



ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΕΘΝΙΚΟ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΕΠΙΧΕΙΡΙΣΙΑΚΗ ΕΡΕΥΝΑ

Διπλωματική Εργασία

Επισκόπηση Multi-Armed Bandit Αλγορίθμων Βελτιστοποίησης
εφαρμοσμένα σε Συστήματα Συστάσεων

Τσατσάνης Αθανάσιος

Αθήνα
Σεπτέμβρης 2022

Περιεχόμενα

Εισαγωγή	1
Συστήματα Συστάσεων	3
Ορισμοί και Έννοιες	3
Ο ρόλος των Συστημάτων Σύστασης	6
Εφαρμογές Συστημάτων Σύστασης	11
Άντληση Δεδομένων	12
Αντικείμενα	13
Χρήστες	14
Συναλλαγές	15
Μαθηματική Μοντελοποίηση	16
Κατηγορίες ΣΣ και Τεχνικές	18
Συστάσεις βάσει Περιεχομένου	21
Συνεργατικό Φιλτράρισμα	22
Υβριδικές τεχνικές	27
Δημογραφικό Φιλτράρισμα	28
Συστάσεις βάσει Γνώσης	28
Συστάσεις βάσει Κοινότητας	29
Συστάσεις βάσει Συμφραζόμενων	29
Προβλήματα και Περιορισμοί των Συστημάτων Σύστασης	30
Μελλοντικές Προκλήσεις	32
Αξιολόγηση των Συστημάτων Σύστασης	36
Μετρικές και Κριτήρια Αξιολόγησης	36
Πλαίσια Αξιολόγησης ΣΣ	43
Κατακλείδα	48
Αλγόριθμοι Multi-Armed Bandit	49
Εισαγωγή και Ορισμοί	49
Μαθηματικός Φορμαλισμός MAB	51
Παραλλαγές του Προβλήματος	55
Στρατηγικές MAB και Αλγόριθμοι	56

Βέλτιστες Στρατηγικές	56
Ασυμπτωτικές Στρατηγικές	57
Στρατηγικές MAB βάσει Περιεχομένου (Contextual bandit)	60
Στρατηγικές MAB σε Ανταγωνιστικό Περιβάλλον (Adversarial Bandit)	64
Άλλες επεκτάσεις του προβλήματος MAB	66
Μια ανάλυση και αξιολόγηση Αλγορίθμων MAB	69
Εφαρμογές Αλγορίθμων MAB	73
Κατακλείδα	78
MAB σε Συστήματα Συστάσεων	79
Αλγόριθμοι MAB στα Συστήματα Συστάσεων	79
Οι κατηγορίες MAB αλγορίθμων στα Συστήματα Συστάσεων	80
Αλγόριθμοι Πολλαπλών Επιλογών	80
Αλγόριθμοι Βελτιστοποίησης Πολλαπλών Στόχων	82
Ομαδοποίηση μέσω MAB Αλγορίθμων	84
Προσεγγίσεις Συνεργατικού Φιλτραρίσματος	86
MAB Πεπερασμένου Χρόνου	88
Συνδυαστικές Μέθοδοι	90
Αντιμετωπίζοντας τις προκλήσεις των Συστημάτων Σύστασης με αλγορίθμους MAB	92
Αραιότητα και Κλιμάκωση Αλγορίθμων	92
Αργή Εκκίνηση	93
Ασφάλεια	94
Διαφάνεια	95
Αξιολόγηση των MAB στο πεδίο των Συστημάτων Σύστασης	97
Κριτήρια Αξιολόγησης	100
MAB σε Συστήματα Συστάσεων στην Βιομηχανία	103
Συμπεράσματα, Κατευθύνσεις και Μελλοντική Έρευνα	108
Σχεδιασμός και μελέτη αλγορίθμων στο μέλλον	108
Κριτήρια Αξιολόγησης	110
Αντιμετώπιση των Προκλήσεων	111
Προτάσεις για την προέκταση της παρούσας διπλωματικής εργασίας	113
Βιβλιογραφία	130

Περίληψη

Η διπλωματική εργασία έχει στόχο την παρουσίαση και την μαθηματική ανάλυση διάφορων αλγορίθμων των *Multi-Armed Bandit (MAB)* προβλημάτων οι οποίοι εφαρμόζονται σε Συστήματα Συστάσεων (*Recommendation Systems*).

Τα προβλήματα MAB αφορούν την εύρεση μιας στρατηγικής επιλογών όπου εξαρχής η ωφέλεια της κάθε επιλογής είναι άγνωστη. Αναλόγως με το πρόβλημα και κατά την διάρκεια της στρατηγικής η ωφέλεια της κάθε επιλογής εκτιμάται, με τελικό στόχο την εύρεση της καλύτερης απόφασης και κατ' επέκταση την μεγιστοποίηση της ωφέλειας.

Τα Συστήματα Συστάσεων είναι συστήματα που στόχο έχουν να προτείνουν στον χρήστη κατάλληλα προϊόντα τα οποία θα εφαρμόζουν στις ανάγκες και προτιμήσεις του. Η γνώση και η χρήση των κριτηρίων σύστασης βασίζεται στην ιστορικότητα της δραστηριότητας του χρήστη μέσα στο σύστημα αλλά και στην ορθή συσχέτισή του με κατάλληλες κατηγορίες οι οποίες συνδέονται με προϊόντα.

Abstract

The goal of this thesis is to present and analyze mathematically various algorithms of Multi-Armed Bandit (MAB) problems that applied to Recommendation Systems (RecSys).

The MAB problems is about finding a selection policy that the reward of any choice is unknown. Considering the problem and during the policy process, the reward of these choices is estimated, so the optimal choice will be discovered and therefore the maximum reward of the process.

A Recommendation System's goal is to suggest various options or products to a user based on his needs and requirements. The knowledge and the usage of suggestion criteria are based on the past interactions of the user inside the system and also the correct user correlation with product categories.

Εισαγωγή

Ο στόχος της παρούσας διπλωματικής εργασίας είναι η παρουσίαση και η κατανόηση της αξίας των αλγορίθμων multi-armed bandit που εφαρμόζονται σε συστήματα συστάσεων. Για να επιτευχθεί αυτό είναι ανάγκη να αποσαφηνιστούν αρκετές έννοιες, τόσο των αλγορίθμων αυτών όσο και των συστημάτων σύστασης. Η διπλωματική εργασία χωρίζεται σε τρία βασικά κεφάλαια.

Το πρώτο κεφάλαιο αφορά τα Συστήματα Συστάσεων (ΣΣ). Δίνονται οι ορισμοί των διάφορων εννοιών που σχετίζονται με τα ΣΣ και γίνεται η παρουσίαση μιας μαθηματικής μοντελοποίησης που μπορεί να αναπαριστά αυτές τις έννοιες. Παρουσιάζονται επίσης οι εφαρμογές των ΣΣ στο τεχνολογικό και εμπορικό πεδίο καθώς και οι διάφορες κατηγορίες τους ανάλογα με την τεχνική προσέγγιση και το πεδίο εφαρμογής τους. Τέλος παρουσιάζονται τα διάφορα προβλήματα που υπάρχουν στις πραγματικές εφαρμογές τους, οι εκτιμώμενες προκλήσεις που θα κληθούν να αντιμετωπίσουν μελλοντικά και ο τρόπος αξιολόγησης της ποιότητας τους. Ο τρόπος αξιολόγησης παίζει πολύ σημαντικό ρόλο ώστε να κρίνεται η εκάστοτε βελτίωση, που επιχειρεί να επιλύσει προβλήματα.

Το δεύτερο κεφάλαιο είναι αφιερωμένο στο πρόβλημα multi-armed bandit και τους αλγορίθμους που καλούνται να το επιλύσουν. Στο κεφάλαιο αυτό δεν υπάρχει ακόμα σαφής συσχέτιση του προβλήματος MAB με τα συστήματα συστάσεων. Είναι αναγκαίο να γίνει ξεχωριστά μια παρουσίαση των εννοιών, των κατηγοριών και των εφαρμογών του γενικού προβλήματος ώστε να τεθούν τα θεμέλια για την κατανόηση της αξίας που προσφέρουν οι αλγόριθμοι MAB στα συστήματα συστάσεων.

Η σύνδεση μεταξύ των δύο (ΣΣ και MAB) πραγματοποιείται στο τρίτο κεφάλαιο. Εδώ γίνεται η αποσαφήνιση των κατηγοριών που έχουν αξία να εφαρμόζονται MAB αλγόριθμοι στα ΣΣ είτε στα πλαίσια βελτιώσεων είτε για την αντιμετώπιση σημαντικών προβλημάτων που έχουν. Τέλος αναφέρονται οι τρόποι που μπορεί να γίνεται η αξιολόγηση της αξίας ενός MAB αλγορίθμου εφαρμοσμένο σε ένα ΣΣ καθώς και η επί του πρακτέου εφαρμογή διάφορων παραδειγμάτων στην τωρινή βιομηχανία.

Η κατακλείδα της εργασίας είναι μια επισκόπηση της εργασίας και κυρίως αναφέρεται στις ενδεχόμενες μελλοντικές κατευθύνσεις που μπορεί να πάρει η έρευνα και ανάπτυξη των αλγορίθμων MAB πάνω σε Συστήματα Σύστασης.

Συστήματα Συστάσεων

Ορισμοί και Έννοιες

Συστήματα Συστάσεων (Recommender Systems – RS) είναι συστήματα λογισμικού που στόχο έχουν την υπόδειξη ή σύσταση *αντικειμένων* σε έναν *χρήστη* που έχουν *αξία* για αυτόν. Η ανάλυση των συστημάτων αυτών έχει στόχο οι συστάσεις να γίνονται με τρόπο χρήσιμο, εύστοχο και αποτελεσματικό. Η έννοια *Χρήστης* χρησιμοποιείται ως γενικός όρος, ο οποίος υποδηλώνει μια οντότητα φυσική ή τεχνητή (κάποιο άλλο λογισμικό) η οποία έχει την δυνατότητα να επιλέγει αντικείμενα. *Αντικείμενο* επίσης ορίζεται ως μια οποιαδήποτε οντότητα, η οποία διαθέτει χαρακτηριστικά τα οποία μπορούν να χρησιμοποιηθούν για την ταξινόμηση και αξιολόγηση του αντικειμένου από τον Χρήστη. Παράδειγμα ενός χρήστη μπορεί να είναι ένας πελάτης σε ένα κατάστημα, ένα σύνολο ανθρώπων, μια εταιρεία, ένα λογισμικό κτλ ενώ παραδείγματα αντικειμένων μπορεί να είναι ένα προϊόν, μια διαφήμιση, μια στρατηγική απόφαση, μια τιμή ενός συντελεστή κτλ.[97]

Η λογική που λειτουργεί κατά βάση ένα σύστημα συστάσεων είναι ότι, αφού ο χρήστης έρθει σε επαφή με τα αντικείμενα που του έχουν προταθεί, θα επικοινωνήσει την *αξία* τους βάσει των προτιμήσεων του δίνοντας κάποια μορφή αξιολόγησης (feedback), το οποίο θα έχει ως αποτέλεσμα το σύστημα να ενημερωθεί για τις προτιμήσεις του χρήστη αυτού. Στόχος είναι οι προτάσεις των αντικειμένων να προσαρμόζονται κατάλληλα στον συγκεκριμένο χρήστη (εξατομίκευση - personalization) και μέσω μιας επαναληπτικής διαδικασίας ο χρήστης να αξιολογεί όλο και περισσότερα αντικείμενα που του συστήνονται και με

όλο και πιο σαφή και αποδοτικό τρόπο. Επίσης ένα σύστημα συστάσεων μπορεί να έχει την δυνατότητα μέσω προϋπάρχουσας πληροφορίας (ιστορικό) των αντικειμένων και του χρήστη να κατασκευάζει συστάσεις που έχουν υψηλή πιθανότητα αποδοχής (θετικής αξιολόγησης) από τον χρήστη, χωρίς να έχει απαραίτητα ανάγκη το feedback του ίδιου. Για παράδειγμα μπορούν να χρησιμοποιηθούν δημογραφικές πληροφορίες του χρήστη (πχ ηλικία, φύλο) αλλά και πιο σύνθετες, όπως ο χρόνος αλληλεπίδρασης, η ώρα της ημέρας και ο τόπος, ώστε ο χρήστης να καταταχθεί σε συγκεκριμένες κατηγορίες οι οποίες μπορεί να συσχετίζονται ήδη με κάποια αντικείμενα βάσει επιλογών άλλων χρηστών του συστήματος στο παρελθόν. Σαφώς για να λειτουργήσει αποτελεσματικά ένα σύστημα συστάσεων αργά ή γρήγορα έχει ανάγκη από την διάδραση πολλών χρηστών, η οποία μπορεί να συμβεί είτε άμεσα (explicit feedback) είτε έμμεσα (implicit feedback).[41]

Ένας χρήστης μπορεί να αλληλεπιδράσει με το σύστημα άμεσα μέσω μιας βαθμολογίας (rating) που θα δώσει σε ένα αντικείμενο ή έμμεσα μέσω της καταγραφής της συμπεριφοράς του εντός του συστήματος, χωρίς ο ίδιος να δίνει κάποια ξεκάθαρη βαθμολογία. Ένα παράδειγμα έμμεσης διάδρασης είναι ο χρόνος που αλληλεπιδρά με ένα αντικείμενο ή η επιλογή ενός αντικειμένου από μόνη της. Το σημαντικό πλεονέκτημα της έμμεσης διάδρασης είναι ότι δεν χρειάζεται επιπλέον ενέργειες από τον χρήστη για να αξιολογηθεί ένα αντικείμενο. Παρόλα αυτά, πολλές φορές δεν είναι σαφής μια έμμεση αξιολόγηση και μπορεί να χάνεται μέσα στην τυχειότητα των διάφορων αλληλεπιδράσεων του χρήστη με το σύστημα (θόρυβος), κάνοντας την λιγότερο αξιόπιστη σε σχέση με την άμεση διάδραση στα πλαίσια της συμπερασματολογίας για τις προτιμήσεις του χρήστη. Αυτό κάνει αναγκαία την πιο εκλεπτυσμένη επεξεργασία των δεδομένων της έμμεσης διάδρασης, ώστε να αξιοποιηθούν πιο εύστοχα και κατάλληλα από τους αλγόριθμους του συστήματος συστάσεων.

Κλασικά παραδείγματα συστημάτων σύστασης είναι οι προτάσεις προϊόντων για αγορά σε ηλεκτρονικά καταστήματα, οι προτάσεις ταινιών και μουσικής από μια πλατφόρμα ψυχαγωγίας, τα εξατομικευμένα άρθρα που εμφανίζονται σε

δημοσιογραφικά sites κ.α.

Ένα σύστημα σύστασης εστιάζει λοιπόν κατά βάση σε μια συγκεκριμένη κατηγορία αντικειμένων (ταινίες, νέα κτλ) και σχεδιάζεται με τέτοιο τρόπο ώστε, το γραφικό του περιβάλλον και οι τεχνικές αλγορίθμων που χρησιμοποιεί, να δημιουργούν χρήσιμες και αξιοποιήσιμες συστάσεις των αντικειμένων αυτών για τον εκάστοτε χρήστη.

