλήθος κριτικών: Ο συνολικός αριθμός κριτικών που έχει λάβει κάθε επιχείρηση. Με βάση αυτή τη μεταβλητή οι επιχειρήσεις μπορούν να χωριστούν με βάση τη δημοφιλία τους, αφού επιχειρήσεις με περισσότερες κριτικές συνήθως είναι δημοφιλέστερες.
- Διεύθυνση καταστήματος: Πληροφορίες για την τοποθεσία της επιχείρησης
- Γεωγραφικό μήκος και πλάτος: Στοιχεία για την ακριβή τοποθεσία στο χάρτη και χρήσιμα για τη δημιουργία οδηγιών και υπολογισμού αποστάσεων.

Για την προσαρμογή των συστάσεων στο χρόνο, και με σκοπό να μη προτείνονται καταστήματα που είναι κλειστά τη χρονική στιγμή που ο χρήστης ζητάει προτάσεις, χρησιμοποιήθηκε επίσης το place details API το οποίο παρέχει περισσότερες λεπτομέρειες σχετικά με την εκάστοτε επιχείρηση, οι οποίες δεν συμπεριλαμβάνονται στο nearby search. Έτσι αντλήθηκαν πληροφορίες για κάθε επιχείρηση ξεχωριστά που αφορούν τις ώρες που είναι ανοιχτές για κάθε ημέρα της εβδομάδας. Με αυτό τον τρόπο οι χρήστες έχουν τη δυνατότητα να γνωρίζουν αν το κατάστημα είναι ανοιχτό εκείνη τη στιγμή, και αν όχι να ενημερώνονται για το πότε θα ανοίξει.

5.1.2 Crowling

Παρόλο που όλα τα παραπάνω δεδομένα είναι αρκετά σημαντικά και συμβάλλουν στη δημιουργία ενός καλού συστήματος συστάσεων, λείπει κάτι βασικό το οποίο είναι οι αλληλεπιδράσεις των χρηστών. Δεν γίνεται να εκπαιδευτεί συνεργατικό μοντέλο (CF) χωρίς την ύπαρξη αξιολογήσεων. Η Google δεν παρέχει κάποιο API για την λήψη τέτοιου είδους πληροφορίας. Συνεπώς ο μόνος τρόπος για την λήψη αυτών των δεδομένων είναι το crawling. Για το σκοπό αυτό δημιουργήθηκε ένα αυτοματοποιημένο script, όπου για το σύνολο των επιχειρήσεων διάβαζε μια-μια της κριτικές και αποθήκευε το id του χρήστη καθώς και την κριτική του μεταξύ 1 και 5. Η συγκεκριμένη εργασία δεν ασχολήθηκε με το κείμενο που άφησε ο κάθε χρήστης διότι αυτό βρισκόταν εκτός των πλαισίων της εργασίας. Οι κριτικές χωρίστηκαν σε θετικές και αρνητικές αναλόγως με την βαθμολογία που άφησαν οι χρήστες. Εδώ επίσης πρέπει να σημειωθεί ότι λόγω του μεγάλου όγκου χρηστών συγκριτικά με τις επιχειρήσεις, ο πίνακας αλληλεπίδρασης ήταν αρκετά αραιός. Σε αυτό βοήθησε αρκετά ο τρόπος που διαχειρίζεται αυτούς τους πίνακες το lightfm χωρίς την προσθήκη αισθητών καθυστερήσεων.

Στη συνέχεια οι κριτικές αυτές ομαδοποιήθηκαν σε επίπεδο επιχείρησης αλλά και χρήστη. Με βάση τα ids των επιχειρήσεων και των χρηστών ήταν εύκολο να συγκεντρωθούν όλες μαζί για την διεξαγωγή στατιστικών και συμπερασμάτων. Αυτό βοήθησε αρκετά, διότι πριν τη συλλογή αυτών των δεδομένων υπήρχε απλά ένας μέσος όρος για κάθε επιχείρηση, χωρίς να υπάρχει πληροφορία για το πως κατανέμονται οι βαθμολογίες. Συνολικά για 2222 επιχειρήσεις συλλέχθηκαν περισσότερες από 1 εκατομμύριο κριτικές(1048366), με περισσότερους από 300 χιλιάδες χρήστες(358822).

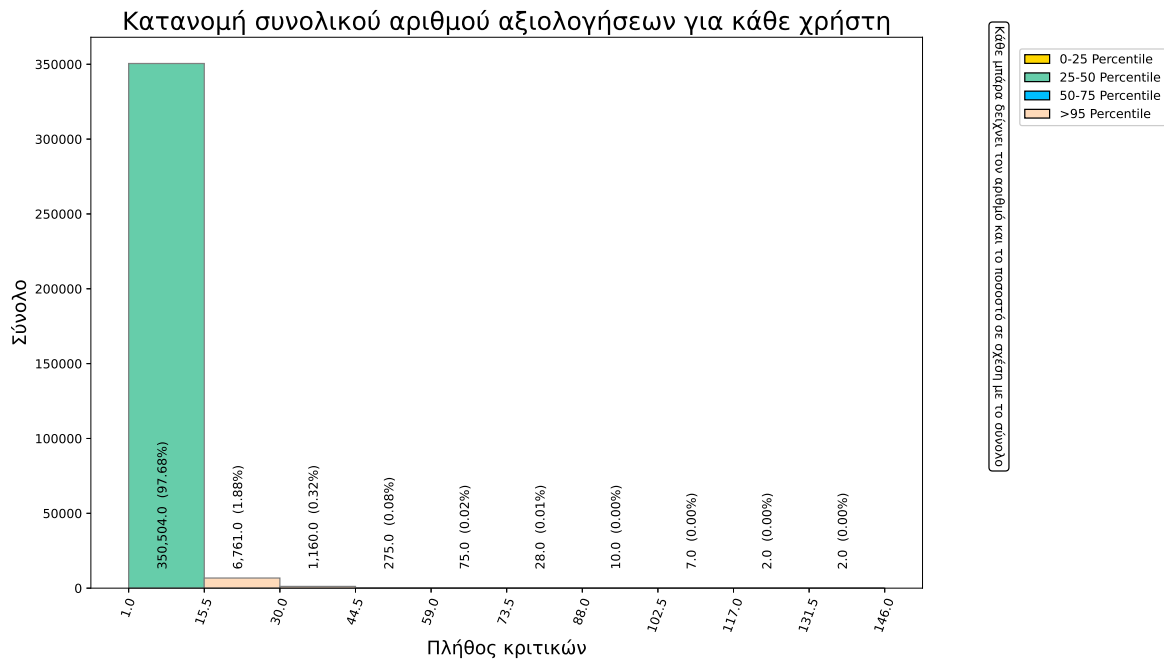
5.1.3 Επεξήγηση Δεδομένων

Για την καλύτερη κατανόηση των δεδομένων παρέχονται ορισμένα σχήματα και κάποιες στατιστικές πληροφορίες. Με αυτό το τρόπο μπορούν να αιτιολογηθούν καλύτερα τα αποτελέσματα που προκύπτουν, αφού ο πυλώνας του συστήματος είναι τα δεδομένα που το απαρτίζουν. Οι 2 βασικές κατηγορίες δεδομένων είναι τα δεδομένα σχετικά με τις επιχειρήσεις και τα δεδομένα σχετικά με τους χρήστες.

Χρήστες

Οι χρήστες είναι ένας πολύ σημαντικός παράγοντας στα συστήματα συστάσεων αφού αυτοί ορίζουν τις αλληλεπιδράσεις. Με βάση τις αλληλεπιδράσεις προκύπτουν πολύ σημαντικά συμπεράσματα όπως παρόμοιοι χρήστες, παρόμοιες επιχειρήσεις κλπ. Αρχικά για να γίνει κατανοητό πόσο αραιός

είναι αυτός ο πίνακας, στο Σχήμα 5.3 παρουσιάζεται ένα ιστόγραμμα σχετικά με το πλήθος κριτικών που κάνουν οι χρήστες. Όπως φαίνεται και στο ιστόγραμμα η συντριπτική πλειοψηφία των χρηστών



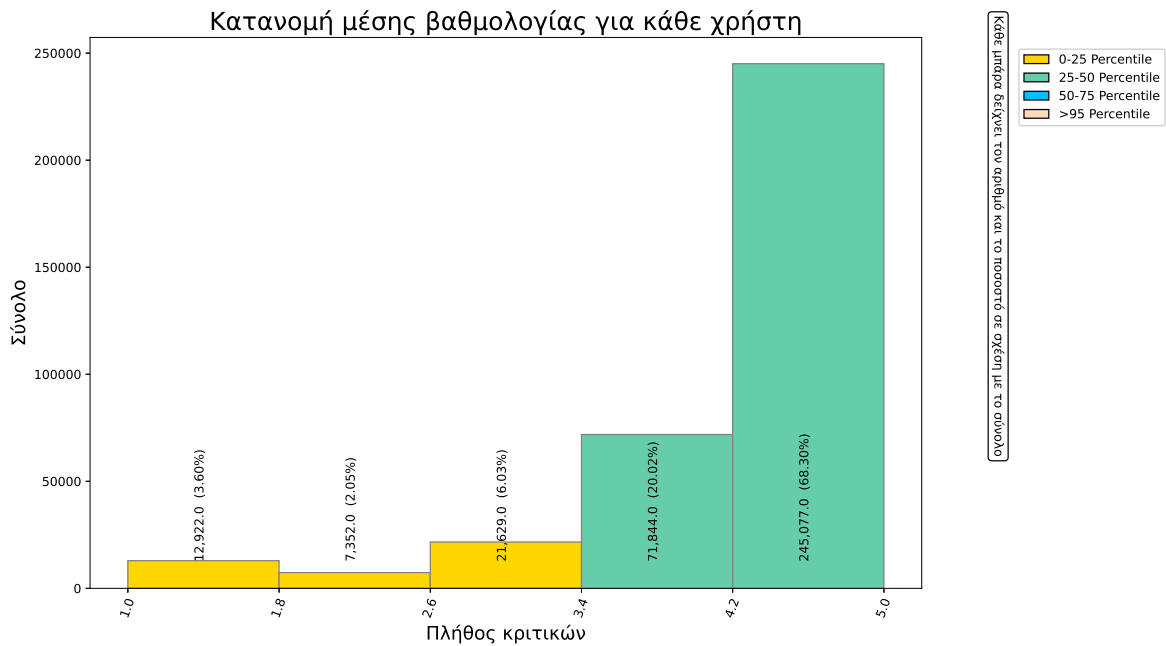
Σχήμα 5.3: Ιστόγραμμα πλήθους κριτικών χρηστών

(97.86%) έχουν μόλις 1 έως 15 κριτικές. Για να γίνει κατανοητό πόσο μικρό είναι αυτό το νούμερο, στην καλύτερη περίπτωση που οι κριτικές είναι 15, αφορούν μόνο το 0.67% των συνολικών επιχειρήσεων. Κατά τη συλλογή των δεδομένων αυτό ήταν αναμενόμενο, αφού όλοι οι χρήστες που υπάρχουν σε αυτή την εργασία πιθανώς να έχουν κάνει κριτικές και σε άλλα πράγματα στο google, όπως άλλα καταστήματα, άλλες περιοχές κλπ. Επειδή αυτά τα είδη καταστημάτων δεν συμπεριλαμβάνονταν στα είδη που αφορούν αυτή την εργασία, τέτοιες κριτικές έμειναν απέξω. Παρά την μεγάλη αραιότητα που παρατηρείται αυτό δεν στάθηκε εμπόδιο στη λειτουργία του συστήματος, αφού πέρα από τις αλληλεπιδράσεις χρησιμοποιούνται και άλλες χαρακτηριστικά όπως τα είδη των καταστημάτων και αρκετές άλλες πληροφορίες επίγνωσης πλαισίου.

Σχετικά με τις βαθμολογίες των χρηστών, όπως φαίνεται και στο Σχήμα 5.4, οι περισσότεροι που μπαίνουν στη διαδικασία να αφήσουν κάποια κριτική, συνήθως αυτή είναι θετική. Πιο συγκεκριμένα το 68.3% βαθμολογεί κατά μέσο όρο μεταξύ 4.2 και 5. Ακολουθεί το 20% με βαθμολογίες μεταξύ 3.4 και 4.2. Οι υπόλοιποι χρήστες (περίπου 10%) τείνουν να βαθμολογούν χαμηλότερα, και συνήθως έχουν μόλις 1 ή 2 κριτικές. Συνεπώς υπάρχουν 2 κατηγορίες χρηστών, η πλειοψηφία που επιβραβεύει τις επιχειρήσεις με τη δημιουργία θετικών αξιολογήσεων, και ένα άλλο μικρό ποσοστό που μπαίνει στη διαδικασία να βαθμολογήσει μόνο στην περίπτωση που έχουν μείνει δυσαρεστημένοι.