Τα τελευταία χρόνια το ενδιαφέρον και η χρήση των συστημάτων σύστασης έχει αυξηθεί δραματικά και αυτό αποδεικνύεται από τους παρακάτω λόγους:

- Παγκοσμίως επιτυχημένες εταιρείες όπως το Youtube, το IMBb και το Netflix χρησιμοποιούν συστήματα συστάσεων τα οποία μάλιστα αποτελούν και βασική υπηρεσία που παρέχουν στους χρήστες - πελάτες τους.
- Γίνονται συνέδρια αφιερωμένα στα συστήματα σύστασης (*ACM Recommender Systems*, *ACM SIGIR Special Interest Group on Information Retrieval (SIGIR)*, *Adaptation and Personalization (UMAP)*, *ACM's Special Interest Group on Management Of Data (SIGMOD)*).
- Οργανισμοί ανώτερης εκπαίδευσης παγκοσμίως παρέχουν μαθήματα μεταπτυχιακής αλλά και προπτυχιακής εκπαίδευσης πάνω στα συστήματα συστάσεων[53]
- Υπάρχουν αρκετά εξειδικευμένα άρθρα σε ακαδημαϊκά περιοδικά τα οποία αναφέρονται στην έρευνα και την ανάπτυξη του πεδίου αυτού (*AI Communications (2008)*, *IEEE Intelligent Systems (2007)*, *International Journal of Electronic Commerce (2006)*, *International Journal of Computer Science and Applications (2006)*, *ACM Transactions on Computer-Human Interaction (2005)*, *ACM Transactions on Information Systems (2004)*).

Ο ρόλος των Συστημάτων Σύστασης

Στην προηγούμενη υποενότητα αναφέρθηκε πως η σύσταση ενός η περισσότερων αντικειμένων σε έναν χρήστη πρέπει να είναι χρήσιμη και να έχει αξία. Εδώ είναι ανάγκη να διευκρινιστεί η διαφορά ανάμεσα στην αξία που μπορεί να έχει η σύσταση για τον χρήστη που την λαμβάνει και στην αξία που μπορεί να έχει για την υπηρεσία η οποία την παρέχει. Για παράδειγμα ένα σύστημα σύστασης εστιατορίων μιας περιοχής μπορεί να χρησιμοποιηθεί από ένα site έκδοσης αεροπορικών εισιτηρίων με στόχο την αύξηση των πωλήσεων των εισιτηρίων ή από μια εταιρεία τουριστικής διαχείρισης με στόχο την αύξηση του τουρισμού σε μια συγκεκριμένη περιοχή. Από την άλλη πλευρά, το κίνητρο ενός χρήστη που θέλει να χρησιμοποιήσει το ίδιο σύστημα μπορεί να είναι για να εντοπίσει εστιατόρια που ανταποκρίνονται στις ανάγκες του από την στιγμή που σκοπεύει να επισκεφτεί μια περιοχή ούτως ή άλλως.

Παρακάτω παρουσιάζονται διάφοροι λόγοι για τους οποίους ένα σύστημα συστάσεων είναι χρήσιμο για τις υπηρεσίες που το χρησιμοποιούν:

- *Αύξηση κατανάλωσης προϊόντων:* Αυτός είναι πιθανόν ο πιο σημαντικός λόγος που χρησιμοποιούνται τα συστήματα συστάσεων στο εμπόριο. Για παράδειγμα μπορεί να αυξάνει την πιθανότητα να πουληθούν επιπλέον αντικείμενα κατά την διάρκεια μιας αγοράς ενός πελάτη σε σχέση με μια αγορά χωρίς συστάσεις προϊόντων. Κάτι τέτοιο επιτυγχάνεται καθώς τα προϊόντα που συστήνονται πιθανότατα ανταποκρίνονται στις ανάγκες του καταναλωτή. Σε υπηρεσίες που ο χρήστης δεν έχει κόστος για την επιλογή του μπορεί επίσης να αξιοποιηθεί αυτή η δυνατότητα. Για παράδειγμα, σε ένα δημοσιογραφικό site, ένα σύστημα συστάσεων μπορεί να αυξήσει τις αναγνώσεις των άρθρων συνολικά μέσω συστάσεων. Επομένως η αξία που προσφέρει ένα σύστημα σύστασης από την οπτική μιας υπηρεσίας είναι ότι αυξάνει την αλληλεπίδραση των χρηστών με την υπηρεσία σε σχέση με την αλληλεπίδραση που έχουν μέσω της απλής πλοήγησης τους εντός των πληροφοριών που παρέχει η υπηρεσία.

- *Κατανάλωση ποικιλίας αντικειμένων:* Επίσης μια πολύ σημαντική αξία που προσφέρεται είναι ότι δίνεται η δυνατότητα στους χρήστες να έρθουν σε επαφή με αντικείμενα που θα ήταν εξαρχής λιγότερο πιθανό να έρθουν. Για παράδειγμα, εντός των εκατοντάδων ταινιών που μπορεί να υπάρχουν σε μια πλατφόρμα, ο χρήστης ενδέχεται να εντοπίσει πολύ πιο εύκολα ταινίες που δεν είναι δημοφιλείς αλλά που είναι πολύ πιθανό να τον ενδιαφέρουν.
- *Αύξηση της ικανοποίησης των χρηστών:* Σαφώς ένα καλά σχεδιασμένο σύστημα σύστασης βελτιώνει σημαντικά την εμπειρία που έχει ο χρήστης με την υπηρεσία. Σε συνδυασμό με ένα κατάλληλα σχεδιασμένο γραφικό περιβάλλον, ο χρήστης θα βρίσκει με μεγάλη πιθανότητα αντικείμενα που τον ικανοποιούν, αφού οι προτάσεις και κατ' επέκταση οι επιλογές των αντικειμένων θα σχετίζονται άμεσα με τα ενδιαφέροντα και τις ανάγκες του.
- *Αύξηση της αφοσίωσης των πελατών:* Ένας χρήστης είναι λογικό να είναι πιστός απέναντι σε μια υπηρεσία όταν εκείνη τον αναγνωρίζει ως παλαιότερο χρήστη και του παρέχει πλεονεκτήματα λόγω της παλαιότητας και αφοσίωσης του. Τα συστήματα συστάσεων εξορισμού ανταμείβουν τους χρήστες οι οποίοι χρησιμοποιούν την υπηρεσία επαναλαμβανόμενα διότι προσαρμόζονται όλο και περισσότερο στις ανάγκες του χρήστη αυτού.
- *Καλύτερη κατανόηση των αναγκών των χρηστών:* Μια πολύ σημαντική λειτουργικότητα των συστημάτων σύστασης είναι ότι δημιουργούν μια περιγραφή των προτιμήσεων των χρηστών η οποία δημιουργείται μέσω των δεδομένων που συλλέγονται από την αλληλεπίδραση των χρηστών με το σύστημα, είτε άμεσα, είτε έμμεσα. Ο πάροχος του συστήματος μπορεί να επαναχρησιμοποιεί αυτήν περιγραφή προτιμήσεων με στόχο την βελτίωση της διαχείρισης των αντικειμένων της υπηρεσίας, όπως την παραγωγή ή την αποθήκευση προϊόντων. Για παράδειγμα μέσω της συλλογής των προτιμήσεων των πελατών ένα κατάστημα μπορεί να συμπεράνει πως χρειάζεται να μειώσει τον αριθμό συγκεκριμένων προϊόντων τα οποία φαίνεται ότι δεν είναι υψηλά στις προτιμήσεις των περισσότερων πελατών του.

Τα παραπάνω πλεονεκτήματα σχετίζονται περισσότερο με την οπτική των υπηρεσιών που χρησιμοποιούν συστήματα συστάσεων όμως υπάρχουν ξεκάθαρα πλεονεκτήματα και από την οπτική των χρηστών. Κατά συνέπεια ένα σύστημα συστάσεων πρέπει να ισορροπεί ανάμεσα στις ανάγκες της υπηρεσίας και των χρηστών έτσι ώστε να έχει αξία και για τις δύο πλευρές.

Στην συνέχεια παρουσιάζονται κάποιες στοχοθεσίες που μπορεί να έχει ένα σύστημα σύστασης έτσι ώστε να παρέχει αξία στους χρήστες:

- *Εύρεση κάποιων ποιοτικών αντικειμένων:* Η σύσταση μιας διατεταγμένης λίστας αντικειμένων σε έναν χρήστη όπου η διάταξη βασίζεται στην πρόβλεψη του πόση αξία έχει το κάθε αντικείμενο για αυτόν τον χρήστη.
- *Εύρεση όλων των ποιοτικών αντικειμένων:* Σύσταση όλων των αντικειμένων που μπορούν να ικανοποιήσουν κάποιες από τις ανάγκες του χρήστη. Υπάρχουν περιπτώσεις που δεν είναι αρκετή η σύσταση μερικών αντικειμένων όπως για παράδειγμα όταν τα αντικείμενα είναι ελάχιστα ή η επιλογή των αντικειμένων αφορά κρίσιμες εφαρμογές (πχ σε ιατρικές και οικονομικές εφαρμογές). Σε τέτοιες περιπτώσεις, πέρα από το πλεονέκτημα του να μελετηθούν προσεκτικά όλες οι επιλογές των προτεινόμενων αντικειμένων, ο χρήστης μπορεί να βρει αξία από το σύστημα σύστασης μέσω της ταξινόμησης που παρουσιάζει για τα αντικείμενα και ενδεχομένως από την επεξήγηση που δίνει το σύστημα για την διάταξη τους αυτή.
- *Επισήμανση εντός συγκεκριμένου πλαισίου:* Δεδομένου ενός πλαισίου, πχ κορυφαία σε πωλήσεις βιβλία, το σύστημα μπορεί να δώσει έμφαση σε κάποια αντικείμενα βασιζόμενο στις μακροπρόθεσμες προτιμήσεις του χρήστη. Για παράδειγμα σε ένα ηλεκτρονικό βιβλιοπωλείο το σύστημα σύστασης μπορεί να επισημάνει ποια βιβλία αξίζουν την προσοχή του χρήστη ανάμεσα στα 10 κορυφαία σε πωλήσεις βιβλία που παρουσιάζονται στην αρχική σελίδα του καταστήματος.
- *Σύσταση ακολουθίας αντικειμένων:* Το σύστημα μπορεί αντί να εστιάζει στην εύρεση του πιο κατάλληλου αντικειμένου για τον χρήστη και να παραθέτει

αντικείμενα με σειρά καταλληλότητας, μπορεί να παρουσιάζει μια λίστα από αντικείμενα που δίνουν αξία στον χρήστη ως σύνολο (πχ μια συλλογή από τραγούδια που ανήκουν στο ίδιο μουσικό άλμπουμ).

- *Σύσταση πακέτων αντικειμένων*: Εδώ το σύστημα μπορεί να προτείνει αντικείμενα που ταιριάζουν καλά μεταξύ τους ως συνδυασμός. Για παράδειγμα ένα ταξιδιωτικό πλάνο μπορεί να συνθέτει διάφορους προορισμούς και τουριστικές υπηρεσίες που υπάρχουν εντός μιας συγκεκριμένης περιοχής. Από την οπτική του χρήστη όλες αυτές οι επιλογές μπορούν να επιλεχθούν από τον χρήστη ως μία επιλογή προορισμού.
- *Αναζήτηση πληροφοριών*: Για τις περιπτώσεις που ο χρήστης πλοηγείται εντός κάποιου καταλόγου χωρίς πρόθεση για κάποια αγορά, το σύστημα συστάσεων μπορεί να τον βοηθήσει έτσι ώστε τα αντικείμενα που του παρουσιάζονται να είναι όσο το δυνατόν πιο σχετικά με τα ενδιαφέροντα του στα πλαίσια της κατηγορίας αντικειμένων που αναζητά. Για παράδειγμα αναζήτηση ταινιών με βάση το είδος τους ή τον σκηνοθέτη.
- *Αναζήτηση αξιόπιστου συστήματος συστάσεων*: Κάποιοι χρήστες δεν εμπιστεύονται τα συστήματα συστάσεων και γι' αυτό τον λόγο παίζουν με τις δυνατότητες τους έτσι ώστε να καταλάβουν πόσο επιτυχημένα μπορούν να κάνουν συστάσεις. Εξαιτίας αυτού, κάποια συστήματα συστάσεων προσφέρουν εργαλεία σε έναν χρήστη εξ αρχής, ώστε να δοκιμάσει τις δυνατότητες τους.
- *Βελτίωση προφίλ*: Αφορά την δυνατότητα του χρήστη να παρέχει πληροφορίες στο σύστημα συστάσεων που αφορούν τις προτιμήσεις του. Αυτή είναι μια βασική προϋπόθεση έτσι ώστε το σύστημα να μπορεί να παρέχει εξατομικευμένες συστάσεις πίσω στον χρήστη. Χωρίς αυτήν την δυνατότητα το σύστημα το μόνο που θα μπορούσε να προτείνει είναι αντικείμενα που αφορούν κατά βάση τον «μέσο» χρήστη της υπηρεσίας.
- *Έκφραση χρηστών*: Κάποιοι χρήστες δεν ενδιαφέρονται για τις συστάσεις του συστήματος αλλά προτιμούν να παρέχουν τις πληροφορίες προτιμήσεων τους

και να συνεισφέρουν στο σύστημα ενώ ταυτόχρονα εκφράζουν τα πιστεύω και τις απόψεις τους. Η ικανοποίηση των χρηστών στα πλαίσια αυτής έκφρασης μπορεί κάλλιστα να διατηρήσει την σχέση που έχουν οι χρήστες αυτοί με την υπηρεσία.

- *Βοήθεια άλλων χρηστών:* Αντίστοιχα με την προηγούμενη περίπτωση πολλοί χρήστες ενδιαφέρονται να παρέχουν πληροφορίες στο σύστημα με στόχο την ενίσχυση της κοινότητας που αξιοποιεί την εκάστοτε υπηρεσία. Κάτι τέτοιο μπορεί να είναι πολύ ισχυρό κίνητρο έτσι ώστε χρήστες να καταχωρούν τις προτιμήσεις τους στο σύστημα. Για παράδειγμα, σε ένα σύστημα συστάσεων μιας ιστοσελίδας αγοραπωλησιών μεταχειρισμένων συσκευών, ένας χρήστης μπορεί να καταχωρήσει μια αξιολόγηση έτσι ώστε να ενημερώσει τους άλλους χρήστες για κάποιο προϊόν, αγοραστή ή πωλητή και όχι γιατί σκοπεύει να επιλέξει κάποιο παρόμοιο αντικείμενο στο μέλλον και χρειάζεται εξατομικευμένες συστάσεις.
- *Επιρροή άλλων χρηστών:* Αν και αυτή η περίπτωση δεν συγκαταλέγεται στην εκ προθέσεως αλληλεπίδραση με τους χρήστες, πολλές φορές οι χρήστες στοχεύουν στην αλλοίωση του συστήματος συστάσεων έτσι ώστε να προωθήσουν ή να δυσφημίσουν κάποιο αντικείμενο, με στόχο να επηρεάσουν τους υπόλοιπους χρήστες στις επιλογές τους. Η συγκεκριμένη περίπτωση σαφώς υποδεικνύει ως στόχο του συστήματος την αποτροπή του παραπάνω φαινομένου έτσι ώστε οι συστάσεις να παραμένουν ποιοτικές και αξιόπιστες για όλους τους χρήστες.

Όπως φάνηκε παραπάνω ο ρόλος ενός συστήματος συστάσεων από την οπτική των χρηστών μπορεί να είναι αρκετά ποικιλόμορφος. Αυτή η ποικιλομορφία επιτάσσει την αξιοποίηση διάφορων πηγών πληροφόρησης και τεχνικών οι οποίες θα περιγραφούν σε επόμενη υποενότητα.

Εφαρμογές Συστημάτων Σύστασης

Η έρευνα για τα συστήματα συστάσεων διεξάγεται με περισσότερη έμφαση σε εμπορικές εφαρμογές καθώς πέρα από την θεωρητική μελέτη, αφιερώνεται μεγάλο μέρος στην βελτίωση των ήδη εφαρμοσμένων εμπορικών ΣΣ. Η βελτίωση αυτή αφορά όλα τα διαφορετικά στάδια της ζωής ενός ΣΣ όπως ο σχεδιασμός του συστήματος, η εφαρμογή, η συντήρηση και η ενίσχυση του κατά την διάρκεια της λειτουργίας του. Κύριο σημείο της έρευνας αφορά το στάδιο του σχεδιασμού ενός ΣΣ και συμπεριλαμβάνει την μελέτη των παραγόντων που παίζουν σημαντικό ρόλο για την επιλογή των αλγορίθμων που εν τέλει διέπουν το ΣΣ. Πρωταρχικός παράγοντας είναι το πεδίο πάνω στο οποίο θα εφαρμοστεί το ΣΣ. Παρακάτω παρουσιάζονται οι βασικά πεδία πάνω βάσει των οποίων γίνεται και η ταξινόμηση των ΣΣ:

- *Ψυχαγωγία* – Σε αυτό το πεδίο συγκαταλέγονται υπηρεσίες που αφορούν την κατανάλωση αντικειμένων από τους χρήστες για ψυχαγωγικούς σκοπούς όπως ταινίες, μουσική κτλ.
- *Ενημέρωση* – Αυτό το πεδίο περιλαμβάνει υπηρεσίες εξατομικευμένης ενημέρωσης όπως άρθρα και νέα γεγονότα αλλά και υπηρεσίες εκμάθησης όπως ιστοσελίδες e-learning.
- *Ηλεκτρονικό εμπόριο* – Εδώ συμπεριλαμβάνονται εμπορικές υπηρεσίες που πουλάνε και διαφημίζουν προϊόντα.
- *Αναζήτηση πληροφοριών* – Αφορά ταξιδιωτικές υπηρεσίες, υπηρεσίες εύρεσης σπιτιού ή αναζήτηση εργασίας
- *Κοινωνική διασύνδεση* – Αφορά την αναζήτηση φίλων, συντρόφων και συνεργατών μέσω διαδικτύου.

Ο σχεδιασμός επομένως ενός ΣΣ είναι αναγκαίο να λαμβάνει υπόψιν το πεδίο εφαρμογής του συστήματος διότι στο κάθε πεδίο υπάρχουν διαφορετικοί περιορισμοί και ανάγκες που αφορούν τους χρήστες και τα αντικείμενα που εμπλέκονται.

Παρακάτω έχουμε μερικά παραδείγματα διάσμηων εταιριών ανά πεδίο οι οποίες χρησιμοποιούν ΣΣ:

Ταινίες	Netflix, IMDb
Βιβλία	Amazon
Νέα	Google News
Βίντεο	Youtube
Μουσική	Spotify
Εστιατόρια	Yelp
Τουρισμός	Tripadvisor
Δωμάτια	Airbnb
Διαφημίσεις	Google Search
Διασύνδεση	Facebook

Κάθε ένα ΣΣ των παραπάνω διάσμηων εταιρειών διαθέτει τα δικά του χαρακτηριστικά, τεχνικές, αλγορίθμους και στόχους που ανταποκρίνονται κατάλληλα στο πεδίο πάνω στο οποίο εφαρμόζονται. Για παράδειγμα στο πεδίο της μουσικής ένα ΣΣ μπορεί να αρκείται στην άμεση αξιολόγηση των τραγουδιών από τους χρήστες ενώ στο πεδίο του τουρισμού να αξιοποιούνται επιπλέον δημογραφικά στοιχεία για την σύσταση αντικειμένων.

Άντληση Δεδομένων

Η πηγές και οι μορφές των δεδομένων που μπορεί να αξιοποιήσει ένα ΣΣ μπορούν να διαφέρουν ποικιλοτρόπως και η δυνατότητα αξιοποίησης τους βασίζεται στην τεχνική που διέπει το ΣΣ. Υπάρχουν τεχνικές ΣΣ όπου αξιοποιούν μόνο απλά δεδομένα, όπως αξιολογήσεις χρηστών για αντικείμενα, ενώ άλλες που βασίζονται εξολοκλήρου σε πολυσύνδετα δεδομένα, όπως οντολογικές περιγραφές χρηστών και αντικειμένων, περιορισμούς που θέτουν οι χρήστες, κοινωνικές συσχετίσεις ή δραστηριότητες των χρηστών μεταξύ τους. Σε κάθε περίπτωση, ως γενική ταξινόμηση τα δεδομένα που αξιοποιούνται από ΣΣ αφορούν τρεις οντότητες: Τα αντικείμενα, τους χρήστες και τις συναλλαγές.

Αντικείμενα

Τα αντικείμενα είναι οι οντότητες οι οποίες συστήνονται από το ΣΣ και χαρακτηρίζονται από την πολυπλοκότητα και την αξία τους. Η αξία ενός αντικειμένου μπορεί να είναι θετική εάν το αντικείμενο κριθεί πως είναι χρήσιμο για τον χρήστη ή αρνητική εάν θεωρηθεί ακατάλληλο για αυτόν. Θεωρείται πάντα πως η επιλογή ενός αντικειμένου από έναν χρήστη συνοδεύεται από ένα κόστος, το οποίο συμπεριλαμβάνει την νοητική κούραση του χρήστη η οποία προκύπτει από την αναζήτηση του αντικειμένου και το οικονομικό αντίτιμο που δίνει ο χρήστης για την αγορά του αντικειμένου.

Για παράδειγμα, ένας σχεδιαστής ΣΣ για σύσταση άρθρων θα πρέπει να λάβει υπόψιν του την πολυπλοκότητα του αντικειμένου «Άρθρο» (δομή, αναπαράσταση τίτλου, επικαιρότητα κτλ) ενώ ταυτόχρονα θα πρέπει να αντιλαμβάνεται ότι ακόμα και εάν δεν πληρώνει ο χρήστης για το αντικείμενο υπάρχει ένα νοητικό κόστος στην αναζήτηση και ανάγνωση του. Εάν το προτεινόμενο από το ΣΣ άρθρο, είναι χρήσιμο για τον χρήστη τότε το κόστος του αντικειμένου υπερκαλύπτεται από το όφελος της απόκτησης μια νέας χρήσιμης πληροφορίας ενώ, εάν το αντικείμενο είναι αδιάφορο για τον χρήστη, η αξία του αντικειμένου θεωρείται αρνητική. Σε άλλα πεδία (πχ αναζήτηση για αγορά αυτοκινήτου ή μετοχών) το οικονομικό κόστος των αντικειμένων έχει μεγαλύτερη βαρύτητα σε σχέση με τον νοητικό κόστος της αναζήτησης του και επομένως ο σχεδιασμός ενός ΣΣ θα πρέπει να επιλέξει τις κατάλληλες προσεγγίσεις

Αντικείμενα με μικρή πολυπλοκότητα θεωρούνται τα άρθρα, οι ιστοσελίδες, οι ταινίες κτλ ενώ με μεγαλύτερη πολυπλοκότητα τα ηλεκτρονικά είδη όπως κινητό, PC κτλ. Αντικείμενα με την μέγιστη πολυπλοκότητα θεωρούνται οι ασφάλειες, οι μετοχές, οι ταξιδιωτικοί προορισμοί, η αναζήτηση εργασίας κτλ.

Τα ΣΣ, σύμφωνα με την βασική τεχνολογία τους, μπορούν να χρησιμοποιήσουν ένα εύρος χαρακτηριστικών και ιδιοτήτων των αντικειμένων. Για παράδειγμα, σε ένα ΣΣ ταινιών, το είδος, η χρονολογία, ο σκηνοθέτης κτλ μπορούν να χρησιμοποιηθούν για να περιγράψουν το αντικείμενο «Ταινία» και η περιγραφή

αυτή χρησιμοποιείται για να μεταφραστεί σε αξία-χρησιμότητα για τον κάθε χρήστη. Τα αντικείμενα μπορούν να αναπαρασταθούν μέσω διάφορων αναπαραστατικών προσεγγίσεων όπως μέσω ενός απλού κωδικού ή μέσω του συνόλου των χαρακτηριστικών τους ή ακόμα πιο σύνθετα μέσω της συσχέτισής τους με μια οντολογική αναπαράσταση του πεδίου (κατηγοριοποίηση και ιεραρχία).

Χρήστες

Οι χρήστες ενός ΣΣ μπορούν να διαφέρουν μεταξύ τους στους στόχους και τα χαρακτηριστικά τους. Δεδομένου του στόχου της εξατομικευμένης σύστασης, ένα ΣΣ μπορεί να αξιοποιήσει ένα μεγάλο εύρος των χαρακτηριστικών αυτών. Η πληροφορία αυτή μπορεί να δομηθεί ποικιλοτρόπως και όπως αντίστοιχα με τα αντικείμενα η κατάλληλη δομή εξαρτάται από την τεχνική προσέγγιση του ΣΣ. Για παράδειγμα, στις τεχνικές *συνεργατικού φιλτραρίσματος*, οι χρήστες μπορούν να μοντελοποιηθούν μέσω μιας λίστας η οποία διαθέτει τις αξιολογήσεις του χρήστη για κάποια αντικείμενα. Σε ΣΣ με *δημογραφική προσέγγιση* ένας χρήστης μπορεί να μοντελοποιηθεί μέσω των δημογραφικών χαρακτηριστικών του (ηλικία, φύλο, επάγγελμα κτλ).

Επομένως τα δεδομένα των χρηστών μαζί με την δομή τους αποτελούν το μοντέλο χρηστών, το οποίο με την σειρά του χρησιμοποιείται για να περιγράψει έναν χρήστη (πχ προτιμήσεις και ανάγκες). Η κάθε τεχνική ΣΣ που στοχεύει στην εξατομικευμένη σύσταση είναι απαραίτητο να αξιοποιήσει το κατάλληλο μοντέλο χρηστών το οποίο παίζει κεντρικό ρόλο για την επιτυχημένη σύσταση αντικειμένων.

Οι χρήστες μπορούν να περιγραφούν επίσης και από τα πρότυπα συμπεριφοράς τους μέσα στο σύστημα καθώς και από τις συσχετίσεις που έχουν με άλλους χρήστες. Ένα ΣΣ μπορεί να αξιοποιήσει τέτοιου είδους πληροφορίες για να προτείνει αντικείμενα που έχουν προτιμηθεί από παρόμοιους χρήστες με τον χρήστη που λαμβάνει την σύσταση.

Συναλλαγές

Συναλλαγή θεωρείται η καταγεγραμμένη αλληλεπίδραση ενός χρήστη με το ΣΣ. Οι συναλλαγές είναι δεδομένα που δημιουργούνται και αποθηκεύονται κατά την διάρκεια της επικοινωνίας ανθρώπου-μηχανής και είναι χρήσιμα για τους αλγορίθμους σύστασης ενός ΣΣ. Για παράδειγμα, ένα δεδομένο συναλλαγής μπορεί να περιέχει μια αναφορά στο αντικείμενο που έχει επιλέξει ένας χρήστης μαζί με τις λέξεις κλειδιά που χρησιμοποίησε ο ίδιος για να το εντοπίσει. Εάν είναι διαθέσιμη, μια συναλλαγή μπορεί να εμπεριέχει και την βαθμολογία του χρήστη προς το αντικείμενο που επέλεξε. Οι βαθμολογίες είναι τα πιο δημοφιλή δεδομένα συναλλαγών των ΣΣ και μπορούν να συλλεχτούν είτε άμεσα είτε έμμεσα όπως έχει αναφερθεί σε προηγούμενη υποενότητα. Στην άμεση συλλογή των βαθμολογήσεων, το σύστημα ζητάει από τους χρήστες να δώσουν τον βαθμό ικανοποίησης τους για τα αντικείμενα που αλληλεπιδρούν. Η βαθμολόγηση μπορεί να έχει τις παρακάτω μορφές:

- *Αριθμητική βαθμολόγηση*, όπου υπάρχει μια κλίμακα (πχ 1 έως 5 αστέρια) και ο χρήστης επιλέγει τον αριθμό που κρίνει ότι είναι πιο αντιπροσωπευτικός για τον ίδιο.
- *Κατηγορική βαθμολόγηση*, όπου υπάρχει μια λίστα επιλογών (πχ «συμφωνώ», «αδιάφορο», «διαφωνώ») και ο χρήστης επιλέγει συνήθως μέσω κάποιου ερωτηματολογίου. Εδώ δεν είναι απαραίτητο να υπάρχει διάταξη μεταξύ των επιλογών όπως στην αριθμητική βαθμολόγηση.
- *Διαδική βαθμολόγηση*, όπου είναι η πιο απλή μορφή της αριθμητικής βαθμολόγησης και ο χρήστης επιλέγει «Ναι» ή «Όχι» υποδηλώνοντας εάν η επιλογή του αντικειμένου ήταν μια καλή για εκείνον επιλογή ή όχι.
- *Ενιαία βαθμολόγηση*, όπου ο χρήστης επιλέγει εάν του αρέσει ένα αντικείμενο αλλά δεν υπάρχει επιλογή για την δυσαρέσκεια. Εδώ η απουσία βαθμολόγησης θεωρείται ως μη ύπαρξη πληροφορίας σχετικά με τον χρήστη και το αντικείμενο.

Μαθηματική Μοντελοποίηση

Αναπαριστούμε τα αντικείμενα του συστήματος ως ένα σύνολο S_M από M διαφορετικά αντικείμενα και τους χρήστες ως ένα σύνολο S_N με N διαφορετικούς χρήστες.

$$S_N = \{u_1, \dots, u_N\}$$

$$S_M = \{i_1, \dots, i_M\}$$

Οι βαθμολογήσεις των χρηστών για τα αντικείμενα αναπαριστώνται μέσω ενός πίνακα R μεγέθους $N \times M$ με στοιχεία $r_{ui} = \langle u, i \rangle$, $u \in S_N$, $i \in S_M$ τα οποία είναι η βαθμολόγηση του χρήστη u για το αντικείμενο i . Το r_{ui} μπορεί να έχει τιμές βάσει του τύπου βαθμολόγησης που διαθέτει το εκάστοτε σύστημα (πχ για αριθμητική βαθμολόγηση τεσσάρων επιπέδων $r_{ui} \in \{1, 2, 3, 4\}$ ενώ για δυαδική $r_{ui} \in \{-1, 1\}$ ή $\{0, 1\}$ κτλ)

Δεδομένου ότι ο πίνακας R δεν είναι συνήθως συμπληρωμένος ορίζουμε το σύνολο S το οποίο αποτελείται από τα στοιχεία r_{ui} του πίνακα που είναι συμπληρωμένα.

$I(u)$ είναι το σύνολο των αντικειμένων που ο χρήστης u έχει δώσει βαθμολόγηση, δηλαδή $I(u) = \{i: (u, i) \in S\}$

$J(i)$ είναι το σύνολο των χρηστών που έχουν δώσει βαθμολόγηση για το αντικείμενο i , δηλαδή $J(i) = \{u: (u, i) \in S\}$

Παράδειγμα εικόνας για $N=3$ και $M=5$ σε αριθμητική βαθμολόγηση τεσσάρων επιπέδων 1-4

		Αντικείμενα				
		i_1	i_2	i_3	i_4	i_5
Χρήστες	u_1	1		3		
	u_2		4		2	1
	u_3			1		

Τα αντίστοιχα σύνολα είναι:

$$S = \{r_{11}, r_{13}, r_{22}, r_{24}, r_{25}, r_{33}\}, I(u_1) = \{i_1, i_3\}, J(i_3) = \{u_1, u_3\} \text{ κτλ}$$

Τα r_{ui} που δεν είναι συμπληρωμένα αναπαριστούν τις μη βαθμολογήσεις (έλλειψη πληροφορίας). Όπως αναφέρθηκε και πριν οι πίνακες R είναι συνήθως αραιοί πίνακες καθώς το πιο πιθανό είναι οι χρήστες να έχουν αλληλεπιδράσει με λίγα αντικείμενα στο σύστημα ενώ το πλήθος αντικειμένων και χρηστών στα πραγματικά συστήματα τείνει είναι πολύ μεγάλο.