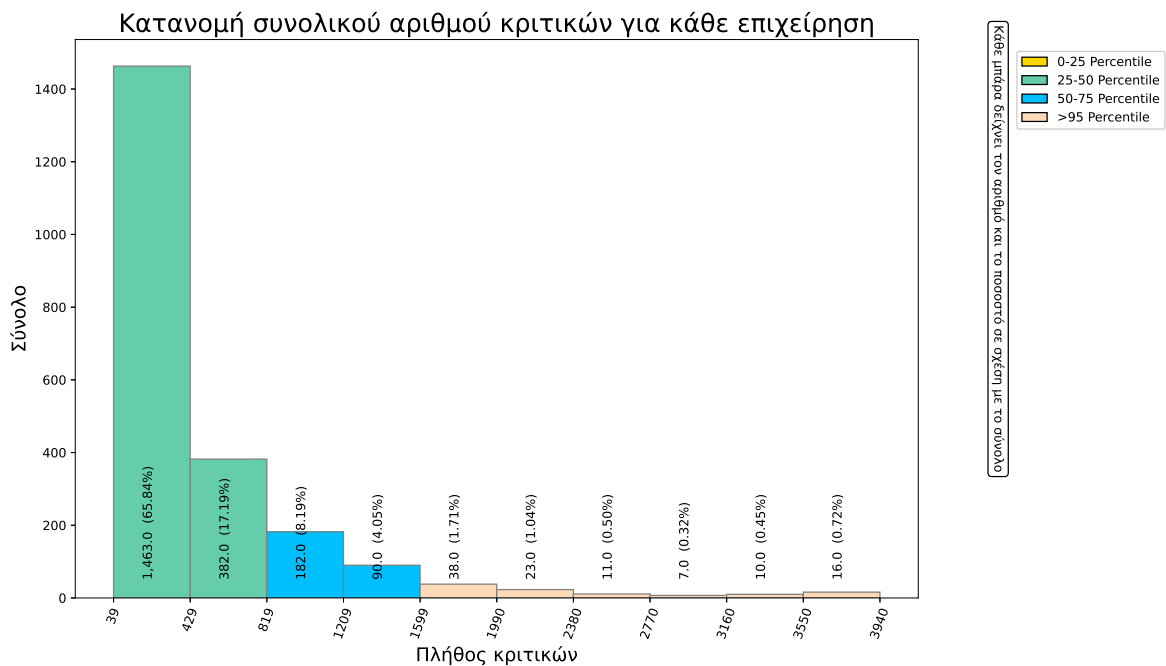
Επιχειρήσεις

Σχετικά με τις επιχειρήσεις, σημαντικοί παράγοντες που πρέπει να ληφθούν υπόψιν με βάση τα διαθέσιμα δεδομένα είναι οι τύποι των επιχειρήσεων, η κατηγορία τιμών τους, και το πλήθος των κριτικών τους που ορίζει πόσο γνωστές είναι στο κοινό. Ξεκινώντας με το συνολικό αριθμό κριτικών για κάθε επιχείρηση, όπως φαίνεται στο ιστόγραμμα 5.5, η πλειοψηφία των καταστημάτων έχει μεταξύ 39 και 819 κριτικές. Πιο συγκεκριμένα το μεγαλύτερο ποσοστό καταλαμβάνει το εύρος 39-429 με ποσοστό 65.84% και ακολουθεί με 17.19% το εύρος 429-819. Με βάση αυτά τα νούμερα προκύπτει το συμπέρασμα ότι οι περισσότερες επιχειρήσεις κυμαίνονται στα ίδια νούμερα και οι χρήστες έχουν την ίδια αντιμετώπιση σε αυτές. Φυσικά υπάρχουν και περιπτώσεις που ορισμένες από αυτές απέχουν αρκετά από αυτές τις τιμές, κάνοντας τες να είναι πολύ δημοφιλέστερες στο κοινό. Αυτό δεν



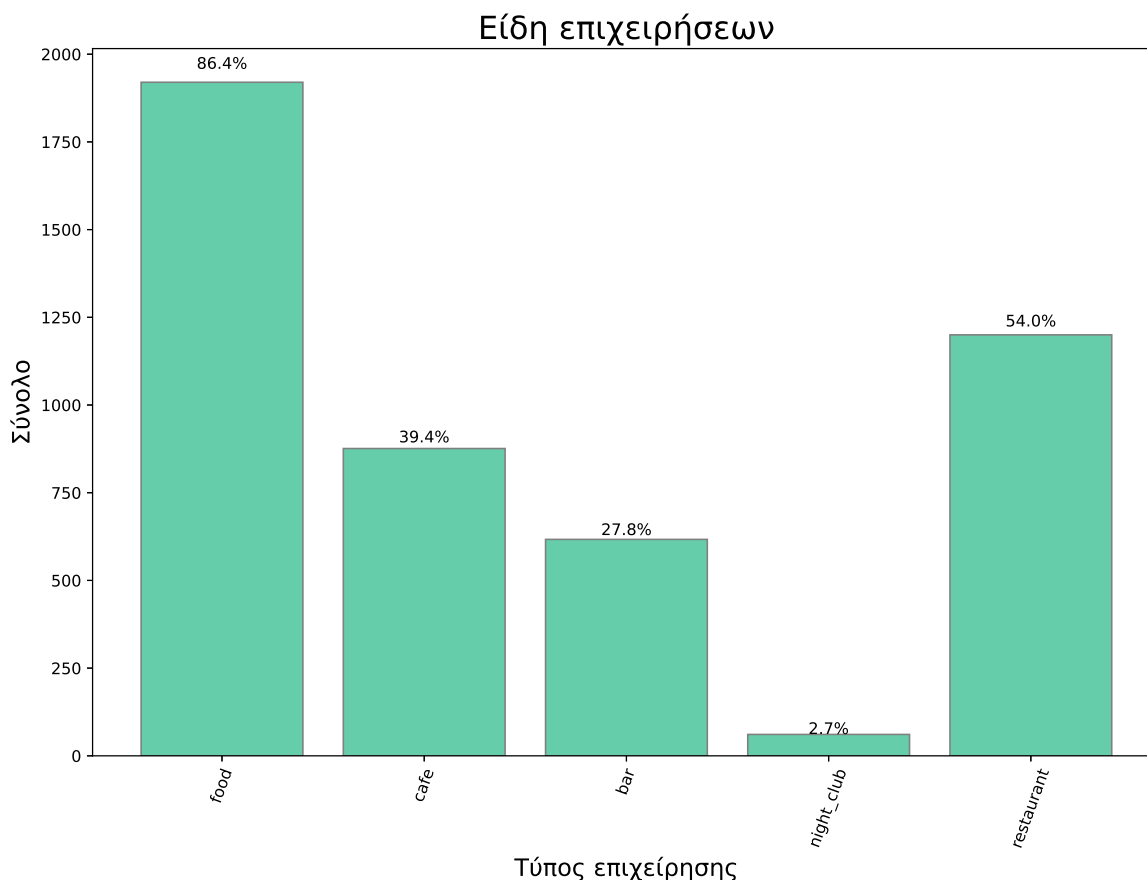
Σχήμα 5.4: Ιστόγραμμα διακύμανσης κριτικών χρηστών

επιρεάζει αναγκαστικά τις συστάσεις που προτείνονται αφού παρέχεται δυνατότητα ο χρήστης να παρουσιάζει τις συστάσεις που δέχεται όπως εκείνος επιθυμεί. Αυτό έχει ως αποτέλεσμα να μην παρέχονται μόνο καταστήματα που είναι ήδη γνωστά, αλλά να δίνονται και ευκαιρίες στις περιπτώσεις που κάποιες επιχειρήσεις δεν έχουν καταφέρει ακόμα να γίνουν τόσο γνωστές στον κόσμο. Φυσικά, δίνεται βάρος και στο πλήθος το κριτικών, χωρίς όμως αυτός να είναι ο μοναδικός παράγοντας που λαμβάνεται υπόψη για τη δημιουργία των προτάσεων. Τέλος παρατηρείται ότι για πλήθος κριτικών μεγαλύτερο από 2000, τα ποσοστά πέφτουν δραματικά και τέτοιου είδους επιχειρήσεις θα μπορούσαν να θεωρηθούν εξαίρεση, χωρίς να ανήκουν στο γενικό κανόνα που περιλαμβάνει την πλειοψηφία.



Σχήμα 5.5: Πλήθος κριτικών ανά επιχείρηση

Σχετικά με τα είδη των επιχειρήσεων υπάρχει μια ομοιομορφία, με την έννοια ότι στο σύνολο τους καλύπτουν όλες τις κατηγορίες σε ικανοποιητικό βαθμό. Πιο συγκεκριμένα όπως φαίνεται και στο σχήμα 5.6 το μεγαλύτερο ποσοστό (86.41%) περιλαμβάνει υπηρεσίες φαγητού. Εδώ πρέπει να

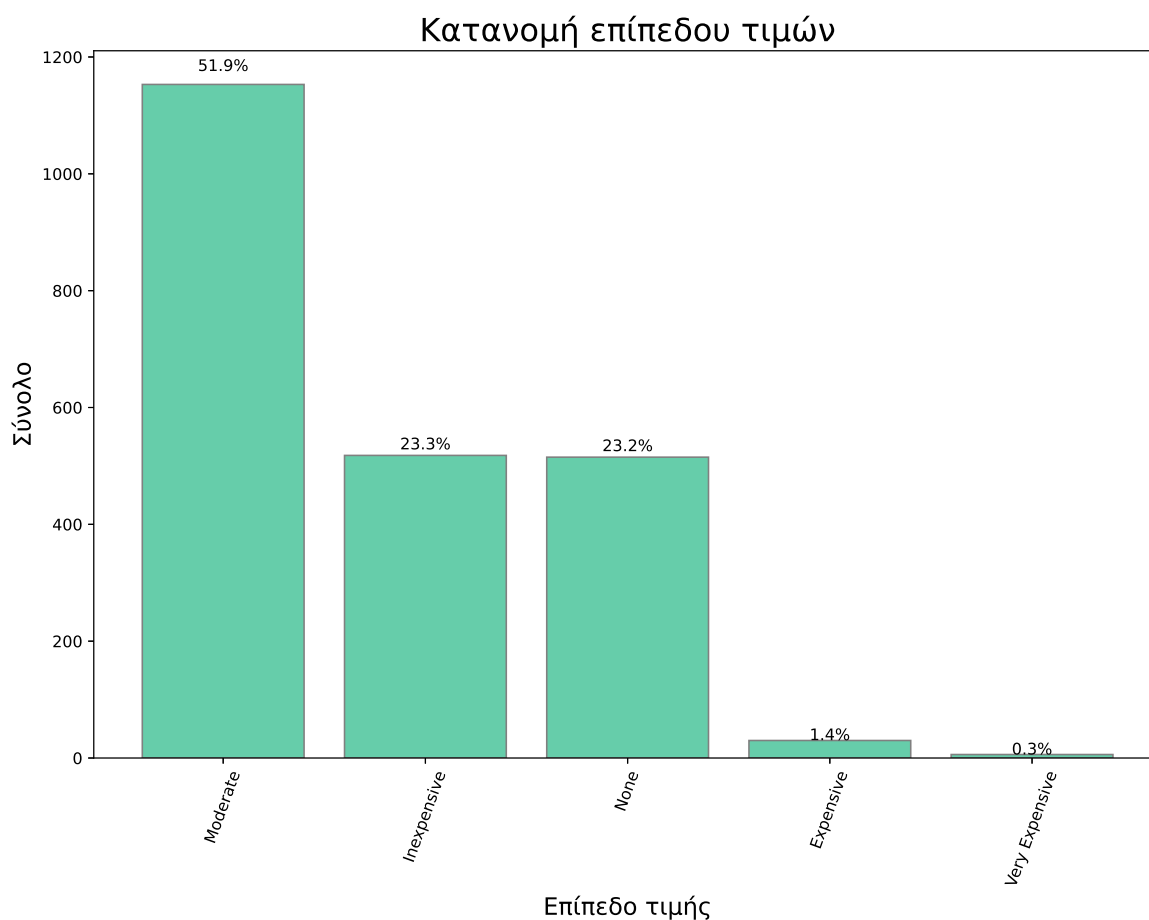


Σχήμα 5.6: Κατηγορίες επιχειρήσεων

τονισθεί ότι μια επιχείρηση μπορεί να περιλαμβάνει περισσότερες από μια κατηγορίες. Παράδειγμα αρκετά καφέ, περιέχουν την κατηγορία φαγητό, ενώ όλα τα εστιατόρια το περιέχουν επίσης. Η μόνη κατηγορία που φαίνεται να υπάρχει λιγότερο σε σχέση με τις υπόλοιπες είναι τα night clubs. Αυτό αιτιολογείται εύκολα αν σκεφτεί κανείς ότι όλες οι υπόλοιπες κατηγορίες συνδυάζονται μεταξύ τους. Για παράδειγμα είναι πολύ συχνό μια καφετέρια να μετατρέπεται το βράδυ σε μπαρ, ενώ αντίστοιχα πιο σπάνιο ένα night club τα πρωινά να λειτουργεί ως καφετέρια. Έτσι μετά το φαγητό ακολουθούν τα εστιατόρια με ποσοστό 54% ενώ ακολουθούν οι καφετέριες και τα μπαρ με ποσοστά 39.4% και 27.77% αντίστοιχα. Με τις παραπάνω παρατηρήσεις γίνεται αντιληπτό ότι το συγκεκριμένο σύστημα συστάσεων είναι σε θέση να καλύψει όλα τα είδη χρηστών με διαφορετικές προτιμήσεις αφού περιέχει ποικιλία καταστημάτων από κάθε κατηγορία.

Επίσης το συγκεκριμένο χαρακτηριστικό είναι κομβικής σημασίας αφού συμβάλλει αρκετά στην κατηγοριοποίηση των επιχειρήσεων, ενώ ταυτόχρονα μπορεί να γίνει πιο εύκολα κατανοητό τι είδη καταστημάτων ενδιαφέρουν κάθε χρήστη. Μέσω αυτών των πληροφοριών γίνεται εφικτό να παρουσιάζονται μηνύματα στην παρουσίαση των συστάσεων, τα οποία μπορούν να εξηγήσουν καλύτερα σε κάθε χρήστη ξεχωριστά γιατί του προτείνεται το κάθε κατάστημα με βάση το ιστορικό των αξιολογήσεων του.

Τέλος ένας τελευταίος παράγοντας που πρέπει να ληφθεί υπόψιν σχετικά με τα καταστήματα είναι τα επίπεδα τιμών τους. Όπως είναι φυσικό κάθε επιχείρηση αναλόγως με τις υπηρεσίες που προσφέρει, υπάρχει και μια διακύμανση στις τιμές. Όπως φαίνεται στο σχήμα 5.7 υπάρχουν επιχειρήσεις για όλες τις περιπτώσεις ανάλογα με το ενδιαφέρον και την οικονομική δυνατότητα του κάθε χρήστη. Σχεδόν 1 στις 4 επιχειρήσεις (23.18%) δεν παρέχει αυτή την πληροφορία, όμως επειδή είναι ένας



Σχήμα 5.7: Επίπεδα τιμών επιχειρήσεων

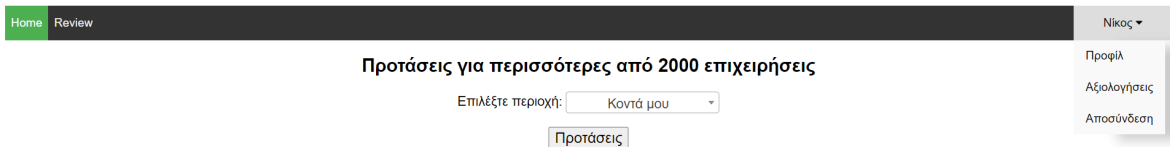
πολύ βασικός παράγοντας που συνεισφέρει στο φιλτράρισμα θεωρήθηκε σωστό να συμπεριληφθεί σαν επιλογή. Οι περισσότερες επιχειρήσεις χαρακτηρίζονται μέτριου οικονομικού επιπέδου, και αυτό ποσοστιαία αντιστοιχεί σχεδόν στις μισές (51.89%). Θεωρητικά είναι ένα αρκετά καλό ποσοστό αν αναλογιστεί κανείς ότι εκτός από τις καφετέριες υπάρχουν και αρκετά εστιατόρια ανάμεσα στα διαθέσιμα καταστήματα. Ίδιο ποσοστό με τις αδιευκρίνιστες, καταλαμβάνουν και οι φθηνές επιχειρήσεις. Τέλος αρκετά μικρά ποσοστά φαίνεται να έχουν οι ακριβές και πολύ ακριβές. Αυτό μπορεί να σώσει αρκετούς χρήστες από δυσάρεστες εκπλήξεις, αποτρέποντας τους να επισκεφτούν κάποιο κατάστημα που ίσως δεν θα μπορούν να υποστηρίξουν οικονομικά.

5.2 Υλοποίηση ως Υπηρεσία

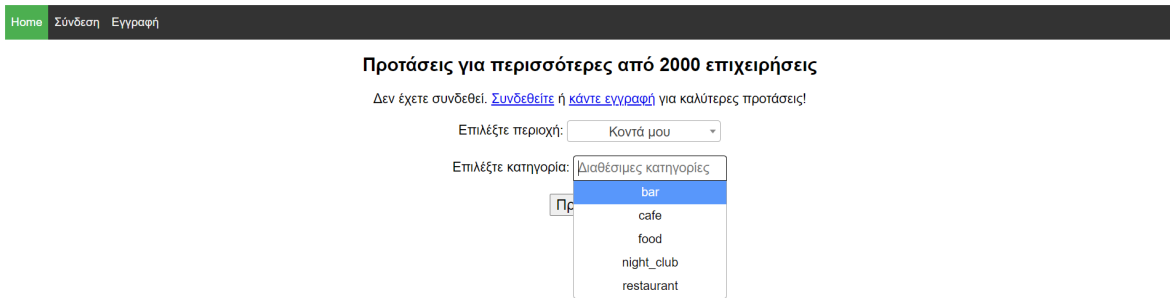
Όπως έχει αναφερθεί αρκετά και στις προηγούμενες ενότητες, στα συστήματα συστάσεων μεγάλη σημασία έχει το περιβάλλον που αλληλεπιδρά ο χρήστης και το πόσο εύκολα γίνεται αυτό. Φυσικά ο αλγόριθμος που παράγει τις συστάσεις είναι εξίσου σημαντικός, όμως και ένας σωστό περιβάλλον για την παροχή των υπηρεσιών είναι μεγάλης σημασίας. Για τη δημιουργία του συγκεκριμένου συστήματος χρησιμοποιήθηκε η Django, προσπαθώντας να δημιουργηθούν όσο το δυνατόν περισσότερες υπηρεσίες γίνεται. Μέσω της Django υποστηρίζονται η εγγραφή και η σύνδεση χρήστη, ενώ ταυτόχρονα πολλές άλλες λειτουργίες με αρκετές σελίδες για κάθε μια από αυτές να δημιουργείται δυναμικά αναλόγως τον χρήστη κάθε φορά. Βασική διαφοροποίηση στη μορφή των σελίδων και τις δυνατότητες που παρέχονται προκύπτει από το αν ο χρήστης είναι συνδεδεμένος ή όχι.

5.2.1 Αρχική σελίδα και συστάσεις

Η βασική διαφορά ξεκινάει στην αρχική σελίδα. Στην περίπτωση που ο χρήστης είναι συνδεδεμένος το σύστημα μπορεί να έχει μια εικόνα για τις προτιμήσεις του, είτε από τις αξιολογήσεις του είτε από το προφίλ του στην περίπτωση που δεν έχει κάνει κάποια αξιολόγηση ακόμα. Σε αυτή τη περίπτωση (σχήμα 5.8a) ο χρήστης χρειάζεται απλά να επιλέξει την περιοχή για την οποία θέλει να λάβει προτάσεις είτε να αφήσει την επιλογή κοντά μου, και με βάση την τοποθεσία του το σύστημα να αποφασίσει μόνο του σε ποιες περιοχές να κοιτάξει για να προτείνει καταστήματα. Στην περίπτωση όμως που ο χρήστης έχει εισέλθει σαν επισκέπτης (σχήμα 5.8b) καμία από αυτές τις πληροφορίες δεν είναι διαθέσιμες και έτσι το σύστημα πρέπει να πάρει κάποιες βασικές πληροφορίες προτού προχωρήσει με τις συστάσεις. Στη συγκεκριμένη περίπτωση δέχεται σαν είσοδο τις κατηγορίες που προτιμάει ο χρήστης να δεχτεί προτάσεις. Στο σχήμα φαίνονται οι 2 διαφορετικές αρχικές σελίδες που παρέχονται αναλόγως με το αν ο χρήστης είναι επισκέπτης ή συνδεδεμένος.



(a) Συνδεδεμένος χρήστης



(b) Χρήστης επισκέπτης

Σχήμα 5.8: Αρχική σελίδα

Επίσης στην περίπτωση του εγγεγραμμένου χρήστη όπως φαίνεται παρουσιάζεται η δυνατότητα δημιουργίας αξιολόγησης στο tab review, όπου από εκεί μπορεί να επιλέξει την επιχείρηση που θέλει να αξιολογήσει και στη συνέχεια να συνεχίσει με τα σχόλια και την βαθμολογία του. Επίσης όταν ο χρήστης έχει συνδεθεί στο σύστημα μπορεί να αποκτήσει μια γενική εικόνα του λογαριασμού του. Υπάρχει ένα αναδυόμενο κουτί με το όνομα του όπου εκεί μπορεί να δει το προφίλ του (στοιχεία που έδωσε κατά την εγγραφή του), καθώς επίσης και τις αξιολογήσεις που έχει ήδη κάνει σε καταστήματα. Έτσι προσφέρεται μια καλύτερη εποπτεία με μεγαλύτερη οργάνωση.

Αφού ο χρήστης πατήσει το κουμπί «προτάσεις», μετά από λίγα δευτερόλεπτα φορτώνει η σελίδα με τις προτάσεις. Όπως και στην αρχική σελίδα έτσι και εδώ υπάρχουν κάποιες μικρές διαφοροποιήσεις με βάση το είδος του χρήστη. Στο σχήμα 5.9 παρουσιάζονται τα αποτελέσματα και για τις 2 περιπτώσεις. Η διαφορά στην εμφάνιση είναι μικρή και έχει να κάνει με την αιτιολόγηση. Επειδή δεν υπάρχουν δεδομένα με προηγούμενες αλληλεπιδράσεις το σύστημα δεν αιτιολογεί το είδος των επιχειρήσεων και τις προηγούμενες επισκέψεις.

Η μεγαλύτερη διαφορά βρίσκεται στον τρόπο που εκτελείται ο αλγόριθμος στην περίπτωση μη εγγεγραμμένου χρήστη.

Στην περίπτωση που ο χρήστης υπάρχει στο σύστημα, εάν έχει κάνει αξιολογήσεις θα υπάρχει και στο εκπαιδευμένο μοντέλο. Συνεπώς οι προτάσεις του θα βασιστούν στο μοντέλο και θα αφορούν το συγκεκριμένο χρήστη. Εάν όμως δεν έχει κάνει αξιολογήσεις, τότε το σύστημα τον βλέπει όπως

Home Results Show Map Νίκος

20 αποτελέσματα

Ταξινόμηση κατά: Προκαθορισμένο

Cafe Restaurant Bar Night club

Επίπεδο Τιμών: Όλα

LOT 51 Έχετε ήδη επισκεφθεί
Παπαδιαμαντοπούλου 24B, Αθήνα
★ 4.4 (2306)
bar, cafe, food Επειδή είναι απόγευμα και σας αρέσουν τα bar
5' 1.67 χλμ Προτείνεται μετάβαση με αυτοκίνητο

My Hood Έχετε ήδη επισκεφθεί
Πανόρμου 16, Αθήνα
★ 4.4 (1586)
cafe, food Επειδή είναι απόγευμα και σας αρέσουν τα cafe
8' 2.76 χλμ Προτείνεται μετάβαση με αυτοκίνητο

Playce Επειδή επισκεφθήκατε το My Hood
Βασίλη Λογοθέτιδη 14, Αθήνα
★ 4.5 (1263)
cafe, food Επειδή είναι απόγευμα και σας αρέσουν τα cafe
8' 2.88 χλμ Προτείνεται μετάβαση με αυτοκίνητο

(a) Συνδεδεμένος χρήστης

Home Results Show Map Νίκος

20 αποτελέσματα

Ταξινόμηση κατά: Προκαθορισμένο

Cafe Restaurant Bar Night club

Επίπεδο Τιμών: Όλα

six d.o.g.s
Αβραμίου 6-8, Αθήνα
★ 4.4 (3940)
bar, cafe, food Επειδή είναι απόγευμα
37' 18.21 χλμ Προτείνεται μετάβαση με αυτοκίνητο

KUKO's The Bar
Καλαμίου 4, Αθήνα
★ 4.5 (2874)
bar, cafe, food Επειδή είναι απόγευμα
34' 18.00 χλμ Προτείνεται μετάβαση με αυτοκίνητο

Black Duck Garden
5, Ιωάννου Παπαρηγοπούλου 7, Αθήνα
★ 4.5 (2214)
bar, cafe, food, restaurant Επειδή είναι απόγευμα
29' 16.03 χλμ Προτείνεται μετάβαση με αυτοκίνητο

(b) Χρήστης επισκέπτης

Σχήμα 5.9: Αποτελέσματα συστάσεων

ακριβώς και ένα μη εγγεγραμμένο χρήστη. Με βάση τις προτιμήσεις του είτε κατά τις προτάσεις είτε κατά την εγγραφή, βρίσκονται χρήστες με παρόμοια χαρακτηριστικά και προτιμήσεις, και οι συστάσεις που προκύπτουν τελικά δεν είναι 100% εξατομικευμένες λόγω της έλλειψης δεδομένων, αλλά έχουν προκύψει με βάση παρόμοιους χρήστες. Το σημαντικό εδώ είναι ότι το σύστημα μπορεί να ανταποκριθεί πλήρως, χωρίς να αντιμετωπίζει πρόβλημα με το φαινόμενο της ψυχρής εκκίνησης.