Ορίζουμε επίσης την συνάρτηση ωφελείας f πάνω στο σύνολο L_u των αντικειμένων που μπορούν να συσταθούν σε έναν χρήστη u . Στόχος της f είναι να προβλέψει πόση αξία έχει ένα αντικείμενο i για τον χρήστη u .

Στην συνέχεια παρουσιάζεται η απλούστερη και η πιο βασική μορφή του αλγορίθμου $\Sigma\Sigma$ βάσει του οποίου γίνονται οι συστάσεις σε έναν χρήστη u :

Αλγόριθμος Σύστασης

Δεδομένα: χρήστης u , σύνολο L_u , μέγεθος της λίστας σύστασης s

Βήμα 1: Πρόβλεψη της $f(u,i)$ για κάθε $I \in L_u$

Βήμα 2: Δημιουργία λίστας L'_u από την επιλογή των πρώτων s αντικειμένων με την υψηλότερη ωφέλεια $f(u,i)$

Βήμα 3: Σύσταση της λίστας L'_u στον χρήστη u

Το σύνολο των αντικειμένων που συστήνονται εξαρτάται σαφώς από την υπηρεσία που χρησιμοποιεί το $\Sigma\Sigma$ και δεν υπονοείται ότι αντικείμενα που έχουν συστηθεί στο παρελθόν δεν συστήνονται και πάλι αργότερα ή το αντίστροφο. Για παράδειγμα, για ένα $\Sigma\Sigma$ ταινιών δεν προτείνεται να συστήνει ταινίες που ο χρήστης έχει βαθμολογήσει πρόσφατα αφού έχει παρατηρηθεί πως οι άνθρωποι συνήθως δεν θέλουν να ξαναδούν την ίδια ταινία σε σύντομο χρονικό διάστημα. Αντίθετα για ένα $\Sigma\Sigma$ μουσικών τραγουδιών η επαναλαμβανόμενη σύσταση ενός υψηλά βαθμολογούμενου τραγουδιού από τον χρήστη αναμένεται, καθώς οι άνθρωποι πολύ πιθανόν να θέλουν να ξανά ακούσουν το ίδιο τραγούδι εάν ταιριάζει με την διάθεση τους.

Εφόσον ένα ΣΣ διαθέτει ως είσοδο τον χρήστη και την λίστα από τα πιθανά αντικείμενα προς σύσταση, το πρόβλημα σύστασης συνήθως λύνεται μέσω δύο τεχνικών:

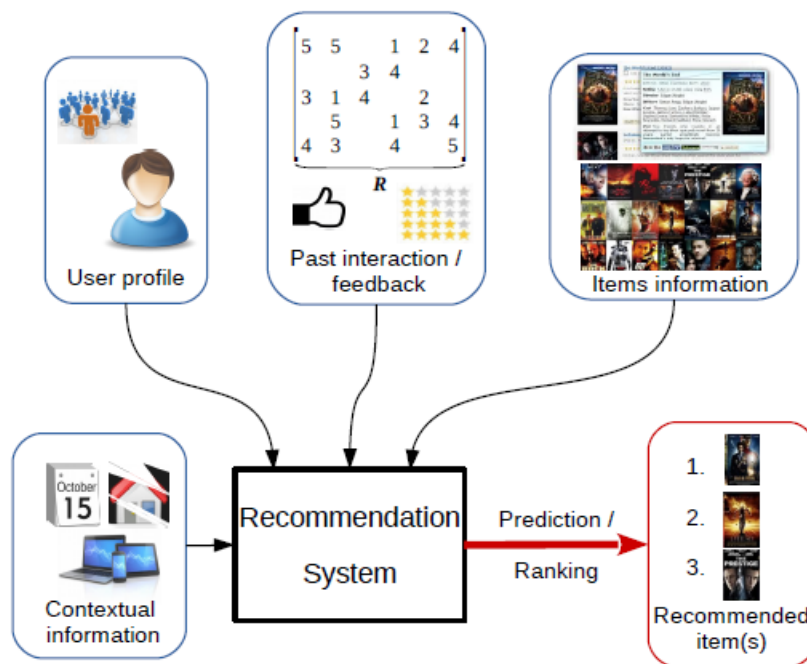
- *Συμπλήρωση του πίνακα συσχετίσεων R* : Αυτό σημαίνει ότι στόχος είναι να συμπληρωθεί όσο το δυνατόν περισσότερο ο πίνακας βαθμολογήσεων για τον χρήστη. Επειδή όμως η συμπλήρωση με τις πραγματικές βαθμολογήσεις του χρήστη για κάθε αντικείμενο είναι αδύνατη, το ΣΣ δημιουργεί προβλέψεις για τις κενές βαθμολογήσεις και βάσει αυτών κάνει τις συστάσεις. Σε αυτήν την περίπτωση η συνάρτηση ωφελείας θα συμπεριλαμβάνει την διαδικασία πρόβλεψης αυτών των βαθμολογήσεων.
- *Διατακτική προσέγγιση*: Οι προσεγγίσεις αυτές αντιμετωπίζουν το πρόβλημα συστάσεων από μια διαφορετική οπτική γωνία, υποθέτοντας ότι δεν υπάρχει ανάγκη για την πρόβλεψη όλων των βαθμολογήσεων των χρηστών για όλα τα αντικείμενα. Αρκεί να δημιουργηθεί η σωστή διάταξη βάσει της οποίας θα προταθούν τα αντικείμενα και παρατηρείται πως είναι αρκετό για να υπάρχει ικανοποίηση των χρηστών με τις συστάσεις που θα λάβουν. Να σημειωθεί πως είναι εφικτό να υπάρχουν διατακτικές προσεγγίσεις που προκύπτουν από την πρώτη μέθοδο, το οποίο σημαίνει πρακτικά ότι η διάταξη προκύπτει βάσει προβλέψεων. Παρόλα αυτά υπάρχουν πιο φυσικές μέθοδοι που στοχεύουν άμεσα στην βελτιστοποίηση της διάταξης.

Κατηγορίες ΣΣ και Τεχνικές

Τα ΣΣ ανάλογα με τις τεχνικές και τις προσεγγίσεις που χρησιμοποιούν διαχωρίζονται στα συστήματα εξατομικευμένης σύστασης και της μη εξατομικευμένης σύστασης.

Τα συστήματα μη εξατομικευμένης σύστασης δεν λαμβάνουν υπόψιν χαρακτηριστικά που αφορούν τον εκάστοτε χρήστη αλλά αξιοποιούν τις πληροφορίες που διαθέτουν με στόχο να εντοπίσουν τα πιο δημοφιλή αντικείμενα

του συστήματος και να τα προτείνουν. Ενώ αξιοποιούν τις βαθμολογήσεις, δεν κάνουν συσχετίσεις μεταξύ των χρηστών και επομένως κάθε χρήστης θεωρείται ίδιος με οποιονδήποτε άλλον όταν πρόκειται να του γίνει κάποια σύσταση. Χρησιμοποιούν πιο απλές δομές και αλγορίθμους από τα συστήματα εξατομικευμένης σύστασης και δεν θα αναλυθούν περισσότερο στα πλαίσια αυτής της διπλωματικής εργασίας.

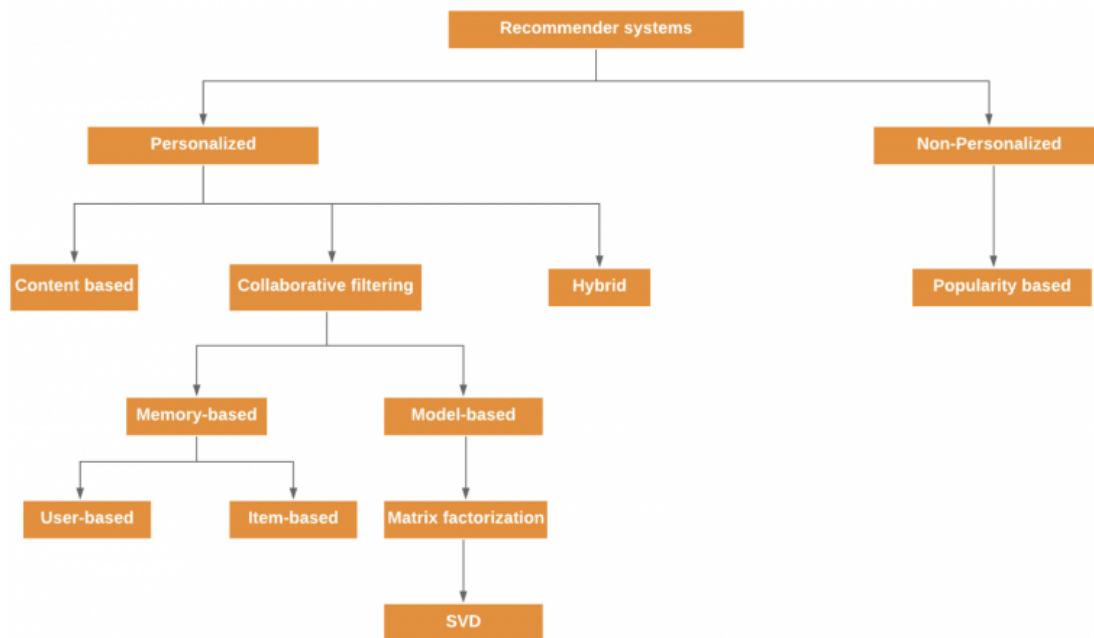


Εικόνα 1: Η διαδικασία σύστασης στην γενικότερη μορφή

Τα συστήματα εξατομικευμένης σύστασης μπορούν να κατηγοριοποιηθούν σε 3 μεγάλες κατηγορίες όσον αφορά τις τεχνικές τους προσεγγίσεις[81]:

- Συστάσεις βάσει περιεχομένου (Content-based Recommending): Σε αυτές τις τεχνικές η κεντρική ιδέα είναι ότι η σύσταση ενός αντικειμένου θα βασιστεί σε αντικείμενα παρόμοιου περιεχομένου που έχει αξιολογήσει υψηλά ο χρήστης στο παρελθόν ή ταιριάζουν με κάποια χαρακτηριστικά του.

- Συνεργατικό φιλτράρισμα (Collaborative Filtering – CF): Αυτές οι τεχνικές αξιοποιούν κατά βάση τις συσχετίσεις μεταξύ των χρηστών για να κάνουν συστάσεις με την λογική ότι ένας χρήστης έχει μεγαλύτερη πιθανότητα να ενδιαφέρεται για ένα αντικείμενο το οποίο έχουν βαθμολογήσει υψηλά άλλοι χρήστες με παρόμοια χαρακτηριστικά.
- Υβριδικές τεχνικές (Hybrid Techniques): Τεχνικές που συνδυάζουν και τις δύο προσεγγίσεις των Συστάσεων βάσει περιεχομένου και του Συνεργατικού φιλτραρίσματος.



Εικόνα 2: Δέντρο Τεχνικών ΣΣ

Θα αναφερθούν και άλλες μικρότερες αλλά σημαντικές κατηγορίες εξατομικευμένης σύστασης που δεν εμπίπτουν ολοκληρωτικά σε καμία από τις 3 μεγάλες κατηγορίες, όπως το Δημογραφικό Φιλτράρισμα, οι Συστάσεις βάσει Γνώσης, Κοινότητας και Συμφραζόμενων.

Συστάσεις βάσει Περιεχομένου

Οι τεχνικές Συστάσεων βάσει περιεχομένου (ή Φιλτράρισμα βάσει περιεχομένου) είναι μέθοδοι που κάνουν συγκρίσεις μεταξύ αναπαραστάσεων περιεχομένων που περιγράφουν τα αντικείμενα και που ενδιαφέρουν τους χρήστες. Παραδείγματα τέτοιων αναπαραστάσεων μπορεί να είναι οι λέξεις-κλειδιά που σχετίζονται με χαρακτηριστικά αντικειμένων πχ το είδος μουσικής και ο καλλιτέχνης για τραγούδια.

Πολλές προσεγγίσεις αυτής της κατηγορίας αντιμετωπίζουν το πρόβλημα σύστασης ως μια διαδικασία ανάκτησης πληροφορίας (Information Retrieval) όπου το περιεχόμενο που σχετίζεται με τις προτιμήσεις του χρήστη θεωρείται ένα ερώτημα και οι βαθμολογήσεις αντικειμένων που δεν έχει δώσει ο χρήστης για αυτά τα αντικείμενα εκτιμώνται βάσει της ομοιότητας των αντικειμένων με αυτό το ερώτημα. Για παράδειγμα σε ένα σύστημα ενημέρωσης, το περιεχόμενο του κάθε άρθρου μπορεί να αναπαρασταθεί με ένα διάνυσμα λέξεων το οποίο να συσχετιστεί με τα αντίστοιχα διανύσματα άρθρων που ο χρήστης έχει βαθμολογήσει στο παρελθόν. Με αυτόν τον τρόπο μπορεί να γίνει μια σύγκριση ομοιότητας μεταξύ των αντικειμένων.

Άλλες προσεγγίσεις στο πρόβλημα σύστασης είναι να αντιμετωπιστεί το πρόβλημα ως διαδικασία κατηγοριοποίησης (classification task) όπου οι παλαιότερες βαθμολογήσεις του χρήστη χρησιμοποιούνται για την εκπαίδευση ενός συστήματος κατηγοριοποίησης για τον εκάστοτε χρήστη και στην συνέχεια η βαθμολόγηση των νέων αντικειμένων να εκτιμάται βάσει της υπάρχουσας κατηγοριοποίησης. Ένα παράδειγμα κατηγοριοποίησης είναι σε μια κλίμακα βαθμολόγησης από το 1 έως το 5 να εκπαιδεύεται το σύστημα ώστε να κατηγοριοποιήσει τα αντικείμενα ανάμεσα σε αυτές τις πέντε κατηγορίες βαθμολόγησης. Φυσικά εδώ μπορούν να αξιοποιηθούν πολλές διαφορετικές τεχνικές κατηγοριοποίησης όπως τα δέντρα απόφασης, οι κ-γείτονες και τα νευρωνικά δίκτυα.

Υπάρχουν διάφορα πλεονεκτήματα των τεχνικών συστάσεων βάσει περιεχομένου:

1. Αξιοποιούν μόνο τις βαθμολογήσεις του εκάστοτε χρήστη και δεν χρειάζονται τις βαθμολογήσεις όλων των υπόλοιπων χρηστών.
2. Ένα νέο αντικείμενο μπορεί να συσταθεί βάσει των χαρακτηριστικών του ακόμα και εάν δεν υπάρχει ακόμα καμία βαθμολόγηση για αυτό από κανέναν χρήστη
3. Είναι εύκολο να παρέχεται επεξήγηση στον χρήστη για τον λόγο που του συστήνεται το κάθε αντικείμενο αφού η αναπαράσταση του περιεχομένου των αντικειμένων βασίζεται σε λέξεις-κλειδιά που είναι σαφή για τον χρήστη.