Και στις 2 περιπτώσεις όπως φαίνεται και στα σχήματα ο χρήστης μπορεί να προσαρμόσει τις συστάσεις στο δικό του στυλ με τους ακόλουθους τρόπους:

- Φιλτράρισμα με βάση τον τύπο της επιχείρησης: Μπορεί να επιλέξει να εμφανίζονται μόνο οι κατηγορίες που επιθυμεί και έτσι να φιλτράρει τα αποτελέσματα με βάση αυτό που θέλει εκείνη τη στιγμή.
- Επίπεδο τιμών: Ομοίως με παραπάνω παρέχεται η δυνατότητα προβολής επιχειρήσεων που ικανοποιούν το επίπεδο τιμών που διαλέγει ο χρήστης.
- Τρόπος ταξινόμησης: Τα αποτελέσματα ταξινομούνται με ένα προεπιλεγμένο τρόπο με βάση το σκορ που έχει προκύψει για κάθε επιχείρηση. Αυτό προκύπτει μέσα από αρκετές μεταβλητές όπως την απόσταση, την βαθμολογία, των αριθμό των κριτικών, το σκορ του LightFm κλπ. Εάν ο χρήστης επιθυμεί τα αποτελέσματα να εμφανιστούν με διαφορετική σειρά έχει τη δυνατότητα να επιλέξει ανάμεσα σε: βαθμολογία, πλήθος κριτικών και απόσταση.

Επίσης κάνοντας κλικ στο όνομα της επιχείρησης παρέχονται συγκεντρωμένες όλες οι πληροφορίες του εκάστοτε καταστήματος μαζί με μια φωτογραφία ώστε να προηδεάζει τον χρήστη για την εμφάνιση σε κάθε περίπτωση. Εκτός από αυτό δίπλα από την εκτιμώμενη απόσταση και το χρόνο μετακίνησης, κάνοντας κλικ εμφανίζονται οι οδηγίες μετάβασης στο συγκεκριμένο μέρος, με το μέσο που προτείνεται κάθε φορά. Στην περίπτωση που κάποιος επιθυμεί να δει όλα τα καταστήματα μαζί στο χάρτη και να τα συγκρίνει, αυτή η δυνατότητα παρέχεται από το tab show map. Εκεί εμφανίζεται ο χάρτης με τα προτεινόμενα καταστήματα, ο οποίος συμπεριλαμβάνει και πληροφορίες όπως αριθμό σύστασης, απόσταση, χρόνο μετάβασης και βαθμολογία.

Τέλος οι συνδεδεμένοι χρήστες μπορούν να οδηγηθούν αμέσως στη σελίδα αξιολόγησης κάθε επιχείρησης πατώντας πάνω στα αστέρια της αξιολόγησης. Αυτό γλιτώνει το χρήστη από τη διαδικασία αναζήτησης του καταστήματος που επιθυμεί να αξιολογήσει.

5.2.2 Αξιολογήσεις και χάρτες

Εκτός από την παροχή συστάσεων το σύστημα πρέπει να προσφέρει κι άλλες λειτουργίες με σκοπό την βελτίωση της εμπειρίας του χρήστη. Η δυνατότητα αξιολογήσεις στοχεύει στη λήψη περισσότερων δεδομένων σχετικά με τις προτιμήσεις των χρηστών ώστε το μοντέλο να μπορεί να εκπαιδευτεί ξανά και να προσαρμόζεται στις αλλαγές των προτιμήσεων που έχουν οι χρήστες. Επίσης η χρήση χαρτών στοχεύει στην διευκόλυνση των χρηστών και την εξοικονόμηση χρόνου. Με αυτόν τον τρόπο δεν είναι αναγκαία η χρήση άλλων λογισμικών ώστε να γνωρίζει ο χρήστης την ακριβή τοποθεσία των επιχειρήσεων και πως θα μεταβεί σε αυτές.

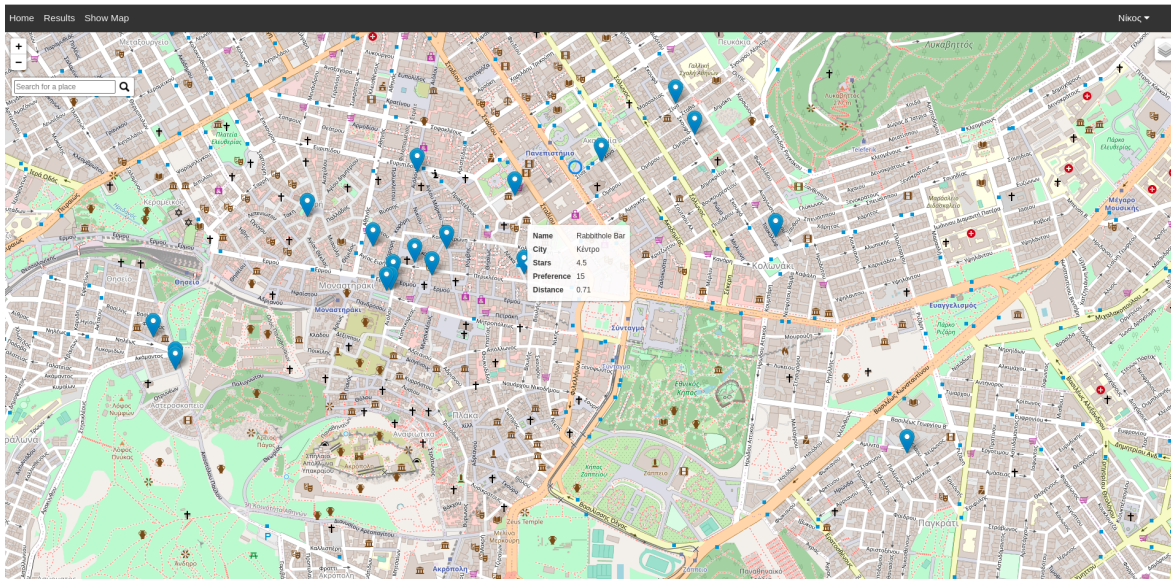
Οδηγίες και προβολή καταστημάτων στο χάρτη

Χρησιμοποιώντας το openrouteserviceAPI αντλούνται πληροφορίες σχετικά με τις αποστάσεις των καταστημάτων από την τοποθεσία του χρήστη, δίνονται εκτιμώμενοι χρόνοι μετάβασης καθώς και προτεινόμενες οδηγίες μετάβασης χρησιμοποιώντας αυτοκίνητο ή ως πεζοί. Είναι πολύ σημαντικό ο χρήστης να γνωρίζει εκ των προτέρων πόσος χρόνος περίπου θα χρειαστεί για να μεταβεί σε μια συγκεκριμένη τοποθεσία. Για το λόγο αυτό παρέχονται συγκεντρωμένες τέτοιου είδους πληροφορίες δίπλα από κάθε πρόταση. Για να υπάρχει επίσης μια καλύτερη γενική εικόνα για το που βρίσκονται όλα τα καταστήματα, παρέχεται ένας χάρτης κάθε φορά με ομαδοποιημένα όλα τα αποτελέσματα όπως φαίνεται στο σχήμα 5.10.

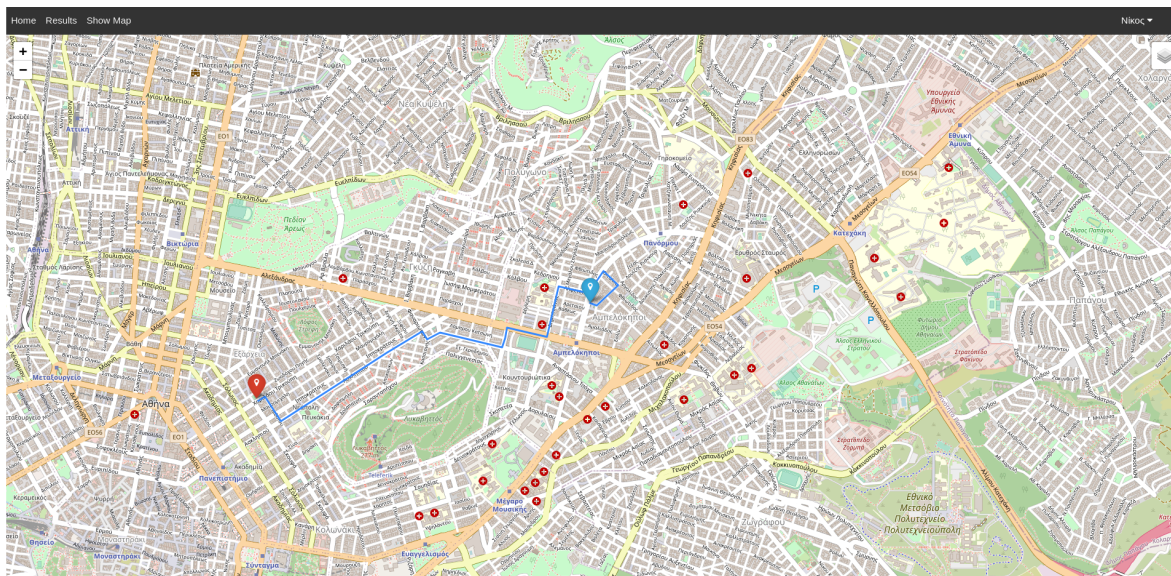
Για να μη χρειάζεται ο χρήστης να επιστρέφει στη σελίδα των αποτελεσμάτων για να δει τις πληροφορίες του κάθε καταστήματος, έχουν συμπεριληφθεί όπως φαίνεται παρακάτω. Πιο συγκεκριμένα παρέχονται: όνομα, περιοχή, σειρά κατάταξης σύστασης, βαθμολογία και απόσταση. Οι συγκεκριμένες πληροφορίες θεωρούνται από τις πιο βασικές. Επίσης αν θέλει ο χρήστης να αναζητήσει ένα συγκεκριμένο κατάστημα που είδε στη λίστα με τις προτάσεις, δεν έχει παρά να αρχίσει να πληκτρολογεί το όνομα του στο κουτί αναζήτησης. Με αυτούς τους τρόπους η εργασία αυτή προσπαθεί να αυξήσει όσο το δυνατόν περισσότερο τη φιλικότητα και ευελιξία που παρέχεται στον χρήστη από το περιβάλλον. Το κυκλικό στίγμα στο χάρτη σηματοδοτεί την τοποθεσία του χρήστη, ώστε να μπορεί να καταλάβει σε τι απόσταση από εκείνον βρίσκονται οι προτεινόμενες επιχειρήσεις. Επίσης με παρόμοιο τρόπο εμφανίζονται και οι οδηγίες μετάβασης (5.10b), έχοντας ως σημεία την τοποθεσία του χρήστη και το σημείο του καταστήματος ώστε να γίνει γνωστό εκ των προτέρων ποια είναι η πιθανή διαδρομή που θα πρέπει να ακολουθηθεί χωρίς να χρειάζεται να κάνει αναζήτηση σε κάποια άλλη εφαρμογή όπως το google maps.

Αξιολογήσεις

Τέλος, η τελευταία δυνατότητα που παρέχεται και απευθύνεται μόνο σε εγγεγραμμένους χρήστες είναι η προσθήκη αξιολογήσεων. Όπως αναφέρθηκε και παραπάνω αυτό γίνεται είτε μέσω του tab review και στη συνέχεια με αναζήτηση του συγκεκριμένου καταστήματος που ο χρήστης επιθυμεί να αξιολογήσει, είτε μέσω των συστάσεων που του έχουν προταθεί κάνοντας κλικ στη μέση βαθμολογία



(a) Συγκεντρωτικά αποτελέσματα σε χάρτη



(b) Οδηγίες μετάβασης

Σχήμα 5.10: Χρήση χαρτών για προβολή και μετάβαση

της επιχείρησης. Το περιβάλλον της αξιολόγησης έχει τη μορφή που φαίνεται στο Σχήμα 5.11. Ο χρήστης προσθέτει έναν τίτλο, τα σχόλια του σχετικά με την εμπειρία του, και τέλος τη βαθμολογία. Κάθε φορά που προστίθενται 5 νέες αξιολογήσεις, το σύστημα επανεκπαιδεύει το μοντέλο συστάσεων. Αυτό είναι ιδιαίτερα σημαντικό για τους χρήστες που κάνουν τις πρώτες τους αξιολογήσεις, αφού μέχρι πριν αυτό οι συστάσεις δεν ήταν τόσο εξατομικευμένες. Όσο οι χρήστες κάνουν αξιολογήσεις το σύστημα ανταποκρίνεται όλο και καλύτερα, προσφέροντας πιο συγκεκριμένες συστάσεις που ανταποκρίνεται καλύτερα στα ενδιαφέροντα τους. Όσο περισσότερες αξιολογήσεις κάνει ένας χρήστης τόσο περισσότερο τον μαθαίνει το σύστημα και βελτιώνει τις συστάσεις.

Επίσης για να μην υπάρχουν διπλές αξιολογήσεις, στην περίπτωση που ένας χρήστης ξανά βαθμολογήσει μια επιχείρηση που είχε αλληλεπιδράσει και στο παρελθόν, δεν δημιουργείται καινούρια. Αντιθέτως ενημερώνεται η υπάρχουσα και το σύστημα κάνει στις αντίστοιχες αλλαγές στη βάση. Με αυτόν τον τρόπο αποφεύγεται επίσης η περίπτωση που κάποιος χρήστης μπορεί να επηρεάσει την συνολική εικόνα της επιχείρησης κάνοντας πολλές θετικές ή αρνητικές αξιολογήσεις αντίστοιχα.

The screenshot shows a review form for 'My Hood'. At the top, there is a navigation bar with 'Home' and 'Reviews' on the left, and 'Nikos' on the right. Below the navigation bar, the product name 'My Hood' is displayed, followed by 'Πανόρμου 16, Αθήνα' and a rating of '4.4' with five stars and '1586 αξιολογήσεις'. The form itself has a title field labeled 'Τίτλος:', a large text area for the review labeled 'Κριτική:', and a five-star rating system below it. A 'Υποβολή' (Submit) button is located at the bottom of the form.

Σχήμα 5.11: Περιβάλλον αξιολόγησης

5.3 Περιπτώσεις χρήσης

Υπάρχουν αρκετές διαφορετικές περιπτώσεις χρήσης, όπου το σύστημα προσαρμόζεται και τα αποτελέσματα του διαφοροποιούνται. Οι παράγοντες που οδηγούν σε αυτές τις αλλαγές έχουν να κάνουν κυρίως με τα δεδομένα επίγνωσης πλαισίου. Οι κυριότερες μεταβλητές που το ορίζουν είναι η τοποθεσία του χρήστη σε συνδυασμό με τη χρονική στιγμή της ημέρας. Για παράδειγμα θα ήταν απαγορευτικό ο χρήστης να θέλει να επισκεφτεί επιχειρήσεις που βρίσκονται κοντά του, και εν τέλει να του προτείνονται καταστήματα 20 χιλιόμετρα μακριά. Επίσης θα ήταν λάθος το πρωί να προτείνονται εστιατόρια, ή ενώ είναι βράδυ να προταθούν μέρη που προσφέρουν μόνο καφέ. Το σύστημα πρέπει να έχει την ικανότητα να λαμβάνει υπόψην όλες αυτές τις πληροφορίες και τους περιορισμούς προτού προχωρήσει στις συστάσεις. Παρακάτω παρουσιάζονται ορισμένες τέτοιες περιπτώσεις ώστε να γίνει κατανοητό πως το σύστημα ανταποκρίνεται σε αυτές τις αλλαγές και τις διαφορετικές περιπτώσεις.

Αρχικά παρουσιάζονται οι περιπτώσεις που αφορούν και τα 2 είδη χρηστών και δεν διαφοροποιούνται σε αυτή την περίπτωση. Στην συνέχεια επειδή υπάρχουν διαφοροποιήσεις, η ανάλυση χωρίζεται σε 2 κατηγορίες.

5.3.1 Χρονική στιγμή της ημέρας

Μεγάλη βαρύτητα για τη δημιουργία συστάσεων έχει δοθεί στη χρονική στιγμή της ημέρας. Στη συγκεκριμένη υλοποίηση η ημέρα χωρίζεται στις εξής κατηγορίες:

- Πρωί(7π.μ - 1μ.μ): Η πρώτη χρονική στιγμή της ημέρας αφορά το πρωί και χωρίζεται με βάση το εύρος που αναφέρθηκε. Σε αυτή τη χρονική στιγμή της ημέρας οι προτάσεις που δίνονται αφορούν μόνο καφετέριες, αφού θα ήταν παράλογο να προταθεί οποιοσδήποτε άλλος τύπος καταστήματος.
- Μεσημέρι(1μ.μ - 5μ.μ): Ακολουθεί το μεσημέρι. Σε αυτή τη χρονική περίπτωση αν και τα επικρατέστερα καταστήματα αφορούν το φαγητό, επειδή δεν λειτουργούν όλοι οι χρήστες με τον ίδιο τρόπο εδώ παρέχονται 2 επιλογές. Χρησιμοποιούνται βάρη, με μεγαλύτερη βαρύτητα να δίνεται στα εστιατόρια που είναι πιθανότερο να προτιμηθούν αυτή τη χρονική στιγμή, ενώ επίσης συμπεριλαμβάνονται και οι καφετέριες επειδή υπάρχει μια ομάδα χρηστών που συνηθίζει να προτιμάει αυτή τη περίοδο της ημέρας για καφέ.
- Απόγευμα (5μ.μ - 9.μ.μ): Συνεχίζοντας με το απόγευμα εδώ οι επιλογές είναι περισσότερες. Ανάλογα το πρόγραμμα και την καθημερινότητα κάθε χρήστη οι προτιμήσεις ποικίλουν. Πιο συγκεκριμένα σε αυτή τη περίοδο της ημέρας κυριαρχούν τα καφέ όπου και δίνεται το μεγαλύτερο βάρος. Ακολουθούν τα εστιατόρια με μικρή διαφορά, ενώ αρκετά μικρότερο βάρος έχουν τα μπαρ. Η τελευταία κατηγορία προστέθηκε αφού υπάρχουν αρκετές ομάδες χρηστών που η συγκεκριμένη ώρα τους βρίσκει να ολοκληρώνουν την εργασία τους και είναι συχνό να πηγαίνουν για μια μπύρα ή ένα χαλαρό ποτό με τους συναδέλφους τους.

- Βράδυ (9μ.μ - 12π.μ) Πλέον βρισκόμαστε σε μια περίοδο που τα καφέ βγαίνουν από το πλάνο. Εδώ κυριαρχούν τα μπαρ, αλλά υπάρχει και η επιλογή των εστιατορίων ώστε να καλύπτονται όλες οι διαφορετικές ομάδες χρηστών.
- Αργά το βράδυ(12π.μ - 7π.μ) Τέλος έχουμε την τελευταία κατηγορία. Εδώ περιλαμβάνονται μόνο επιχειρήσεις που αφορούν την νυχτερινή ζωή. Αυτά είναι τα μπαρ και τα night clubs, με τα δεύτερα να έχουν ελαφρώς μεγαλύτερο βάρος από τα πρώτα.

Εδώ πρέπει να σημειωθεί το εξής: Πληθώρα των καταστημάτων που έχουν χρησιμοποιηθεί σε αυτή την εργασία ανήκουν σε παραπάνω από μια κατηγορίες. Αυτό είναι αρκετά βοηθητικό αφού δεν περιορίζει το σύστημα με την έννοια ότι ένα κατάστημα ανάλογα με την ώρα της ημέρας καλύπτει διαφορετικές κατηγορίες.

Επίσης λαμβάνοντας την ακριβή χρονική στιγμή της ημέρας, και χρησιμοποιώντας τις διαθέσιμες πληροφορίες όλων των επιχειρήσεων τα αποτελέσματα φιλτράρονται αποκλείοντας εκείνα που είναι κλειστά την συγκεκριμένη χρονική στιγμή. Όμως επειδή είναι πολύ πιθανόν κάποια από αυτά να ανοίγουν άμεσα, όσα ανοίγουν σύντομα περιλαμβάνονται στις προτάσεις ενημερώνοντας όμως το χρήστη για τη συγκεκριμένη ώρα που το μαγαζί θα είναι ανοικτό.

5.3.2 Τοποθεσία χρήστη

Η τοποθεσία του χρήστη είναι μια ακόμη σημαντική μεταβλητή που παίζει κομβικό ρόλο στις συστάσεις. Χρησιμοποιείται με 2 τρόπους. Αρχικά στην περίπτωση που ο χρήστης αναζητά μαγαζιά κοντά του, ορίζεται ένα όριο ακτίνας πέρα από το οποίο τα μαγαζιά δεν συμπεριλαμβάνονται αφού δεν θεωρούνται κοντά στη τοποθεσία του χρήστη. Υπάρχουν 2 ειδών αποστάσεις που υπολογίζονται: Η απόσταση σε ευθεία γραμμή που δεν απαιτεί ερωτήματα σε κάποιο API, αλλά υπολογίζεται μόνο από το γεωγραφικό πλάτος και έχει να κάνει με την ακτίνα που αναφέρθηκε. Η δεύτερη απόσταση που υπολογίζεται έχει να κάνει με το αν χρησιμοποιηθεί αυτοκίνητο ή μετάβαση με τα πόδια. Ο υπολογισμός του μεταφορικού μέσου προκύπτει αναλόγως με την απόσταση σε ευθεία γραμμή. Αν το σύστημα θεωρήσει ότι ο προορισμός βρίσκεται αρκετά μακριά από το χρήστη προτείνει αυτοκίνητο. Αυτή η επιλογή χρησιμοποιείται στη συνέχεια για να γίνει το ερώτημα σε ένα συγκεκριμένο API που παρέχει λεπτομέρειες σχετικά με το χρόνο μετάβασης και την ακριβή απόσταση σε πραγματικό χρόνο.

Επίσης η τοποθεσία εκτός από το φιλτράρισμα των αποτελεσμάτων χρησιμοποιείται στα αποτελέσματα ώστε να μπορεί ο χρήστης να τα ταξινομήσει με βάση αυτή. Έτσι μπορεί να γνωρίζει με λεπτομέρεια ποια από τα καταστήματα που προτάθηκαν βρίσκονται πιο κοντά του και ποια όχι. Επιπροσθέτως όταν το σύστημα υπολογίζει τα βάρη για κάθε μεταβλητή δίνει άλλη βαρύτητα όταν ένα μέρος έχει προταθεί για μετάβαση με τα πόδια επειδή ο χρήστης ψάχνει κάτι κοντά του, και διαφορετικό βάρος όταν έχει επιλεγεί μια περιοχή που δεν είναι τόσο κοντά του από άποψη απόστασης. Η απόσταση του χρήστη από όλες τις επιχειρήσεις που πρόκειται να προταθούν κοινωνικοποιείται με βάση την ελάχιστη και τη μέγιστη απόσταση και στη συνέχεια εφαρμόζεται το βάρος ανάλογα με τον τρόπο μετάβασης που έχει προταθεί.

5.3.3 Χρήστες

Επειδή υπάρχουν αρκετές διαφοροποιήσεις σχετικά με το πως δημιουργούνται συστάσεις μεταξύ των 2 διαφορετικών ειδών χρηστών, για την καλύτερη εξήγηση και κατανόηση οι περιπτώσεις έχουν χωριστεί σε 2 κατηγορίες. Στην περίπτωση των συνδεδεμένων χρηστών υπάρχουν 2 περιπτώσεις. Στην πρώτη ο χρήστης ενώ υπάρχει στο σύστημα δεν έχει κάνει ακόμα αξιολογήσεις με αποτέλεσμα να μην έχει συμπεριληφθεί στο μοντέλο εκπαίδευσης. Σε αυτή την περίπτωση λαμβάνεται υπόψιν μόνο το προφίλ του με τις προτιμήσεις που έδωσε κατά την εγγραφή του. Έτσι χρησιμοποιούνται δεδομένα από χρήστη με παρόμοιες προτιμήσεις σε επιχειρήσεις και γίνονται προτάσεις. Στην δεύτερη περίπτωση ο χρήστης έχει κάνει αξιολογήσεις και έχει συμπεριληφθεί στο μοντέλο. Έτσι οι

προτάσεις είναι πλήρως εξατομικευμένες. Κάθε φορά που αξιολογεί κι άλλα καταστήματα το μοντέλο ενημερώνεται και προσαρμόζεται στις αλλαγές.

Σε αυτό το είδος χρηστών το σύστημα λειτουργεί ακριβώς όπως στην πρώτη περίπτωση της προηγούμενης κατηγορίας. Η μόνη διαφορά σε σχέση με πριν είναι ότι πλέον οι χρήστες πρέπει να εισάγουν χειροκίνητα τις προτιμήσεις τους αντιθέτως με πριν που αυτές υπήρχαν καταχωρημένες στη βάση του συστήματος.