Παρόλα αυτά οι τεχνικές αυτές παρουσιάζουν και σημαντικές αδυναμίες:

1. Δεν είναι εφικτό να γίνονται καλές συστάσεις σε νέους χρήστες εφόσον οι ίδιοι δεν έχουν παράσχει στο σύστημα ακόμα αρκετές βαθμολογήσεις για να μπορεί να δημιουργηθεί το προφίλ περιεχομένου τους.
2. Ένας βασικός στόχος των ΣΣ που είναι η κατανάλωση ποικιλίας αντικειμένων δεν μπορεί να επιτευχθεί αφού αυτές οι τεχνικές τείνουν να δημιουργούν ένα υπερ-εξειδικευμένο προφίλ για τους χρήστες που σχετίζεται εν τέλει με αντικείμενα συγκεκριμένων κατηγοριών.
3. Είναι αναγκαία η πληρέστερη γνώση σχετικά με το περιεχόμενο των αντικειμένων διότι με ελλιπείς πληροφορίες είναι δύσκολο να κατασκευαστεί μια εύστοχη αναπαράσταση. Σε πολλά πεδία η πλήρης γνώση είναι ανέφικτη.

Συνεργατικό Φιλτράρισμα

Η κεντρική ιδέα πίσω από τις τεχνικές Συνεργατικού Φιλτραρίσματος (ΣΦ) είναι πως οι χρήστες που έχουν δείξει στο παρελθόν παρόμοιες προτιμήσεις μεταξύ τους, τείνουν να συμφωνούν και στο μέλλον. Ο όρος «Συνεργατικό» χρησιμοποιείται για να υποδηλώσει ότι το σύστημα ανακαλύπτει τις συσχετίσεις μεταξύ των χρηστών για να δημιουργήσει εξατομικευμένες συστάσεις για κάθε χρήστη. Οι ΣΦ τεχνικές συνήθως αξιοποιούν μόνο την ιστορικότητα των αλληλεπιδράσεων και

βαθμολογήσεων για να κατασκευάσουν το μοντέλο σύστασης τους και για τον λόγο αυτό θεωρούνται οι πιο δημοφιλείς τεχνικές στα ΣΣ. Επομένως η έννοια «παρόμοιες προτιμήσεις» είναι ένα ξεκάθαρο μέτρο που υπολογίζεται. Οι τεχνικές ΣΦ διαχωρίζονται περαιτέρω σε δύο μεγάλες υπο-κατηγορίες: Τα ΣΦ βάσει μνήμης και τα ΣΦ βάσει μοντέλων.

Συνεργατικό Φιλτράρισμα Μνήμης

Στα ΣΦ μνήμης ένα υποσύνολο χρηστών επιλέγεται βάσει τις «ομοιότητας» με τον ενεργό χρήστη και παράγεται μια πρόβλεψη βάσει των αξιολογήσεων τους. Γενικός αλγόριθμος πρόβλεψης:

1. Ορίζεται ένας συντελεστής βάρους σε όλους τους χρήστες βάσει της «ομοιότητας» τους με τον ενεργό χρήστη
2. Επιλέγονται οι k χρήστες με την υψηλότερη «ομοιότητα» (οι χρήστες αυτοί αναφέρονται και ως «γείτονες»)
3. Υπολογίζεται μια πρόβλεψη βάσει των βαθμολογήσεων των γειτόνων.

Στο βήμα 1 ο συντελεστής βάρους είναι ένα μέτρο «ομοιότητας» μεταξύ ενός χρήστη u και του ενεργού χρήστη a . Το πιο συνηθισμένο μέτρο «ομοιότητας» είναι ο συντελεστής συσχέτισης Pearson:

$$w_{a,u} = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2 \sum_{i \in I} (r_{u,i} - \bar{r}_u)^2}}$$

όπου I είναι το σύνολο των αντικειμένων που έχουν βαθμολογηθεί και από τους 2 χρήστες και \bar{r}_u , \bar{r}_a είναι οι μέσοι όροι βαθμολόγησης αντικειμένων του χρήστη u και του χρήστη a , αντίστοιχα.

Ένα εναλλακτικό μέτρο ομοιότητας μπορεί να είναι το \cos της γωνίας μεταξύ των διανυσμάτων βαθμολογήσεων των 2 χρηστών στο χώρο m διαστάσεων. Υπάρχουν πολλοί τρόποι υπολογισμού ενός μέτρου ομοιότητας όπως *Spearman rank correlation*,

Kendall's τ correlation, mean squared differences, entropy και adjusted cosine similarity.

Στο βήμα 3 ένα μέτρο πρόβλεψης είναι ο ισοσταθμισμένος μέσος όρος των αποκλίσεων από τον μέσο όρο των γειτόνων:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in K} (r_{u,i} - \bar{r}_u) \times w_{a,u}}{\sum_{u \in K} w_{a,u}}$$

όπου $p_{a,i}$ είναι η πρόβλεψη της βαθμολόγησης του αντικειμένου i από τον ενεργό χρήστη a , $w_{a,u}$ τα μέτρα ομοιότητας μεταξύ χρήστη a,u και K είναι το σύνολο των γειτόνων του a .

Παρακάτω παραθέτουμε κάποιες επεκτάσεις του $\Sigma\Phi$ μνήμης:

- $\Sigma\Phi$ βάσει αντικειμένων: Ένα πρόβλημα του γενικού $\Sigma\Phi$ μνήμης όπως περιγράφηκε παραπάνω είναι ότι για μεγάλα σύνολα χρηστών αυξάνεται σημαντικά η υπολογιστική πολυπλοκότητα της διαδικασίας εύρεσης «ομοιότητας» μεταξύ των χρηστών. Ένας εναλλακτικός τρόπος $\Sigma\Phi$ έχει προταθεί να είναι το $\Sigma\Phi$ βάσει αντικειμένων όπου αντί να συλλέγει όμοιους χρήστες να συλλέγει όμοια αντικείμενα. Αυτό σημαίνει πως σε αυτήν την περίπτωση το μέτρο «ομοιότητας» μεταξύ 2 αντικειμένων μπορεί να υπολογίζεται offline

$$w_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2 \sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}},$$

όπου U είναι το σύνολο όλων των χρηστών που έχουν βαθμολογήσει και τα δύο αντικείμενα i, j και r_i είναι ο μέσος όρος βαθμολόγησης του αντικειμένου i μεταξύ όλων των χρηστών. Επομένως αντίστοιχα η πρόβλεψη βαθμολόγησης για το αντικείμενο i μπορεί να είναι

$$p_{a,i} = \frac{\sum_{j \in K} r_{a,j} w_{i,j}}{\sum_{j \in K} |w_{i,j}|},$$

όπου K είναι οι γείτονες του αντικειμένου βάσει των συντελεστών ομοιότητας.

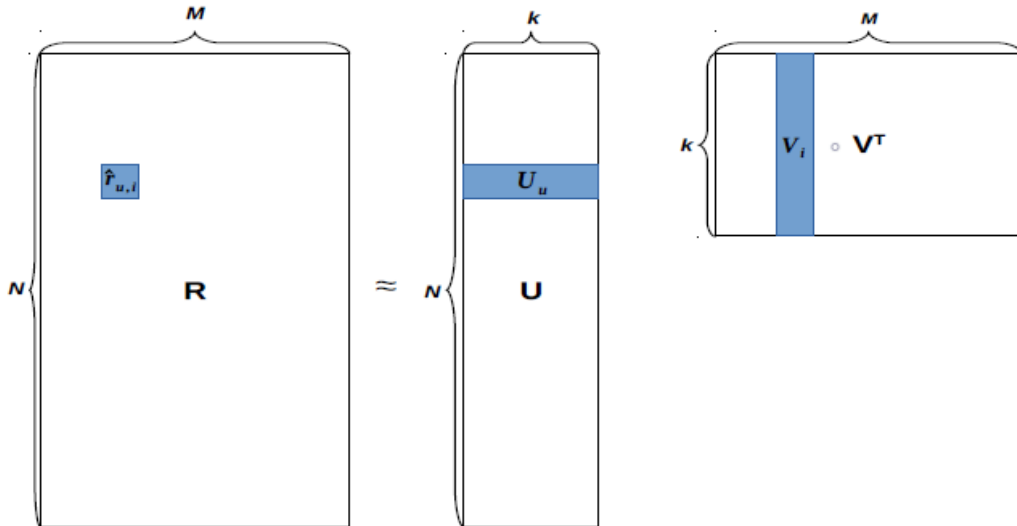
- **Συντελεστές Σημαντικότητας (Significance Weighting):** Είναι σύνηθες για τον ενεργό χρήστη να έχει γείτονες με πολύ υψηλό μέτρο ομοιότητας που έχει προκύψει από ελάχιστα κοινώς βαθμολογούμενα αντικείμενα. Τέτοιες περιπτώσεις γειτόνων που βασίζονται σε πολύ λίγα αντικείμενα τείνουν να δημιουργούν αδύναμες προβλέψεις. Μια τεχνική που αντιμετωπίζει αυτό το φαινόμενο είναι η ενσωμάτωση ενός συντελεστή σημαντικότητας όπου βάσει αυτού το μέτρο ομοιότητας τέτοιων γειτόνων μειώνεται.
- **Συντελεστής Συχνότητας:** Αντίστοιχα, αντικείμενα που έχουν βαθμολογηθεί από όλους τους χρήστες, σε πολλές περιπτώσεις δεν είναι τόσο χρήσιμα για προβλέψεις όσο τα λιγότερο δημοφιλή αντικείμενα. Μια τεχνική που λαμβάνει υπόψιν το παραπάνω είναι ο υπολογισμός κάποιου συντελεστή συχνότητας της αξιολόγησης του κάθε αντικειμένου όπου πολλαπλασιάζεται με την βαθμολόγηση του και έτσι όσο πιο δημοφιλές είναι ένα αντικείμενο τόσο λιγότερη επιρροή θα έχει στην φόρμουλα πρόβλεψης. Η εφαρμογή της τεχνικής αυτής βασίζεται στην υπόθεση ότι τα αντικείμενα που είναι πολύ δημοφιλή είτε με θετικό είτε με αρνητικό τρόπο τείνουν να βαθμολογούνται και πιο συχνά.

Υπάρχουν και πολλές άλλες τεχνικές που ενισχύουν τις ΣΦ προσεγγίσεις επιλύοντας διάφορα προβλήματα που οι κλασικές ΣΦ δεν μπορούν να αντιμετωπίσουν όπως το *case amplification*, *weighted majority prediction* και *imputation-boosted collaboration filtering*.

Συνεργατικό Φιλτράρισμα Μοντελοποίησης

Οι τεχνικές ΣΦ μοντελοποίησης βασίζονται στην εκτίμηση παραμέτρων στατιστικών μοντέλων για τις βαθμολογήσεις του χρήστη. Για παράδειγμα το σύστημα μπορεί να κατασκευάσει έναν ταξινομητή (classifier) για κάθε χρήστη ο οποίος αξιοποιεί τα αντικείμενα ως διανύσματα χαρακτηριστικών (feature vectors) και τις βαθμολογήσεις ως τις κατηγορίες ταξινόμησης (labels). Οι πιο διάσημες υποκατηγορίες ΣΦ μοντελοποίησης είναι η latent factor (κρυμμένοι παράγοντες) και

η matrix factorization. Σε αντίθεση με τις τεχνικές ΣΦ μνήμης οι οποίες βασίζονται στην «ομοιότητα» μεταξύ αντικειμένων και χρηστών, τα μοντέλα Latent Factor υποθέτουν ότι η «ομοιότητα» προκύπτει μέσω κάποιας δομής μικρής διάστασης στα δεδομένα. Για παράδειγμα, η βαθμολόγηση που δίνει ένας χρήστης σε μια ταινία μπορεί να θεωρηθεί εξαρτημένη από παράγοντες όπως η γενικότερη προτίμηση του χρήστη για το είδος που ανήκει η ταινία. Η Matrix Factorization τεχνική θεωρείται κατηγορία της Latent Factor όπου τα αντικείμενα και οι χρήστες αναπαριστούνται από διανύσματα άγνωστων χαρακτηριστικών U_u, V_i . Αυτά τα χαρακτηριστικά μαθαίνονται έτσι ώστε το γινόμενο $U_u \times V_i^T$ να προσεγγίζει την γνωστή αξιολόγηση r_{ui} βάσει ενός περιθωρίου σφάλματος.



Εικόνα 3: Αναπαράσταση Matrix Factorization

Οι διάφορες επιλογές για τις συναρτήσεις απώλειας, τεχνικές κανονικοποίησης και περιορισμοί στο μοντέλο δημιουργούν μια τεράστια γκάμα υποκατηγοριών για MF.

Συνεχίζοντας το παράδειγμα, έστω ότι ο πίνακας V αναπαριστά το είδος μιας ταινίας ενώ ο πίνακας U αναπαριστά πόσο ο χρήστης προτιμάς ένα είδος ταινιών. Τότε η αξιολόγηση του χρήστη για μια ταινία θα είναι το άθροισμα όλων των γινομένων

μεταξύ της προτίμησης του χρήστη σε ένα είδος και του αντικειμένου στο ίδιο είδος, δηλαδή $U_u \times V_i^T$. Στο παράδειγμα αυτό στόχος είναι να γίνει αντιληπτή η ιδέα πίσω από τα κρυμμένα χαρακτηριστικά των χρηστών και των αντικειμένων αλλά στην πράξη πολλές φορές δεν είναι εύκολο να υπάρχει ερμηνεία για κάθε γινόμενο χαρακτηριστικών.

Σε σχέση με τις τεχνικές ΣΦ μνήμης, οι τεχνικές ΣΦ μοντελοποίησης παρουσιάζουν τα εξής πλεονεκτήματα:

- *Ταχύτητα*: Η μοντελοποίηση μέσω αυτών των τεχνικών γίνεται πολύ πιο γρήγορα
- *Εξοικονόμηση χώρου*: Λόγω της μείωσης διαστάσεων ο χώρος αποθήκευσης δεδομένων είναι μικρότερος
- *Ισχυρότερη γενίκευση*: Μέσω τις αξιοποίησης τεχνικών κανονικοποίησης κατά της διάρκειας της εκμάθησης του μοντέλου, υπάρχει σημαντική μείωση *overfitting*

Υβριδικές τεχνικές

Με στόχο να συνδυαστούν τα πλεονεκτήματα των τεχνικών Σύστασης βάσει Περιεχομένου και Συνεργατικού Φιλτραρίσματος κατασκευάζονται διάφοροι συνδυασμοί που χαρακτηρίζονται ως Υβριδικές Τεχνικές. Μια απλή προσέγγιση είναι ένα σύστημα που δημιουργεί δύο διαφορετικές λίστες συστάσεων αντικειμένων που η κάθε μία έχει δημιουργηθεί από μία τεχνική και τελικώς να συστήνεται ένας συνδυασμός αντικειμένων και από τις δύο λίστες. Άλλες Υβριδικές τεχνικές μπορούν να χρησιμοποιούν κατά βάση τεχνικές ΣΦ αλλά να διατηρούν προφίλ χρηστών βάσει τεχνικών Περιεχομένου.

Δημογραφικό Φιλτράρισμα

Αυτή η κατηγορία τεχνικών κάνει εξαρχής μια κατηγοριοποίηση των χρηστών βάσει των δημογραφικών τους χαρακτηριστικών όπως η χώρα, η ηλικία, το φύλο κτλ (Demographic Filtering). Στην συνέχεια εφαρμόζονται διαφορετικοί κανόνες για την διεξαγωγή των συστάσεων σε κάθε κατηγορία. Αυτές οι τεχνικές είναι παρόμοιες με τεχνικές κατηγοριοποίησης (classification) ή παλινδρόμησης (regression) στις οποίες τα χαρακτηριστικά εισόδου είναι τα δημογραφικά χαρακτηριστικά ως οι ανεξάρτητες μεταβλητές ενώ οι εξαρτημένες μεταβλητές είναι οι προτιμήσεις του χρήστη. Αυτές οι τεχνικές δεν παρουσιάζουν ισχυρή εκτιμητική ικανότητα σε σχέση με τις προηγούμενες τεχνικές αλλά μπορούν να προσφέρουν σημαντική ενίσχυση στο σύστημα εάν συνδυαστούν μαζί με άλλες.