Βασικό από όλα τα παραπάνω είναι ότι το σύστημα καταφέρνει να προσαρμόζεται αυτοματοποιημένα σε όλες τις καταστάσεις χρησιμοποιώντας πληροφορίες επίγνωσης πλαισίου, χωρίς να απαιτούνται περισσότερες και έξτρα ενέργειες από τους χρήστες. Όλες οι διεργασίες γίνονται στο παρασκήνιο χωρίς ο χρήστης να καταλαβαίνει τις διαφοροποιήσεις που υπάρχουν. Σημασία έχει αυτό να φαίνεται στον τρόπο που προκύπτουν τα αποτελέσματα σε κάθε διαφορετική περίπτωση με σκοπό την καλύτερη παραγωγή συστάσεων.

Κεφάλαιο 6

Συμπεράσματα και Επεκτάσεις

6.1 Συμπεράσματα

Συνολικά, αυτή η διπλωματική επικεντρώνεται στο πρόβλημα της μη στασιμότητας στα Συστήματα Συστάσεων (RS) Αρχικά, προσπαθώντας να προσφέρει σχετικές προτάσεις, το σύστημα προσπαθεί να αποκαλύψει τις προτιμήσεις των χρηστών για στοιχεία που βασίζονται στη συμπεριφορά του χρήστη στο παρελθόν, δεδομένου ότι αυτό ορίζει τα ενδιαφέροντα των χρηστών. Κατά μήκος αυτού του κύματος, ένας αμέτρητος αριθμός έργων έχουν προτείνει προσεγγίσεις προτάσεων που μαθαίνουν προτιμήσεις με βάση τις αλληλεπιδράσεις χρηστών και μοντέλα χρηστών και αντικειμένων σε έναν κοινό χώρο. Αυτές οι προσεγγίσεις υποθέτουν ότι τα ενδιαφέροντα των χρηστών και οι αντιλήψεις των αντικειμένων παραμένουν στατικές με την πάροδο του χρόνου και σε όλες τις περιστάσεις.

Παρόλο που είναι αλήθεια ότι η ανάλυση της συμπεριφοράς των χρηστών παρέχει πληροφορίες σχετικά με τις προτιμήσεις για αντικείμενα, το περιβάλλον που οριοθετείται από ένα σύστημα συστάσεων είναι περίπλοκο και δυναμικό και περιλαμβάνει πολλές οντότητες που εξελίσσονται η καθεμιά από μόνη της και με μοναδικό τρόπο. Οι δυναμικές που συμβαίνουν στους χρήστες και τα στοιχεία μπορούν να εξηγηθούν από την παρουσία διαφορετικών παραγόντων που επηρεάζουν το περιβάλλον. Προηγούμενες έρευνες έχουν αξιολογήσει τη συνάφεια του περιβάλλοντος στα συστήματα συστάσεων και έχουν προτείνει το συστήματα συστάσεων επίγνωσης πλαισίου (CARS). Αυτά τα συστήματα ενσωματώνουν παράγοντες περιβάλλοντος για καλύτερη ποιότητα προτάσεων. Σε αυτή την εργασία, συζητάμε τους περιορισμούς των συμβατικών CARS στο πεδίο εφαρμογής πραγματικών εφαρμογών: συστάσεις καταστημάτων εστίασης. Ενώ η παραδοσιακή άποψη του CARS θεωρεί πλήρως παρατηρήσιμο πλαίσιο, υποστηρίζουμε ότι το πλαίσιο είναι εν μέρει παρατηρήσιμο στον τομέα της εστίασης και μελετάμε ανεπιτήρητο πλαίσιο στο διαδικτυακό περιβάλλον, απαιτώντας την ανάπτυξη νέων κατάλληλων προσεγγίσεων.

Στο ευρύτερο επίπεδο, οι συνεισφορές αυτής της διατριβής εμπίπτουν σε μία από τις ακόλουθες τρεις κατηγορίες:

- **Ανάλυση:** Αναλύουμε διεξοδικά και παρουσιάζουμε τα προβλήματα που αντιμετωπίζονται σε αυτή τη διατριβή, τα οποία είναι τα προβλήματα σύστασης εστίασης. Διατυπώνουμε το πρόβλημα προτάσεων εστιατορίων και καφέ, συζητάμε τη σχέση του με άλλα γνωστά προβλήματα προτάσεων και παρουσιάζουμε τα χαρακτηριστικά και τις συγκεκριμένες προκλήσεις που καθιστούν ανεπαρκή τη χρήση παραδοσιακών προσεγγίσεων προτάσεων. Παρέχουμε επίσης μια περιγραφή της έννοιας του πλαισίου επίγνωσης ως αναδύμενου στον τομέα της εστίασης και συζητάμε γιατί χαρακτηρίζεται ως εν μέρει παρατηρήσιμο, σύμφωνα με τον ορισμό μας στο Κεφάλαιο 2. Από την άλλη πλευρά, αφού υπογραμμίσουμε τη σημασία του να εξετάσουμε τη διαδικτυακές προτάσεις, διατυπώνουμε το πρόβλημα της διαδικτυακής προσαρμοστικής πρότασης βάσει ενός πλαισίου εμπνευσμένου από την προσαρμοστική μάθηση στην εξόρυξη ροής δεδομένων. Συζητάμε τους περιορισμούς των σχετικά λίγων υπάρχουσών προσεγγίσεων που προσπαθούν να λάβουν υπόψη τις μετακινήσεις που συμβαίνουν λόγω μη παρατηρήσιμου περιβάλλοντος στο διαδικτυακό RS. Εξηγούμε πώς υιοθετούμε μια ουσιαστικά διαφορετική κατεύθυνση για την αντιμετώπιση του προβλήματος.
- **Αλγόριθμοι:** Σχεδιάζουμε νέες προσεγγίσεις και αλγόριθμους με γνώμονα το περιβάλλον που

λαμβάνουν υπόψη τη δυναμική που υπάρχει στις αλληλεπιδράσεις των χρηστών. Λαμβάνοντας υπόψη τις προτάσεις εστιατορίων, προτείνουμε προσεγγίσεις προτάσεων που ενσωματώνουν παράγοντες περιβάλλοντος που επηρεάζουν τη διαδικασία λήψης αποφάσεων του χρήστη. Αυτοί οι παράγοντες περιλαμβάνουν τη τοποθεσία και τις χρονικές διαστάσεις, κριτικές που εξάγονται από το διαδικτυακό περιβάλλον. Ορισμένοι από αυτούς τους παράγοντες παρέχονται ρητά από το χρήστη, ενώ άλλοι συλλέγονται μόνο από το σύστημα και μερικοί από αυτούς παρατηρούνται σε βοηθητικούς τομείς διαφορετικούς από τον τομέα της εστίασης. Οι προτεινόμενες προσεγγίσεις επιτρέπουν την ενσωμάτωση δεδομένων που σχετίζονται με το περιβάλλον χρήστη στην εστίαση προκειμένου να βελτιωθεί η ποιότητα των προτάσεων. Λαμβάνοντας υπόψη τις διαδικτυακές προτάσεις, προτείνουμε νέους διαδικτυακούς αλγόριθμους που λαμβάνουν υπόψη τις μετακινήσεις στις προτιμήσεις των χρηστών, τις αντιλήψεις στοιχείων και τις περιγραφές στοιχείων. Η καινοτομία των προσεγγίσεων μας βασίζεται στην ενεργή ανίχνευση αλλαγών στις προτιμήσεις και στη δυναμική ενημέρωση μοντέλων σύμφωνα με τις παρατηρούμενες αλλαγές, και όχι μόνο παθητική ενημέρωση μοντέλων με συνεχή αύξηση. Οι προσεγγίσεις μας επιτρέπουν τη συνεχή ενσωμάτωση σχολίων που δημιουργούνται από τους χρήστες και διατηρώντας τα μοντέλα προτάσεων ενημερωμένα σε πραγματικό χρόνο.

- **Διδάγματα.** Αξιολογούμε τις προτεινόμενες προσεγγίσεις σε σύνολα δεδομένων πραγματικού κόσμου, τις συγκρίνουμε με τις υπάρχουσες προσεγγίσεις και εξάγουμε συμπεράσματα σχετικά με την απόδοση των μεθόδων και τα οφέλη της συνεκτίμησης της δυναμικής πλευράς των δεδομένων που δημιουργούνται από τον χρήστη. Τα πειράματα που πραγματοποιήθηκαν στο πλαίσιο της σύστασης καταστημάτων εστίασης δείχνουν το ενδιαφέρον της χρήσης πληροφοριών επίγνωσης πλαισίου για τη βελτίωση της ποιότητας των προτάσεων. Υπογραμμίζουν περαιτέρω την ευαισθησία διαφορετικών υποομάδων χρηστών στους θεωρούμενους παράγοντες περιβάλλοντος. Τα πειράματα που πραγματοποιήθηκαν στο πλαίσιο της διαδικτυακής πρότασης δείχνουν τη σημασία της ενεργού λογιστικής καταχώρησης διαφορετικών τύπων κινήσεων σε επίπεδο χρήστη και στοιχείου. Τα μοντέλα διατηρούνται προσαρμοσμένα στις τρέχουσες καταστάσεις των χρηστών και των αντικειμένων, καθώς και στη σύσταση η ποιότητα βελτιώνεται σε σύγκριση με άλλες σταδιακές προσεγγίσεις. Τα αποτελέσματα δείχνουν επίσης πώς οι συμβατικές προσεγγίσεις προτάσεων που μαθαίνουν με σταθερό ρυθμό αποτυγχάνουν να προσαρμοστούν στις μετακινήσεις και να συμβάλλουν στην επιδείνωση της απόδοσης του συστήματος.

Τώρα ολοκληρώνουμε τη διπλωματική εργασία παρουσιάζοντας ανοιχτές προκλήσεις και κατευθύνσεις για μελλοντική έρευνα, που σχετίζονται με καθένα από τα προβλήματα που μελετήθηκαν.

6.2 Μελλοντικές Επεκτάσεις

Υπάρχουν πολλές πιθανές κατευθύνσεις για μελλοντική εργασία, μερικές από τις οποίες αναφέρονται παρακάτω.

Σύσταση στην εστίαση

Το πρόβλημα της σύστασης επιχειρήσεων εστίασης μπορεί πρώτα να επωφεληθεί από την ενσωμάτωση επιπλέον παραγόντων επίγνωσης πλαισίου που επηρεάζουν τους χρήστες και που δεν εξετάστηκαν σε αυτήν την εργασία. Αναφέρουμε συγκεκριμένα τον αντίκτυπο της πρόγνωσης του καιρού. Η ελκυστικότητα των προορισμών είναι γνωστό ότι αλλάζει υπό διαφορετικές καιρικές συνθήκες και οι χρήστες συχνά ελέγχουν την πρόγνωση καιρού όταν θέλουν να πάνε κάπου. Αν και θεωρούμε ότι η χρονικότητα και η εποχικότητα είναι ένα πρώτο βήμα για την αντιμετώπιση αυτού του προβλήματος, δεν είναι επαρκής λύση. Αυτό ισχύει ιδιαίτερα δεδομένου ότι ο καιρός δεν είναι σταθερός κατά τη διάρκεια μιας σεζόν και ότι δεν είναι ο ίδιος σε πολλές εμφανίσεις της ίδιας σεζόν. Η ενσωμάτωση

της πρόγνωσης καιρού στα συστήματα συστάσεων στον τομέα της εστίασης θα μπορούσε να ωφεληθεί η σύσταση, αλλά απαιτεί την κατάλληλη διαμόρφωση της διάστασης του καιρού εκτός από το σχεδιασμό κατάλληλων μεθόδων.

Μια άλλη μεταβλητή που θα μπορούσε να ωφεληθεί η σύσταση στον τομέα της εστίασης και η οποία δεν μελετήθηκε σε αυτή τη διατριβή αφορά τον αντίκτυπο της κυκλοφοριακής κίνησης. Αρκετοί παράγοντες μπορούν να επηρεάσουν αυτή τη μεταβλητή όπως την ημέρα της εβδομάδας, την ώρα της ημέρας και την εποχή. Ιδανικά η καλύτερη λύση θα ήταν να αντλούνται δεδομένα σχετικά με την κίνηση στους δρόμους σε πραγματικό χρόνο ώστε να μπορεί ο χρήστης να ενημερωθεί ότι κάποιοι προορισμοί θα είναι δύσκολο να επισκεφθούν λόγω των συνθηκών που επικρατούν. Η συνεκτίμηση όλων των σχετικών παραγόντων σε ένα σύστημα συστάσεων απαιτεί επίσης περαιτέρω εξελίξεις που θα συνδυάζουν τα οφέλη που αποκτά ο καθένας από αυτούς.

Επίσης σχετικό με αυτό και αρκετά χρήσιμο θα ήταν να υπάρχουν πληροφορίες σχετικά με την πληρότητα της κάθε επιχείρησης τη χρονική στιγμή που επιθυμεί να λάβει προτάσεις ο χρήστης. Είναι αρκετές οι περιπτώσεις που ενώ οι χρήστες επισκέπτονται μέρη που τους αρέσουν, στη συνέχεια το μετανιώνουν λόγω του μεγάλου όγκου κόσμου που έχει επισκεφτεί επίσης αυτή τη χρονική στιγμή. Αυτό ισοδυναμεί με την αύξηση του χρόνου αναμονής και με την αρνητική εμπειρία του χρήστη. Συνεπώς είναι αναγκαία η λήψη δεδομένων πληρότητας ή η πρόβλεψη της για τη συγκεκριμένη χρονική στιγμή με βάση ιστορικά δεδομένα.

Με βάση όλα τα παραπάνω, ένας μελλοντικός στόχος θα ήταν η συνεργασία μεταξύ του συστήματος μας και των επιχειρήσεων. Η ανταλλαγή δεδομένων μεταξύ του συστήματος και της επιχείρησης θα μπορούσε να επωφεληθεί ταυτόχρονα και την επιχείρηση, αυξάνοντας τα κέρδη της αλλά και βελτιώνοντας την οργάνωση της. Άμεσο αποτέλεσμα αυτής της ενέργειας θα ήταν και η παροχή προτάσεων λαμβάνοντας υπόψη δεδομένα πραγματικού χρόνου που δεν θα βασιζόνταν μόνο σε προβλέψεις αλλά θα είχαν φυσική υπόσταση και μεγάλη αξιοπιστία αφού θα τα παρείχαν οι ίδιες οι επιχειρήσεις.

Online προτάσεις

Όπως αναφέρθηκε σε αυτή την εργασία, η διαδικτυακή προσαρμοστική πρόταση παραμένει ένα ανεξερεύνητο πρόβλημα, παρόλο που σχετίζεται με πραγματικές εφαρμογές. Καθίσταται απαραίτητο να υιοθετηθεί η διαδικτυακή ρύθμιση για σύσταση, ειδικά δεδομένου ότι θα ακυρώσει πολλές τεχνικές προτάσεων που έχουν καλή απόδοση σε παρτίδες. Ενώ αποδεικνύουμε σε αυτήν τη διατριβή τη σημασία του ενεργού εντοπισμού και προσαρμογής στις μετακινήσεις σε RS σε πραγματικό χρόνο, υπάρχουν πολλές κατευθύνσεις για τη βελτίωση της διαδικτυακής σύστασης που παρουσιάζουμε παρακάτω.

Η ενότητα ανίχνευσης αλλαγών. Ενώ η εργασία που παρουσιάζεται σε αυτή τη διατριβή εξετάζει αυξητικές και απότομες μετακινήσεις, οι επαναλαμβανόμενες μετακινήσεις εμφανίζονται συχνά στο online RS. Οι χρήστες μπορεί να επηρεαστούν από επαναλαμβανόμενους παράγοντες που προέρχονται από μεταβλητές εποχικότητας. Ο χειρισμός τέτοιων κινήσεων επιτρέπει στο σύστημα να επωφεληθεί από προηγούμενες παρατηρήσεις που αντιστοιχούν στις ίδιες ιδέες. Απαιτεί τη διαχείριση μιας μνήμης παλαιότερων καταστάσεων προκειμένου να αξιοποιηθεί η γνώση που αποκτήθηκε προηγουμένως. Οι μελλοντικές εργασίες πρέπει να εξετάσουν το ενδεχόμενο ανάπτυξης διαδικτυακών προσεγγίσεων προτάσεων που θα αντιστοιχούν σε αυτούς τους τύπους αλλαγών. Πρέπει επίσης να σχεδιαστούν συγκεκριμένες μέθοδοι που να χειρίζονται όλους τους τύπους αλλαγών που εμφανίζονται στο επίπεδο χρήστη και στοιχείου ταυτόχρονα και σε ένα πλαίσιο.

Η ενότητα ανάκτησης. Το πρόβλημα της αποτελεσματικής ανάκτησης συστάσεων προέκυψε σε πρόσφατες έρευνες ειδικά δεδομένου ότι τα συστήματα πραγματικού κόσμου χειρίζονται μεγάλο αριθμό χρηστών και αντικειμένων. Σε αυτές τις ρυθμίσεις και λαμβάνοντας υπόψη το πλαίσιο Matrix Factorization (MF), η κατασκευή ολόκληρης της μήτρας ανατροφοδότησης δεδομένου ενός μοντέλου που έχει μάθει απαιτεί μεγάλους υπολογισμούς και πόρους. Ως εκ τούτου, μια σειρά από έργα έχουν προτείνει λύσεις σε αυτό το πρόβλημα, αλλά η εστίαση έχει δοθεί στην αποτελεσματική ανάκτηση συστάσεων μόλις εκπαιδευτεί το μοντέλο. Αυτές οι προσεγγίσεις δεν μπορούν να εφαρμοστούν για online συστήματα όπου τα μοντέλα ενημερώνονται συνεχώς. Ωστόσο, η υιοθέτηση παρόμοιων στρα-

τηγικών είναι απαραίτητη στο διαδικτυακό περιβάλλον που υπόκειται σε περιορισμούς όσον αφορά τους πόρους και τον χρόνο υπολογισμού. Οι μελλοντικές εργασίες πρέπει να εξετάσουν την ανάπτυξη αποτελεσματικών μεθοδολογιών για την ανάκτηση συστάσεων σε ένα διαδικτυακό περιβάλλον όπου τα μοντέλα ενημερώνονται συνεχώς.

Η ενότητα αξιολόγησης. Η προκαταρκτική αξιολόγηση αντιμετωπίζει πολλούς περιορισμούς στο πλαίσιο της ηλεκτρονικής σύστασης. Συγκεκριμένα, οι συστάσεις αξιολογούνται μόνο με βάση τη μοναδική παραληφθείσα παρατήρηση. Σχετικά προτεινόμενα αντικείμενα που επιλέχθηκαν στο εγγύς παρελθόν ή που θα επιλεγούν στο εγγύς μέλλον μπορεί να τιμωρηθούν. Επιπλέον, η προηγούμενη εργασία που σχετίζεται με τις προτάσεις σε διαδικτυακά περιβάλλοντα λαμβάνει υπόψη μόνο την αξιολόγηση της ακρίβειας και της κατάταξης των προτάσεων, π.χ. χρήση ανάκλησης και DCG. Οι μετρήσεις που σχετίζονται με άλλα κριτήρια όπως η διαφορετικότητα και η καινοτομία πρέπει να προσαρμοστούν στο διαδικτυακό περιβάλλον και να αξιολογηθούν για υπάρχουσες διαδικτυακές προσεγγίσεις προτάσεων. Επιπλέον, πρέπει να αξιολογηθεί η εξέλιξη των μετρήσεων με την πάροδο του χρόνου σε μια διαδοχική ρύθμιση. Τα σχόλια που ελήφθησαν από αυτές τις αξιολογήσεις θα μπορούσαν επίσης να αξιοποιηθούν για την προσαρμογή των συστάσεων σε πραγματικό χρόνο. Οι μελλοντικές εργασίες θα πρέπει να εξετάζουν λύσεις στους περιορισμούς που αντιμετωπίζει η προκαταρκτική αξιολόγηση και πρόσθετες μετρήσεις για την αξιολόγηση της διαδικτυακής σύστασης.

Συνδέσεις με άλλες περιοχές των συστάσεων. Θα μπορούσε να είναι ενδιαφέρον να διερευνήσουμε συνδέσεις μεταξύ διαδικτυακών προσαρμοστικών προτάσεων και άλλων τομέων της σύστασης. Συγκεκριμένα, μια πολλά υποσχόμενη κατεύθυνση θα εξετάσει το σχεδιασμό της διαδικτυακής προσαρμοστικής σύστασης πολλαπλού τομέα. Το RS μεταξύ τομέων είναι γνωστό για την ικανότητά του να εμπλουτίζει έναν τομέα-στόχο μεταφέροντας γνώσεις από βοηθητικούς τομείς. Οι γνώσεις που αποκτώνται σε πραγματικό χρόνο θα μπορούσαν στη συνέχεια να ωφελήσουν τον τομέα-στόχο και θα μπορούσαν να βοηθήσουν στην πρόβλεψη γεγονότων ή μετακινήσεων. Περαιτέρω πρόοδοι θα μπορούσαν να εξετάσουν το σχεδιασμό τέτοιων μεθόδων και την αξιολόγηση των επιδόσεών τους.

Συνολικά, το έργο που παρουσιάζεται σε αυτή τη διατριβή υπογραμμίζει τη σημασία της δυναμικής που υπάρχει στο οικοσύστημα της σύστασης και τα οφέλη από τη λήψη τους. Υπογραμμίζει επίσης ότι η ανάπτυξη συστημάτων συστάσεων σε πραγματικό περιβάλλον υπόκειται σε συγκεκριμένους περιορισμούς που πρέπει να προσδιορίζονται και να λαμβάνονται υπόψη κατά την ανάπτυξη νέων προσεγγίσεων. Παρά τις προόδους που έχουν σημειωθεί στον τομέα αυτό, το πρόβλημα της σύστασης εξακολουθεί να είναι πλούσιο σε προκλήσεις πραγματικού κόσμου, φέρνοντας συναρπαστικές νέες κατευθύνσεις για έρευνα.

Βιβλιογραφία

- [1] Share of e-commerce companies worldwide that will invest more in personalization in 2020. <https://www.statista.com/statistics/1174164/investing-personalization-e-commerce-companies-worldwide/>. (Ημερομηνία πρόσβασης 11 April 2021).
- [2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.
- [3] How personalised product recommendations can multiply your profit. <https://social.com/resources/how-personalised-product-recommendations-can-multiply-your-profit/>. (Ημερομηνία πρόσβασης 11 April 2021).
- [4] Alexander Tuzhilin, Gediminas Adomavicius. *Recommender Systems Handbook*. Springer, 2011 edition, 2010.
- [5] Recommender system. <https://group.accor.com/-/media/Corporate/Group/PDF-for-pages/Brand-book/Accor-Bands-And-Services-July-2020-def.pdf>.
- [6] Google maps directions and recommendations. <https://maps.google.com/>. (Ημερομηνία πρόσβασης 11 April 2021).
- [7] Food and more recommendation. <https://www.yelp.ie/dublin>. (Ημερομηνία πρόσβασης 11 April 2021).
- [8] Zagat. <https://www.zagat.com>. (Ημερομηνία πρόσβασης 11 April 2021).
- [9] Michelin guide - the official website. <https://guide.michelin.com/en>. (Ημερομηνία πρόσβασης 11 April 2021).
- [10] Netflix – watch tv programmes online, watch films online. <https://www.netflix.com>. (Ημερομηνία πρόσβασης 11 April 2021).
- [11] Marko Balabanović and Yoav Shoham. Fab: Content-based, collaborative recommendation. *Commun. ACM*, 40(3):66–72, March 1997.
- [12] X. Su. Collaborative filtering: A survey. In *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, pages 1–1, 2015.
- [13] Yue Shi, Martha Larson, and Alan Hanjalic. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. 47(1), 2014.
- [14] Context-aware recommender systems. 32:217–253, Oct. 2011.
- [15] Umberto Panniello, Alexander Tuzhilin, Michele Gorgoglione, Cosimo Palmisano, and Anto Pedone. Experimental comparison of pre- vs. post-filtering approaches in context-aware recommender systems. In *Proceedings of the Third ACM Conference on Recommender Systems, RecSys '09*, page 265–268, New York, NY, USA, 2009. Association for Computing Machinery.

- [16] H. Liu and H. Motoda. *Instance selection and construction for data mining*. Springer, 2013.
- [17] James Bennett, Charles Elkan, Bing Liu, Padhraic Smyth, and Domonkos Tikk. Kdd cup and workshop 2007. *SIGKDD Explor. Newsl.*, 9(2):51–52, December 2007.
- [18] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Byers. Big data: The next frontier for innovation, competition, and productivity, 05 2011.
- [19] Jeffrey Mervis. Agencies rally to tackle big data. *Science*, 336(6077):22–22, 2012.
- [20] Paul Resnick and Hal R. Varian. Recommender systems. *Commun. ACM*, 40(3):56–58, March 1997.
- [21] A. Felfernig, Sergiu Gordea, Dietmar Jannach, Erich Teppan, and Markus Zanker. A short survey of recommendation technologies in travel and tourism. *OGAI Journal (Oesterreichische Gesellschaft fuer Artificial Intelligence)*, 25:17–22, 01 2007.
- [22] Sarabjot Singh Anand and Bamshad Mobasher. *Contextual Recommendation*, page 142–160. Springer-Verlag, Berlin, Heidelberg, 2007.
- [23] Michael Schrage. Great digital companies build great recommendation engines. *Harvard Business Review*, 08 2017.
- [24] Carlos A. Gomez-Uribe and Neil Hunt. The netflix recommender system: Algorithms, business value, and innovation. 6(4), December 2016.
- [25] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '99*, page 230–237, New York, NY, USA, 1999. Association for Computing Machinery.
- [26] G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.
- [27] G. Semeraro, P. Lops, Pierpaolo Basile, and M. Degemmis. Knowledge infusion into content-based recommender systems. In *RecSys '09*, 2009.
- [28] R. S. Bhadoria, D. Sain, and R. Moriwai. Data mining algorithms for personalizing user' s profiles on web. *Technology and Electronics Engineering (IJCTEE)*, 7(1):120–125, 2011.
- [29] Boris Rousseau, Parisch Browne, Paul Malone, and Mícheál Ó Foghlu. Abstract user profiling for content personalisation in information retrieval. 01 2003.
- [30] Bo Xiao and Izak Benbasat. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly*, 31:137–209, 03 2007.
- [31] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW '94*, page 175–186, New York, NY, USA, 1994. Association for Computing Machinery.
- [32] Thomas Malone, Kenneth Grant, Franklyn Turbak, Stephen Brobst, and Michael Cohen. Intelligent information-sharing systems. *Commun. ACM*, 30:390–402, 05 1987.
- [33] Robin Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12, 11 2002.