Συστάσεις βάσει Γνώσης

Τα ΣΣ Γνώσης (Knowledge Based Recommendations) λειτουργούν βάσει ενός εννοιολογικού περιεχομένου μέσα στο οποίο ο χρήστης έχει θέσει κάποιους περιορισμούς για την αναζήτηση των αντικειμένων. Αυτές οι τεχνικές χρησιμοποιούνται κυρίως σε συστήματα όπου η αλληλεπίδραση ενός χρήστη με το σύστημα είναι σπάνια (πχ site αγοραπωλησίας αυτοκινήτων). Στόχος είναι να αντικατασταθεί η έλλειψη ιστορικότητας των βαθμολογήσεων του χρήστη από τους περιορισμούς που θέτει ο ίδιος ώστε να γίνουν εύστοχες συστάσεις. Οι μέθοδοι που χρησιμοποιούν τέτοια συστήματα είναι παρόμοιες με τις μεθόδους της Σύστασης βάσει Περιεχομένου με την διαφορά ότι δεν αξιοποιείται η ιστορικότητα του χρήστη αλλά αναζητείται η ομοιότητα μεταξύ των αντικειμένων και των περιορισμών που έχουν τεθεί. Τα ΣΣ Γνώσης τείνουν να λειτουργούν καλύτερα στην αρχή τους αλλά εάν δεν συνδυάζονται στην συνέχεια και με άλλες τεχνικές εκμάθησης πληροφοριών δεν μπορούν να ξεπεράσουν σε ευστοχία άλλες τεχνικές που είναι πολύ πιο απλές όπως για παράδειγμα την απλούστερη μορφή του Συνεργατικού Φιλτραρίσματος.

Συστάσεις βάσει Κοινότητας

Αυτές οι τεχνικές ΣΣ (Community Based Recommendations) δημιουργούν λίστες συστάσεων βάσει των βαθμολογήσεων που έχουν κάνει οι χρήστες που έχουν στενή σύνδεση με τον ενεργό χρήστη (πχ φίλοι, συνεργάτες κτλ). Βασίζονται στην υπόθεση ότι οι άνθρωποι τείνουν να εμπιστεύονται συστάσεις από πρόσωπα που ανήκουν στην ίδια κοινότητα με αυτούς. Φυσικά τέτοιες τεχνικές αξιοποιούνται κυρίως εντός των κοινωνικών δικτύων (τα οποία κυριαρχούν τα τελευταία χρόνια) ή αντλώντας πληροφορίες από αυτά.

Συστάσεις βάσει Συμφραζόμενων

Μια ειδική κατηγορία τεχνικών σύστασης είναι οι τεχνικές που αποσκοπούν να εξάγουν τα συμφραζόμενα της αλληλεπίδρασης του χρήστη με το σύστημα (Context-Aware Recommendations). Αυτό είναι εφικτό μέσω μιας σειράς μετασχηματισμών πριν, μετά και κατά την διάρκεια της επεξεργασίας των δεδομένων που υπάρχουν για τον ενεργό χρήστη. Τα συμφραζόμενα μπορούν να αναπαριστώνται μέσω δύο τρόπων:

- Μέσω χαρακτηριστικών που παρατηρούνται στις οντότητες που συμμετέχουν και δεν αλλάζουν καθ' όλη την διάρκεια του συστήματος
- Μέσω των αλληλεπιδράσεων ή σχέσεων των οντοτήτων μεταξύ τους και αλλάζουν δυναμικά.

Προβλήματα και Περιορισμοί των Συστημάτων Σύστασης

Παρακάτω παρουσιάζονται τα βασικότερα φαινόμενα και οι συνθήκες που εμποδίζουν τα ΣΣ να λειτουργήσουν αποτελεσματικά:

1. *Αραιότητα*: Το πιο σύνηθες είναι οι περισσότεροι χρήστες να μην βαθμολογούν τα περισσότερα αντικείμενα που σημαίνει ότι ο πίνακας βαθμολογήσεων R είναι συχνά ένας πολύ αραιός πίνακας. Αυτό το φαινόμενο δημιουργεί κατά βάση πρόβλημα στις τεχνικές Συνεργατικού Φιλτραρίσματος αφού μειώνει σημαντικά την πιθανότητα να βρεθεί σύνολο χρηστών με παρόμοιες βαθμολογήσεις. Αυτό το πρόβλημα προκύπτει συχνά σε συστήματα όπου υπάρχει υψηλό ποσοστό αναλογίας όλων των αντικειμένων προς έναν χρήστη ή όταν το σύστημα είναι στην αρχή της λειτουργίας του.
2. *Τάση και Έκλειψη Αντικειμένων*: Στα δυναμικά συστήματα είναι πιθανόν πολλά αντικείμενα να προστίθενται και να αφαιρούνται με γρήγορους ρυθμούς ή να δημιουργείται μια τάση υψηλής βαθμολόγησης από τους χρήστες για κάποιο αντικείμενο η οποία διαρκεί μικρό χρονικό διάστημα. Ένα ξεκάθαρο παράδειγμα αυτού του φαινομένου είναι τα συστήματα δημοσιογραφικής ενημέρωσης όπου τα νέα πολλές φορές παλιώνουν γρήγορα και αντικαθίστανται από τα επόμενα. Ένα ΣΣ είναι απαραίτητο να προσαρμόζεται πολύ γρήγορα μέσω της συνεχόμενης ενημέρωσης του μοντέλου του. Η χρήση πληροφοριών περιεχομένου (contextual information) είναι μια σημαντική παράμετρος (πχ η ημερομηνία εμφάνισης του αντικειμένου στο σύστημα) η οποία βοηθάει ένα ΣΣ να επιλύσει τέτοιου είδους φαινόμενα.
3. *Αργή Εκκίνηση*: Η Αργή Εκκίνηση (Cold-Start) είναι το κύριο φαινόμενο που αφορά την έλλειψη πληροφορίας κατά την εισαγωγή ενός νέου χρήστη ή νέου αντικειμένου στο σύστημα. Το πρόβλημα αυτό εντοπίζεται κυρίως στις τεχνικές Συνεργατικού Φιλτραρίσματος όπου αξιοποιείται η ιστορικότητα του χρήστη και του αντικειμένου για τις συστάσεις. Διαχωρίζεται από το

πρόβλημα της Αραιώσης διότι, αν και παρόμοιας φύσεως, σε συστήματα που ο πίνακας αξιολογήσεων R δεν είναι αραιός, το πρόβλημα της Αργής Εκκίνησης εξακολουθεί να υπάρχει. Η αξιοποίηση υβριδικών τεχνικών είναι μια λύση στο πρόβλημα που παρουσιάζεται πολύ συχνά στην πράξη.

4. *Βελτιστοποίηση Μετρικών Κατάταξης*: Ενώ έχει γίνει σαφές ότι στις τεχνικές $\Sigma\Sigma$ ο στόχος είναι να γίνουν προβλέψεις για τις άγνωστες βαθμολογήσεις αντικειμένων, στην πράξη η ικανοποίηση του χρήστη εξαρτάται από την επιτυχή σύσταση των πρώτων αντικειμένων της λίστας που του παρουσιάζεται. Για αυτόν τον λόγο τα τελευταία χρόνια έχει στραφεί το ενδιαφέρον της έρευνας $\Sigma\Sigma$ στην εύστοχη κατάταξη των αντικειμένων σύστασης. Εφευρίσκονται και μελετώνται τεχνικές επομένως, που εστιάζουν στην βελτιστοποίηση των μετρικών κατάταξης των αντικειμένων και όχι στην βελτιστοποίηση των μετρικών πρόβλεψης.
5. *Ασφάλεια και Εξαπάτηση*: Η ανάπτυξη των $\Sigma\Sigma$ στον κλάδο του εμπορίου έχει κάνει σαφές πως τα $\Sigma\Sigma$ παίζουν σημαντικό ρόλο στην κερδοφορία των επιχειρήσεων. Για το λόγο αυτό πολλοί πωλητές έχουν κίνητρο να διαβρώσουν το σύστημα μέσω στοχευόμενων βαθμολογήσεων ώστε να αναδείξουν ένα προϊόν (push attacks) ή να υποτιμήσουν προϊόντα του ανταγωνισμού (nuke attacks). Τέτοιου είδους επιθέσεις διάβρωσης πραγματοποιούνται μέσω εικονικών χρηστών που σε μεγάλο πλήθος είναι πολύ εύκολο να αλλάζουν τον πίνακα αξιολόγησης R προς το συμφέρον του υποκινητή. Είναι επομένως επιτακτική ανάγκη τα $\Sigma\Sigma$ να διαθέτουν μηχανισμούς όπου είτε εντοπίζουν και αγνοούν είτε αντισταθμίζουν τέτοιου είδους εικονικές βαθμολογήσεις.

Ενώ φαίνεται ότι οι τεχνικές Σύστασης βάσει Περιεχομένου αποφεύγουν πολλά από τα παραπάνω προβλήματα, οι τεχνικές Συνεργατικού Φιλτραρίσματος προσφέρουν σημαντικά πλεονεκτήματα σε σχέση με αυτές. Αρχικά οι τεχνικές $\Sigma\Phi$ μπορούν να λειτουργήσουν επιτυχώς σε πεδία όπου δεν υπάρχει σαφής τρόπος να συσχετιστεί περιεχόμενο με τα αντικείμενα ή όταν το περιεχόμενο είναι δύσκολο να

αναλυθεί από υπολογιστικής πλευράς (πχ ιδέες και απόψεις σε άρθρα). Επίσης οι τεχνικές ΣΦ έχουν την δυνατότητα να παρέχουν συστάσεις αντικειμένων που δεν προκύπτουν άμεσα από την ιστορικότητα του ενεργού χρήστη αλλά ταυτόχρονα να έχουν υψηλή πιθανότητα να είναι επιτυχημένα και έτσι να φέρουν σε επαφή τον χρήστη με αντικείμενα που δεν θα μπορούσε να έρθει διαφορετικά. Φαίνεται επομένως ότι για την μέγιστη αξιοποίηση των δυνατοτήτων των ΣΣ, ένα πλήρως ποιοτικό ΣΣ τείνει να αξιοποιεί υβριδικές τεχνικές έτσι ώστε να αντιμετωπίζει τα προβλήματα που παρατέθηκαν ενώ ταυτόχρονα να λαμβάνει την μέγιστη αξία ανάλογα με το πεδίο στο οποίο εφαρμόζεται.

Μελλοντικές Προκλήσεις

Παρακάτω αναφέρονται περιληπτικά κάποιες προκλήσεις που προβλέπεται να αντιμετωπίσουν τα ΣΣ στο μέλλον:

1. *Κλιμάκωση των αλγορίθμων ΣΣ στα πλαίσια των «μεγάλων δεδομένων»:*
Καθώς η έρευνα των τεχνικών ΣΣ εξελίσσεται, έχει γίνει σαφές πως ίσως η βασικότερη πρόκληση που έχουν να αντιμετωπίσουν είναι τα τεράστια και δυναμικά σύνολα δεδομένων που προκύπτουν από τις αλληλεπιδράσεις των χρηστών με τα συστήματα υπολογιστών. Μια λύση που μπορεί να φαίνεται πολύ αποδοτική σε ένα σχετικά μικρό και ελεγχόμενο περιβάλλον εργαστηρίου πολλές φορές αποδεικνύεται ανεπαρκής όταν πρόκειται να αντιμετωπίσει τα δεδομένα του πραγματικού κόσμου.
2. *Προληπτικά Συστήματα Σύστασης:* Τα περισσότερα ΣΣ μέχρι τώρα λειτουργούν μέσω της λογικής πως ο χρήστης ζητάει άμεσα συστάσεις. Στην νέα εποχή όπου οι υπολογιστές και οι χρήστες είναι διαρκώς συνδεδεμένοι μεταξύ τους φαίνεται λογικό επαχόλουθο να υπάρχουν ΣΣ που κάνουν συστάσεις έμμεσα χωρίς να χρειάζονται τους χρήστες να το ζητήσουν. Με αυτόν τον τρόπο τα ΣΣ μπορούν να λειτουργούν προβλεπτικά χωρίς να θεωρούνται ενοχλητικά από τους χρήστες.

3. *Ιδιωτικότητα*: Τα ΣΣ αξιοποιούν δεδομένα χρηστών για να δημιουργούν εξατομικευμένες συστάσεις. Επειδή όσο περισσότερα δεδομένα είναι διαθέσιμα τόσο πιο αποτελεσματικό θα είναι ένα ΣΣ τίθεται το ζήτημα των προσωπικών δεδομένων όπου για πολλούς χρήστες θεωρούνται ευαίσθητα δεδομένα. Αυτό έχει ως αποτέλεσμα πολλοί χρήστες να αισθάνονται καχυποψία απέναντί τους και επομένως μια πρόκληση για τον σχεδιασμό ενός ΣΣ είναι η εξασφάλιση της αποτελεσματικότητας δίνοντας στους χρήστες την δυνατότητα να επιλέγουν τα δεδομένα που θέλουν οι ίδιοι να διαθέτουν καθώς και την προστασία των δεδομένων τους από κακόβουλους χρήστες.
4. *Ποικιλομορφία Συστάσεων*: Σε μια λίστα συστάσεων, το πιο πιθανό είναι, ο χρήστης να εντοπίζει κατάλληλα αντικείμενα τα οποία όμως είναι διαφορετικά σε ένα βαθμό μεταξύ τους. Δεν έχει τόσο αξία μια λίστα που τα αντικείμενα που συστήνονται είναι αυστηρά όμοια μεταξύ τους, με εξαίρεση όταν ο ίδιος ο χρήστης έχει αναζητήσει αντικείμενα με αυστηρά κριτήρια. Μάλιστα ειδικά στα αρχικά στάδια της αλληλεπίδρασης με ένα ΣΣ, οι χρήστες ενδιαφέρονται να αναζητήσουν νέα και διαφορετικά αντικείμενα, που σημαίνει ότι αξιοποιούν το ΣΣ ως εργαλείο αναζήτησης επιλογών και διεύρυνσης των προτιμήσεων τους.
5. *Ενσωμάτωση των μακροπρόθεσμων και των βραχυπρόθεσμων προτιμήσεων*: Τα ΣΣ μπορούν να διαχωριστούν σε εκείνα που στοχεύουν να δημιουργήσουν ένα μακροπρόθεσμο προφίλ χρηστών μέσω της επαναλαμβανόμενης συνάθροισης δεδομένων από όλους τους χρήστες του συστήματος και σε εκείνα που στόχο έχουν την αξιοποίηση των εφήμερων προτιμήσεων. Προφανώς και οι δύο προσεγγίσεις είναι χρήσιμες, ανάλογα με το πεδίο εφαρμογής τους. Μάλιστα υπάρχουν έρευνες που υποδεικνύουν υβριδικές τεχνικές, όπου το σύστημα μπορεί να αποφασίζει εάν θα αποκλίνει ή όχι από μια συνεχόμενη και μακροπρόθεσμη κατασκευή προφίλ προτιμήσεων όταν υπάρχουν αρκετά στοιχεία που υποδηλώνουν ότι οι προτιμήσεις του χρήστη τείνουν να γίνονται βραχυπρόθεσμες.
6. *Αξιοποίηση Μοντέλων μεταξύ διαφορετικών συστημάτων*: Λόγω του ότι οι χρήστες είναι διασυνδεδεμένοι σε πολλά συστήματα ταυτόχρονα, υπάρχει η

πρόκληση της αξιοποίησης δεδομένων και προφίλ χρηστών από το ένα σύστημα σε ένα άλλο. Δηλαδή, οι πληροφορίες που υπάρχουν σε ένα σύστημα για έναν χρήστη και έχουν αξιοποιηθεί για να κατασκευαστεί ένα προφίλ προτιμήσεων, να χρησιμοποιηθούν για να κατασκευάσουν ένα άλλο προφίλ σε ένα διαφορετικό σύστημα.