- [34] Geoffrey I. Sammut, Claude; Webb. *Encyclopedia of Machine Learning*. Springer, 2010 edition, 2011.
- [35] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132, 2013.
- [36] Lalita Sharma and Anju Gera. A survey of recommendation system: Research challenges. *IJETT International Journal of Engineering Trends and Technology*, 4(3):1990–1992, may 2013.
- [37] Abowd Gregory Abowd Anind K. Dey. *Towards a Better Understanding of Context and Context-Awareness*. Springer, 2011 edition, 1999.
- [38] Yehuda Koren. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, page 426–434, New York, NY, USA, 2008. Association for Computing Machinery.
- [39] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. Recommender system application developments: A survey. *Decision Support Systems*, 74, 04 2015.
- [40] C. C. Aggarwal. *Neighborhood-Based Collaborative Filtering*. Springer International Publishing, 2016 edition, 2016.
- [41] C. C. Aggarwal. *An Introduction to Recommender Systems*. Springer, 2016.
- [42] Ben Schafer, Joseph Konstan, and John Riedl. E-commerce recommendation applications. 5, 08 2000.
- [43] Alexander Tuzhilin. Gediminas Adomavicius. *Recommender Systems Handbook*. Springer, 2011 edition, 2010.
- [44] M.D. Ekstrand, F.M. Harper, M.C. Willemsen, and J.A. Konstan. User perception of differences in recommender algorithms. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys'14), October 6–10, 2014, Silicon Valley, California*, pages 161–168, United States, 2014. Association for Computing Machinery, Inc.
- [45] C. C. Aggarwal. *Model-Based Collaborative Filtering*. Springer, 2016.
- [46] C. C. Aggarwal. *Data mining: the textbook*. Springer, 2016.
- [47] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*, pages 263–272, 2008.
- [48] Michael J. A. Berry Gordon S. Linoff. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons, 3rd edition, 2011.
- [49] C. Palmisano, A. Tuzhilin, and M. Gorgoglione. Using context to improve predictive modeling of customers in personalization applications. *IEEE Transactions on Knowledge and Data Engineering*, 20(11):1535–1549, 2008.
- [50] B. N. Schilit and M. M. Theimer. Disseminating active map information to mobile hosts. *IEEE Network*, 8(5):22–32, 1994.
- [51] P.J. Brown, J.D. Bovey, and Xian Chen. Context-aware applications: from the laboratory to the marketplace. *Personal Communications, IEEE [see also IEEE Wireless Communications]*, 4(5):58–64, 1997.

- [52] N. S. Ryan, J. Pascoe, and D. R. Morse. Enhanced reality fieldwork: the context-aware archaeological assistant. In V. Gaffney, M. van Leusen, and S. Exxon, editors, *Computer Applications in Archaeology 1997*, British Archaeological Reports, pages 182–196, Oxford, October 1998. Tempus Reparatum.
- [53] Anind Dey, Gregory Abowd, and Daniel Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction*, 16, 04 2001.
- [54] James Bettman, Mary Luce, and John Payne. Constructive consumer choice process. *Journal of Consumer Research*, 25:187–217, 12 1998.
- [55] Richard W. Olshavsky. Task complexity and contingent processing in decision making: A replication and extension. *Organizational Behavior and Human Performance*, 24(3):300–316, 1979.
- [56] C. K. Prahalad. Beyond crm: Ck prahalad predicts customer context is the next big thing. *American Management Association McWorld*, may 2004.
- [57] Gregory D. Abowd, Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton. Cyberguide: A mobile context-aware tour guide. 3(5):421–433, October 1997.
- [58] Federica Cena, Luca Console, Cristina Gena, Anna Goy, Guido Levi, Sonia Modeo, and Ilaria Torre. Integrating heterogeneous adaptation techniques to build a flexible and usable mobile tourist guide. *Ai Communications - AICOM*, 19, 01 0002.
- [59] Mark Setten, Stanislav Pokraev, and Johan Koolwaaij. Context-aware recommendations in the mobile tourist application compass. volume 3137, pages 235–244, 01 2004.
- [60] Gediminas Adomavicius, Ramesh Sankaranarayanan, Shahana Sen, and Alexander Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Trans. Inf. Syst.*, 23(1):103–145, January 2005.
- [61] Kenta Oku, Shinsuke Nakajima, J. Miyazaki, and Shunsuke Uemura. Context-aware svm for context-dependent information recommendation. pages 109 – 109, 06 2006.
- [62] Umberto Panniello, Alexander Tuzhilin, Michele Gorgoglione, Cosimo Palmisano, and Anto Pedone. Experimental comparison of pre- vs. post-filtering approaches in context-aware recommender systems. pages 265–268, 01 2009.
- [63] Zhiwen Yu, Xingshe Zhou, Daqing Zhang, Chung-Yau Chin, Xiaohang Wang, and Ji Men. Supporting context-aware media recommendations for smart phones. *Pervasive Computing, IEEE*, 5:68– 75, 08 2006.
- [64] Pedro G. Campos, Ignacio Fernández-Tobías, Iván Cantador, and Fernando Díez. Context-aware movie recommendations: An empirical comparison of pre-filtering, post-filtering and contextual modeling approaches. In Christian Huemer and Pasquale Lops, editors, *E-Commerce and Web Technologies*, pages 137–149, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [65] Joeran Beel and Stefan Langer. A comparison of offline evaluations, online evaluations, and user studies in the context of research-paper recommender systems. volume 9316, pages 153–168, 09 2015.
- [66] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. 22(1):5–53, January 2004.

- [67] Hyung Jun Ahn. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, 178:37–51, 01 2008.
- [68] John Hannon, Kevin McCarthy, and Barry Smyth. Finding useful users on twitter: Twittomender the followee recommender. pages 784–787, 04 2011.
- [69] Pablo Castells, Neil J. Hurley, and Saul Vargas. Novelty and diversity in recommender systems. In Francesco Ricci, Lior Rokach, and Bracha Shapira, editors, *Recommender Systems Handbook*, pages 881–918. Springer, 2015.
- [70] Zhenkun Wen and Tianrui Li, editors. *Knowledge Engineering and Management*. Springer Berlin Heidelberg, 2014.
- [71] Mohammed Wasid and Rashid Ali. An improved recommender system based on multi-criteria clustering approach. *Procedia Computer Science*, 131:93–101, 2018.
- [72] Guy Shani and Asela Gunawardana. Evaluating recommendation systems. In *Recommender Systems Handbook*, pages 257–297. Springer US, October 2010.
- [73] Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M. Henne. Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1):140–181, July 2008.
- [74] Florent Garcin and Boi Faltings. PEN recsys. In *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge on - NRS '13*. ACM Press, 2013.
- [75] Guy Shani, Ronen I. Brafman, and Solomon E. Shimony. Model-based online learning of POMDPs. In *Machine Learning: ECML 2005*, pages 353–364. Springer Berlin Heidelberg, 2005.
- [76] J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel. Lars: A location-aware recommender system. In *2012 IEEE 28th International Conference on Data Engineering*, pages 450–461, 2012.
- [77] Pedro Campos, Fernando Rubio, and Iván Cantador. Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24, 02 2014.
- [78] Hao Ma, Haixuan Yang, Michael Lyu, and Irwin King. Sorec: Social recommendation using probabilistic matrix factorization. pages 931–940, 01 2008.
- [79] J. Swarbrooke and S. Horner. *Consumer Behaviour in Tourism*. Butterworth-Heinemann, 2007.
- [80] Hannes Werthner, Aurkene Alzua-Sorzabal, Lorenzo Cantoni, Astrid Dickinger, Ulrike Gretzel, Dietmar Jannach, Julia Neidhardt, Birgit Pröll, Francesco Ricci, Miriam Scaglione, Brigitte Stangl, Oliviero Stock, and Markus Zanker. Future research issues in IT and tourism. *J. Inf. Technol. Tour.*, 15(1):1–15, 2015.
- [81] Hannes Werthner and Francesco Ricci. E-commerce and tourism. *Commun. ACM*, 47(12):101–105, December 2004.
- [82] Zheng Xiang, Vincent P. Magnini, and Daniel R. Fesenmaier. Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services*, 22(C):244–249, 2015.

- [83] Fuzheng Zhang, Nicholas Jing Yuan, Kai Zheng, Defu Lian, Xing Xie, and Yong Rui. Exploiting dining preference for restaurant recommendation. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, April 2016.
- [84] Daniel Herzog, Christopher Laß, and Wolfgang Wörndl. Tourrec: A tourist trip recommender system for individuals and groups. *RecSys '18*, page 496–497, New York, NY, USA, 2018. Association for Computing Machinery.
- [85] Michael Luca. Reviews, reputation, and revenue: The case of yelp.com. *SSRN Electronic Journal*, 2011.
- [86] Christoph Trattner, Alexander Oberegger, Lukas Eberhard, Denis Parra, and Leandro Balby Marinho. Understanding the impact of weather for poi recommendations. In *Proceedings of the ACM RecSys Workshop on Recommenders in Tourism (RecTour)*, 2016.
- [87] Maciej Kula. Metadata embeddings for user and item cold-start recommendations. In Toine Bogers and Marijn Koolen, editors, *Proceedings of the 2nd Workshop on New Trends on Content-Based Recommender Systems co-located with 9th ACM Conference on Recommender Systems (RecSys 2015), Vienna, Austria, September 16-20, 2015.*, volume 1448 of *CEUR Workshop Proceedings*, pages 14–21. CEUR-WS.org, 2015.
- [88] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. *UAI '09*, page 452–461, Arlington, Virginia, USA, 2009. AUAI Press.
- [89] Jason Weston, Hector Yee, and Ron J. Weiss. Learning to rank recommendations with the k-order statistic loss. In *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, page 245–248, New York, NY, USA, 2013. Association for Computing Machinery.

Παράρτημα Α

Ευρετήριο Ακρωνυμίων και Συντμήσεων

- CF:** Collaborative Filtering (Συνεργατική Διήθηση)
- RS:** Recommender Systems (Συστήματα Συστάσεων)
- CBF:** Content based Filtering (Βάσει Περιεχομένου)
- MF:** Matrix factorization (Παραγοντοποίηση Πίνακα)
- TF-IDF** term frequency-inverse document frequency
- Internet of Things** (IoT)
- SVM** Support Vector Machines (Μηχανές Διανυσμάτων Υποστήριξης)
- MB-CFRS** Model Based Collaborative Filtering Recommendation Systems (Συστήματα Συστάσεων Βάσει Μοντέλων)
- CARS** Context Aware Recommendation Systems (Συστήματα Συστάσεων Επίγνωσης Πλαισίου)
- MCRS** Multi Criteria Recommender Systems (Συστήματα Συστάσεων Πολλαπλών Κριτηρίων)
- MCDM** Multi Criteria Decision Making (Πρόβλημα Λήψης Αποφάσεων Πολλαπλών Κριτηρίων)
- MAE** Mean Absolute Error (Μέσο Απόλυτο Σφάλμα)
- RMSE** Root Mean Square Error (Τετραγωνικό Σφάλμα)
- CTR** Clickthrough rate (Κλικ προς Αριθμό Εμφανίσεων)
- BPR** Bayesian Personalized Ranking (Εξατομικευμένη κατάταξη Bayesian)
- ROC** Receiver Operating Characteristic (Χαρακτηριστικό Λειτουργίας Δέκτη)
- AUC** Area Under the Curve (Περιοχή Κάτω από την Καμπύλη)
- WRAP** Weighted Approximate-Rank Pairwise (Ζυγισμένη κατά προσέγγιση κατάταξη κατά ζεύγη)
- SVD** Singular Value Decomposition (Αποσύνθεση Μοναδικής Αξίας)
- FA** Factor Analysis (Ανάλυση Κύριων Συνιστωσών)
- PCA** Principal Component Analysis (Ανάλυση Κύριων Εξαρτημάτων)