7. *Κατανεμημένα Συστήματα Συστάσεων εντός ανοικτών δικτύων:* Το αρχιτεκτονικό μοντέλο των περισσότερων ΣΣ διαθέτει την συνηθισμένη δομή πελάτη-διακομιστή (client-server) όπου οι χρήστες (πελάτες) αλληλεπιδρούν όλοι με το ίδιο κεντρικό σύστημα (διακομιστής) για να λάβουν συστάσεις. Αυτή η δομή φυσικά αντιμετωπίζει όλα τα προβλήματα που αντιμετωπίζουν όλες οι κεντροποιημένες αρχιτεκτονικές. Η πρόκληση επομένως αφορά το σενάριο των διαδικτυακών υπολογιστικών συστημάτων τα οποία μπορούν να υποστηρίξουν πιο αποτελεσματικά και ευέλικτα υπολογιστικά μοντέλα για τα συστήματα συστάσεων.
8. *Βελτιστοποίηση Συστάσεων σε Ακολουθία:* Τα Ακολουθιακά Συστήματα Συστάσεων ή ΣΣ Διαλόγου (Conversational RS) είναι μια βελτίωση των ΣΣ όπου ενσωματώνουν δυνατότητες εκμάθησης, βελτιστοποιώντας όχι μόνο τις συστάσεις αντικειμένων αλλά και την ακολουθιακή αλληλεπίδραση του χρήστη με αυτά (διάλογος του χρήστη με το σύστημα).
9. *Συστάσεις σχεδιασμένες πάνω σε κινητές εφαρμογές:* Καθώς η τεχνολογία των κινητών υπολογιστών εξελίσσεται, γίνεται προφανές ότι υπάρχει ανάγκη οι συστάσεις να γίνονται «εν κινήσει» (πχ όταν ο χρήστης βρίσκεται στο δρόμο και αναζητά μαγαζιά ή ξενοδοχεία). Αυτό σημαίνει ότι ο σχεδιασμός υπολογιστικών διεπαφών πάνω σε κινητούς υπολογιστές απαιτεί την βέλτιστη αξιοποίηση των περιορισμένων πόρων (υπολογιστική ισχύς, μέγεθος οθόνης κτλ) των κινητών υπολογιστών.
10. *Διαφάνεια:* Τα ΣΣ μπορούν να ενσωματώσουν τρόπους ούτως ώστε να επεξηγούν με διαφάνεια την λογική που κάνουν τις συστάσεις τους στον εκάστοτε χρήστη. Αυτό βοηθάει αρχικά στην εμπιστοσύνη που μπορεί να έχει

ο χρήστης στο σύστημα και επίσης βοηθάει στην ανάπτυξη μιας πιο επιτυχημένης αλληλεπίδρασης αφού ο χρήστης θα κατανοεί πώς η συμπεριφορά του μέσα στο σύστημα επηρεάζει τις συστάσεις και επομένως θα την προσαρμόζει συνειδητά με στόχο να λαμβάνει την μέγιστη αξία από το σύστημα συστάσεων.

11. *Αξιολόγηση*: Μια σημαντική πρόκληση είναι ο τρόπος που ένα ΣΣ αξιολογείται. Υπάρχουν διάφορες μέθοδοι και μετρικές βάσει των οποίων μπορεί να αξιολογηθεί ένα σύστημα. Το πεδίο εφαρμογής του ΣΣ παίζει σημαντικό ρόλο για την επιλογή της κατάλληλης μεθόδου αξιολόγησης.
12. *Αναζήτηση εναντίον Αξιοποίησης*: Η πρόκληση αυτή αφορά το θεμελιώδες δίλημμα μεταξύ αναζήτησης και αξιοποίησης (exploration vs exploitation) πάνω στο οποίο ένας σχεδιαστής ΣΣ πρέπει να πάρει αποφάσεις. Το δίλημμα αφορά τις διάφορες πτυχές ενός ΣΣ, από την αξιοποίηση της υπάρχουσας πληροφορίας ή την αναζήτηση επιπλέον δεδομένων μέχρι τις συστάσεις που κάνει οι οποίες μπορούν μέσω μιας μη βέλτιστης εκτίμησης (αναζήτηση) να διερευνήσουν περαιτέρω πληροφορίες για τον χρήστη. Η πρόκληση αυτή είναι και το αντικείμενο της παρούσας διπλωματικής εργασίας η οποία θα παρουσιάσει διάφορους αλγόριθμους τεχνικών Αναζήτησης – Αξιοποίησης και τον συνδυασμό τους με τα Συστήματα Συστάσεων.
13. *Σχέση Έρευνας και Βιομηχανίας*: Τέλος μια ενδιαφέρουσα πρόκληση για τα ΣΣ είναι η συνεργασία μεταξύ της έρευνας που γίνεται από τα πανεπιστήμια και της βιομηχανίας που εφαρμόζει στην πράξη συστήματα συστάσεων. Η βιομηχανία αντιμετωπίζει πολύ συγκεκριμένα προβλήματα τα οποία όμως δεν γίνονται ξεκάθαρα για λόγους ανταγωνισμού. Από την άλλη η έρευνα εστιάζει σε προβλήματα ΣΣ που μπορούν να αντιμετωπιστούν εντός του πλαισίου των ακαδημαϊκών πόρων και ανάπτυξης τα οποία έχουν κυρίως ακαδημαϊκή αξία. Η συνεργασία μεταξύ των δύο πλευρών είναι μια πρόκληση η οποία εάν αντιμετωπιστεί εκτιμάται ότι πολλά πρακτικά ζητήματα θα μπορέσουν να αντιμετωπιστούν από την συλλογική έρευνα εφόσον γίνουν σαφέστερα από την βιομηχανία.

Αξιολόγηση των Συστημάτων Σύστασης

Για να αξιολογήσει κανείς την απόδοση ενός ΣΣ χρειάζεται σαφή κριτήρια και μετρικές, καθώς και την ορθή μεθοδολογία αξιολόγησης. Η επιλογή των παραπάνω είναι κρίσιμης σημασίας καθώς η λάθος αξιολόγηση ενός συστήματος έχει σημαντικές συνέπειες για την υπηρεσία που εφαρμόζει, όπως η απώλεια πελατών. Σε αυτήν την ενότητα θα παρουσιαστούν οι πιο βασικές μετρικές καθώς και οι μέθοδοι αξιολόγησης για τα ΣΣ.

Μετρικές και Κριτήρια Αξιολόγησης

Υπάρχει μεγάλο εύρος μετρικών που έχουν χρησιμοποιηθεί για την έρευνα των ΣΣ. Η κάθε μετρική εστιάζει συνήθως και σε μια διαφορετική συνάρτηση ωφέλειας που έχει κατασκευαστεί καταλλήλως ώστε να ταιριάζει με το πεδίο και τους στόχους εφαρμογής του συστήματος. Είναι πιθανόν, και πολλές φορές προτείνεται, η αξιολόγηση ενός ΣΣ να χρησιμοποιεί διαφορετικές μετρικές ώστε να αξιολογηθεί η αποτελεσματικότητα του με διαφορετικά κριτήρια.

Κριτήριο Ευστοχίας στην Εκτίμηση

Οι μέθοδοι Συνεργατικού Φιλτραρίσματος έχουν στόχο την εκτίμηση των βαθμολογήσεων που λείπουν για τους χρήστες. Αυτό έχει ως συνέπεια, οι μετρικές της ευστοχίας εκτίμησης να είναι οι βασικές μετρικές αξιολόγησης ενός ΣΣ που χρησιμοποιεί τεχνικές ΣΦ. Έχοντας ένα σύνολο από κρυφές πραγματικές βαθμολογήσεις μπορεί κάποιος να μετρήσει πόσο κοντά είναι οι εκτιμήσεις του συστήματος για αυτές. Οι πιο συνήθεις μετρικές εκτίμησης είναι οι παρακάτω:

- Το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error – MAE) μετρά την μέση απόλυτη απόκλιση μεταξύ της πραγματικής βαθμολόγησης r_{ui} από την εκτίμηση

της \hat{r}_{ui} δεδομένου ενός συνόλου κρυφών βαθμολογήσεων T :

$$MAE = \frac{1}{T} \sum_{(u,i) \in T} |r_{ui} - \hat{r}_{ui}|$$

- Το Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error – MSE) μετρά την τετραγωνική απόκλιση και εξ' αυτού εστιάζει στα πιο μεγάλα σφάλματα σε σχέση με το MAE.

$$MAE = \frac{1}{T} \sum_{(u,i) \in T} (r_{ui} - \hat{r}_{ui})^2$$

- Η ρίζα του MTΣ χρησιμοποιείται αρκετές φορές αντί του MTΣ

$$RMSE = \sqrt{MSE}$$

Σε όλα αυτά τα μέτρα, όσο πιο μικρό σε τιμή είναι το σφάλμα, τόσο πιο καλή προβλεπτική ισχύ έχει το σύστημα. Παρόλα αυτά, τέτοιες μετρικές δίνουν μεγάλη έμφαση στα δημοφιλή αντικείμενα και έτσι αντικείμενα που έχουν σπάνιες βαθμολογήσεις δεν μπορούν να τις επηρεάσουν σημαντικά. Για παράδειγμα, σε ένα σύστημα όπου όλες οι βαθμολογήσεις είναι γύρω στο 3 σε μια κλίμακα από 1 έως 5, η εκτίμηση θα είναι ευστοχότερη εάν δοθεί το 3 ως τιμή εκτίμησης για κάθε μη βαθμολογημένο αντικείμενο ενώ θα ήταν πολύ πιο σημαντικό από την οπτική του χρήστη να δοθεί μεγαλύτερη βαρύτητα σε βαθμολογήσεις πιο στοχευόμενων αντικειμένων.

Κριτήριο Ανάκτησης Πληροφοριών (Information Retrieval)

Ενώ το κριτήριο ευστοχίας εκτίμησης μετρά πόσο καλά μπορεί το σύστημα να εκτιμήσει τις βαθμολογήσεις, δεν εστιάζει στην ικανοποίηση των χρηστών (πχ αν τελικά ο χρήστης θα βαθμολογήσει υψηλά το αντικείμενο που του συστήθηκε). Οι μετρικές Ανάκτησης Πληροφοριών μετρούν πόσο το $\Sigma\Sigma$ είναι ικανό να δημιουργήσει σχετικές με τον χρήστη συστάσεις, συγκρίνοντας την λίστα συστάσεων με τα αποδεκτά δεδομένα για τις προτιμήσεις του χρήστη. Η σχετικότητα ενός αντικειμένου μπορεί να καθοριστεί με τους εξής τρόπους: Εάν η βαθμολόγηση είναι

έμμεση στο σύστημα και αναπαριστάται μέσω του κλικ στο αντικείμενο, τότε το αντικείμενο θα θεωρείται σχετικό εάν ο χρήστης κάνει όντως κλικ πάνω του. Από την άλλη εάν η βαθμολόγηση είναι άμεση και αναπαρίσταται από μια κλίμακα, τότε ορίζεται ένα κατώφλι όπου βάσει αυτού κρίνεται εάν το αντικείμενο είναι σχετικό ή όχι (πχ σε μια κλίμακα 1 έως 5 μπορεί το κατώφλι να οριστεί το 3 και επομένως ένα αντικείμενο είναι σχετικό εάν βαθμολογηθεί πάνω από 3).

Έστω ότι ο αλγόριθμος συστήνει μια λίστα s από αντικείμενα στον χρήστη, και το σύνολο των συστάσεων ορίζεται ως $L(s)$. Έστω επίσης R το σύνολο των σχετικών αντικειμένων για τον χρήστη. Ορίζονται οι παρακάτω δύο μετρικές:

- *Ακρίβεια (Precision)*: Μετρά το ποσοστό των σχετικών αντικειμένων εντός της λίστας σύστασης ως

$$P = \frac{|L(s) \cap R|}{|L(s)|}$$

- *Ανάκληση (Recall)*: Μετρά το ποσοστό των σχετικών αντικειμένων εντός της λίστας σύστασης σε σχέση με το σύνολο όλων των σχετικών αντικειμένων:

$$R = \frac{|L(s) \cap R|}{|R|}$$

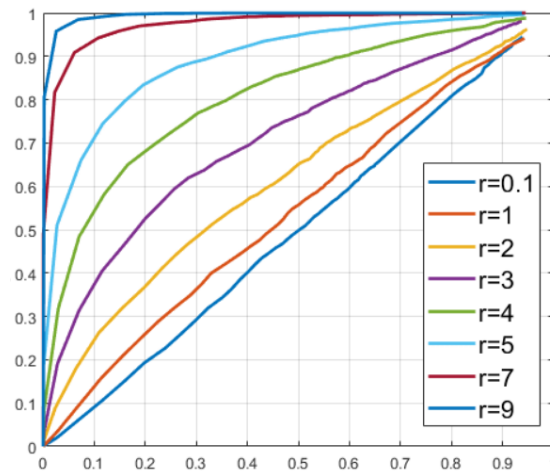
Παράδειγμα: Έστω το σύνολο των σχετικών αντικειμένων $R(u) = \alpha, \beta, \gamma, \delta, \epsilon$ και μια λίστα σύστασης με τα αντικείμενα $L = \alpha, \beta, \kappa$ τότε η ακρίβεια της λίστας είναι $2/3$ και η ανάκληση $2/5$.

Οι βαθμολογίες της ακρίβειας και της ανάκλησης αθροίζονται για όλους τους χρήστες έτσι ώστε να μετρηθεί η συνολική επίδοση του συστήματος σύστασης.

Η Ακρίβεια και Ανάκληση μπορούν πολλές φορές να είναι αντιφατικές μεταξύ τους διότι αυξάνοντας το μέγεθος της λίστας σύστασης αυξάνεται η Ανάκληση ενώ ταυτόχρονα μειώνεται η Ακρίβεια. Μια λύση για την αξιολόγηση της ισορροπίας μεταξύ των δύο είναι ο υπολογισμός του:

$$F1 = 2 \times \frac{P \times R}{P + R}$$

Τέλος, έστω ότι το A αναπαριστά το σύνολο όλων των διαθέσιμων αντικειμένων στο σύστημα, υπάρχει η Χαρακτηριστική Καμπύλη Λειτουργίας Δέκτη (Receiver Operating Characteristic Curve - ROC) η οποία παρουσιάζει την σχέση ανάμεσα στον Ψευδώς Θετικό Δείκτη (False Positive Rate) $FPR = \frac{|L(s) \setminus R|}{|A \setminus R|}$ (όπου αναπαριστά το λόγο ανάμεσα στα μη σχετικά αντικείμενα τα οποία προτείνονται σε σχέση με όλα τα μη σχετικά αντικείμενα) και στον Αληθώς Θετικό Δείκτη (True Positive Rate) $TPR = |L(s) \cap R|/|R|$ (δηλαδή την Ανάκληση - Recall). Μέσω της σταδιακής αυξομείωσης του μεγέθους της λίστας σύστασης, μπορούν να υπολογίζονται αυτές οι δύο τιμές και να καταγραφεί η καμπύλη ROC με τον FPR στον οριζόντιο άξονα και τον TPR στον κάθετο.



Εικόνα 4: Καμπύλη ROC
Πιθανότητα Ψευφώς Θετικού (οριζόντια)
Πιθανότητα Αληθώς Θετικού (κάθετα)

Κριτήριο Ακρίβειας Βαθμολογίας

Στις πραγματικές εφαρμογές ΣΣ ο χρήστης λαμβάνει μια λίστα από τα πρώτα k αντικείμενα που έχουν εκτιμηθεί με την μέγιστη βαθμολόγηση από το σύστημα. Γενικότερα, τα αντικείμενα χρειάζεται να έχουν μια βαθμολογία σχετική με τις προτιμήσεις του χρήστη, και όσο πιο σχετικά είναι τόσο υψηλότερη βαθμολογία έχουν. Για να μετρηθεί πόσο καλά το ΣΣ μπορεί να κάνει αυτήν την βαθμολόγηση

χρησιμοποιούνται μετρικές που ορίζουν πόσο κοντά είναι μια βαθμονομημένη λίστα συστάσεων με την πραγματική σειρά προτιμήσεων του χρήστη.

Για όλες τις παρακάτω μετρικές ορίζεται ο χρήστης u στον οποίο συστήνεται η λίστα αντικειμένων i_1, \dots, i_l μεγέθους l . Έστω $r_{u,i}^*$ είναι η πραγματική αξιολόγηση του χρήστη σχετικά με το αντικείμενο i . Επιπλέον υποθέτουμε ότι οι βαθμολογήσεις είναι εντός του εύρους 0 και r_{max} .

- *Συντελεστής Kendall*: Υπολογίζεται αρχικά ο συντελεστής $C(i,j)$ ο οποίος είναι ίσος με $+1$ εάν το αντικείμενο i είναι στην ίδια σχετική θέση με το αντικείμενο j (σύμφωνο ζεύγος) στις δύο λίστες (σύσταση και πραγματική) και -1 διαφορετικά (ασύμφωνο ζεύγος). Εάν τα αντικείμενα ισοβαθμούν τότε $C(i,j) = 0$. Ο συντελεστής Kendall για τον χρήστη u υπολογίζεται ως ο αριθμός των σύμφωνων ζευγαριών μείον τον αριθμό των ασύμφωνων ζευγαριών προς το αριθμό όλων των ζευγαριών (l ανά 2)

$$\tau_u = \frac{\sum_{i < j} C(i,j)}{l(l-1)/2}$$

- *Συντελεστής Spearman*: Ο συντελεστής Spearman είναι όμοιος με το συντελεστή συσχέτισης Pearson εφαρμοσμένος στις βαθμονομημένες μεταβλητές (μεταξύ λίστας σύστασης και πραγματικής). Το εύρος της υπολογισμένης τιμής είναι από -1 έως $+1$ όπου το 1 σημαίνει πως οι λίστες είναι πανομοιότυπες.
- *Normalized Discounted Cumulative Gain (NDCG)*:

Ο συντελεστής NDCG εστιάζει περισσότερο στην βαθμολογική συσχέτιση των αντικειμένων που βρίσκονται πιο πάνω στην λίστα βασιζόμενος στο επιχείρημα ότι οι χρήστες ασχολούνται περισσότερο με αυτά τα αντικείμενα.

$$DCG = \sum_{r=1}^l \frac{2^{r_{u,i_r}^*} - 1}{\log_2(r+1)}$$

$$NDCG = \frac{DCG}{DCG^*}$$

όπου ο DCG^* είναι ο καλύτερος δυνατός DCG.

- *Μέσος Συντελεστής Αμοιβαιότητας (ΜΣΑ)*: Ο ΜΣΑ ή Mean Reciprocal Rank (MRR) συνήθως ορίζεται για δυαδικές συσχετίσεις (η βαθμολόγηση είναι 0 ή 1), αλλά μπορεί να οριστεί και για βαθμολογήσεις κλίμακας. Η τιμή του ΜΣΑ υποδηλώνει την αμοιβαιότητα της βαθμολογίας που προκύπτει από το ΣΣ με την βαθμολογία σύγκρισης για το πρώτο σχετικό αντικείμενο που θα εντοπιστεί. ΜΣΑ με τιμή 1 σημαίνει ότι το πρώτο αντικείμενο είναι σχετικό με την λίστα σύστασης. Για N αριθμό χρηστών ορίζεται ως:

$$MRR = \frac{1}{N} \sum_{u=1}^N \frac{1}{rank_u}$$

όπου $rank_u$ είναι η θέση της βαθμολογίας του πρώτου σχετικού αντικειμένου για τον χρήστη u. Για παράδειγμα έστω ΣΣ με τρεις χρήστες και λίστες σύστασης μεγέθους 5. Σε δυαδική βαθμολόγηση έχουμε για κάθε χρήστη τις λίστες $\{0, 0, 1, 1, 0\}$, $\{0, 0, 0, 1, 0\}$, $\{1, 1, 0, 1, 0\}$, και επομένως το πρώτο σχετικό αντικείμενο για τον κάθε ένα είναι στις θέσεις 3, 4 και 1 αντίστοιχα. Άρα $ΜΣΑ = \frac{1}{3} \sum_{u=1}^3 \frac{1}{rank_u} = \frac{1}{3}(\frac{1}{3} + \frac{1}{4} + \frac{1}{1}) = 0.527$

- *Αναμενόμενος Συντελεστής Αμοιβαιότητας (ΑΣΑ)*:

Αντίστοιχος με τον ΜΣΑ, αλλά με πιο σύνθετο τύπο, ο ΑΣΑ ή Expected Reciprocal Rank (ERR) κάνει την υπόθεση ότι ο χρήστης ελέγχει κάθε αντικείμενο με την σειρά (σε αλληλουχία) μέχρι να εντοπίσει το κατάλληλο όπου και θα σταματήσει να ελέγχει περαιτέρω.

$$ERR = \sum_{r=1}^l \frac{1}{r} p(u, i_r) \prod_{s=1}^{r-1} (1 - p(u, i_s))$$

$$p(u, i) = \frac{2^{r_{u,i}^*} - 1}{2^{r_{max}}}$$

Η $p(u, i)$ είναι η πιθανότητα ο χρήστης να ικανοποιηθεί από το αντικείμενο i. Ο ΑΣΑ είναι η αναμενόμενη θέση όπου θα εντοπίσει ένας χρήστης σχετικό αντικείμενο και θα σταματήσει να ψάχνει παρακάτω. Το γινόμενο στην εξίσωση αντιπροσωπεύει την πιθανότητα ο χρήστης να μην έχει ικανοποιηθεί μέχρι και το αντικείμενο r. Για παράδειγμα, έστω μια λίστα μεγέθους 3 που συστήνεται

στην u σε βαθμολόγηση κλίμακας 1 έως 5 ($r_{max} = 5$). Ο χρήστης δίνει βαθμολόγηση $\{3,5,2\}$. Ο παρακάτω πίνακας δίνει τον τρόπο υπολογισμού της πιθανότητας $p(u,i)$ ώστε να υπολογιστεί ο $\Lambda\Sigma A = \frac{1}{1}p_1 + \frac{1}{2}p_2 + \frac{1}{3}p_3 = 0.598$

r	$1/r$	r_{u,i_r}^*	$p(u,i)$	πιθανότητα να σταματήσει στο r
1	1/1	3	7/32	$p_1 = 7/32$
2	1/2	5	31/32	$p_2 = 31/32(1 - 7/32)$
3	1/3	2	3/32	$p_3 = 3/32(1 - 31/32)(1 - 7/32)$

Άλλα Κριτήρια Μέτρησης

Τα προηγούμενα 3 κριτήρια μέτρησης είναι σημαντικά για την αξιολόγηση ενός $\Sigma\Sigma$ συγκρίνοντας μαθηματικά διαφορετικές τεχνικές σύστασης, αλλά δεν εστιάζουν στο πώς ένας χρήστης αντιλαμβάνεται τις συστάσεις που παράγονται. Υπάρχουν διάφορες μετρικές που δίνουν την δυνατότητα να μετρήσουν κάτι τέτοιο οι οποίες είναι πολύ πιο σύνθετες και δίνονται με διάφορους τύπους στην βιβλιογραφία.

- *Καινοτομία*: Αφορά την δυνατότητα του $\Sigma\Sigma$ να συστήνει αντικείμενα τα οποία δεν είναι γνωστά στον χρήστη [21, 143, 119]
- *Ποικιλία*: Αφορά την δυνατότητα του $\Sigma\Sigma$ να προτείνει αντικείμενα τα οποία διαφέρουν μεταξύ τους. Συχνά ένα $\Sigma\Sigma$ μπορεί να προτείνει όμοια αντικείμενα στον χρήστη (πχ βιβλία του ίδιου συγγραφέα), ενώ ένα καλά σχεδιασμένο $\Sigma\Sigma$ θα έπρεπε να συστήνει αντικείμενα που καλύπτουν το σύνολο των προτιμήσεων του. [144, 138]
- *Ενδιαφέρον*: Αφορά την ικανότητα του $\Sigma\Sigma$ να προτείνει αντικείμενα που θα εκπλήξουν τον χρήστη ενώ ταυτόχρονα να είναι εντός των προτιμήσεων του. Δηλαδή οι προτάσεις που γίνονται να μην είναι πάντα οι προφανείς. [38, 82]
- *Πλήρης Κάλυψη*: Μετρά το ποσοστό των στοιχείων (χρήστες, αντικείμενα ή αξιολογήσεις) για το οποίο το $\Sigma\Sigma$ κάνει συστάσεις. Για παράδειγμα η κάλυψη χρηστών μετρά πόσο μεγάλο είναι το σύνολο των χρηστών για τους οποίους το $\Sigma\Sigma$ κάνει συστάσεις, ενώ η κάλυψη καταλόγου μετρά το ποσοστό ζευγαριών

χρήστης-αντικείμενο για το οποίο γίνεται πρόβλεψη σε σχέση με το σύνολο δυνατών ζευγαριών. Μια τέτοια μετρική μπορεί να βοηθήσει να εντοπιστεί μια πιθανή μεροληψία του συστήματος απέναντι σε κάποιο σύνολο στοιχείων.[38]

- *Εγκατάλειψη*: Χρησιμοποιείται όταν μια λίστα αντικειμένων συστήνεται σε χρήστη. Θεωρείται ότι ο χρήστης «εγκαταλείπει» το σύστημα όταν η λίστα αυτή δεν περιέχει κανένα αντικείμενο που σχετίζεται με τις προτιμήσεις του. Η μετρική εγκατάλειψης είναι ουσιαστικά η πιθανότητα ο χρήστης να φύγει από το σύστημα κατά την διάρκεια μιας επίσκεψης και φυσικά πρέπει να ελαχιστοποιείται. [95, 59]

Πλαίσια Αξιολόγησης ΣΣ

Δεδομένων κάποιων συμφωνημένων μετρικών αξιολόγησης ενός ΣΣ, στην συνέχεια στόχος είναι να στηθεί το κατάλληλο περιβάλλον έτσι ώστε να μπορούν να γίνουν οι μετρήσεις και να εξαχθούν οι τιμές των μετρικών που θα χρησιμοποιηθούν. Υπάρχουν τρεις κατηγορίες εντός των οποίων εμπίπτει η μεθοδολογία αξιολόγησης: Η online, η offline και τα user studies.

Online Αξιολόγηση

Η Online Αξιολόγηση είναι εφικτή όταν υπάρχει ένα σύστημα που είναι ενεργό ήδη. Σε αυτήν την περίπτωση η αξιολόγηση του ΣΣ αξιοποιεί τις βαθμολογήσεις πραγματικών χρηστών που χρησιμοποιούν έτσι και αλλιώς το σύστημα. Συνήθως δημιουργείται ένα πρωτότυπο ΣΣ το οποίο παρουσιάζεται σε ένα τυχαία διαμορφωμένο σύνολο χρηστών του συστήματος. Στην συνέχεια μέσω της ανατροφοδότησης και της βελτίωσης του σχεδιασμού οι επόμενες εκδόσεις του ΣΣ παρουσιάζονται σε διαφορετικό σύνολο, διατηρώντας το παλαιότερο ΣΣ στο αρχικό σύνολο χρηστών. Με αυτόν τον τρόπο τα δεδομένα που συλλέγονται συγκρίνονται ανάμεσα στις διαφορετικές εκδόσεις του ΣΣ και επομένως αποφεύγονται οι αλλαγές με δραστική επιρροή σε ολόκληρο το σύστημα. Έτσι δεν υπάρχει μεγάλο ρίσκο να

δημιουργηθούν σοβαρά προβλήματα για το σύστημα σε περίπτωση που μια έκδοση του ΣΣ είναι ελαττωματική ή δεν λειτουργήσει όπως ήταν αναμενόμενο. Η μέθοδος αυτή αναφέρεται στην βιβλιογραφία ως A/B test. Έστω ότι έχουμε δύο διαφορετικούς αλγόριθμους σύστασης: οι χρήστες χωρίζονται σε δύο σύνολα A και B και κάθε σύνολο έρχεται σε επαφή με έναν μόνο από τους δύο αλγόριθμους. Έτσι στο τέλος της δοκιμαστικής περιόδου οι αλγόριθμοι αξιολογούνται για το σύνολο τους.

Μια άλλη οπτική είναι η αξιοποίηση αλγορίθμων Multi-Armed-Bandit (MAB) οι οποίοι βρίσκουν τον βέλτιστο αλγόριθμο από ένα σύνολο αλγορίθμων. Η βασική ιδέα είναι να επιλέγεται ένας αλγόριθμος σύστασης κάθε φορά που κάποιος χρήστης ζητά κάποια σύσταση. Η επιλογή αυτή μπορεί να είναι μια επιλογή διερεύνησης όπου ο αλγόριθμος επιλέγεται τυχαία ή η επιλογή του βέλτιστου μέχρι τώρα αλγορίθμου. Η αξιοποίηση των αλγορίθμων MAB αναλύεται περισσότερο στο κεφάλαιο "MAB σε Συστήματα Συστάσεων" της παρούσας διπλωματικής εργασίας.

Το πρόβλημα με την online αξιολόγηση είναι πως τα εμπορικά ΣΣ δεν είναι δημοσίως προσβάσιμα και ο έλεγχος τους μπορεί να γίνει μόνο από ανθρώπους που έχουν πρόσβαση. Αυτό υποδηλώνει ότι τα τεστ αυτά δεν μπορούν να αναπαραχθούν από άλλους ερευνητές. Ένα άλλο πρόβλημα επίσης είναι χρειάζεται ένας πολύ υψηλός αριθμός χρηστών για να συγκριθούν διαφορετικοί αλγόριθμοι και να γίνει η αξιολόγηση τους.

User Studies

Σε περίπτωση που η online αξιολόγηση δεν μπορεί να διεξαχθεί διότι είναι είτε πολύ επικίνδυνη είτε γιατί δεν υπάρχει πρόσβαση σε ένα πραγματικό σύστημα, τα user studies είναι μια πιθανή λύση έτσι ώστε να ληφθούν βαθμολογήσεις από πραγματικούς χρήστες. Σε αυτήν την περίπτωση, συγκεντρώνεται ένα μικρό σύνολο πραγματικών χρηστών από το οποίο ζητείται να αλληλεπιδράσει με το υπό δοκιμή ΣΣ. Η διαδικασία γίνεται υπό ελεγχόμενο περιβάλλον και με αυτόν τον τρόπο οι ερευνητές μπορούν να συλλέγουν δεδομένα από τις αλληλεπιδράσεις των χρηστών και ύστερα να ζητάνε

από τους ίδιους να βαθμολογήσουν την εμπειρία τους. Με αυτόν τον τρόπο είναι εφικτό να εξαχθούν μη ποσοτικές πληροφορίες και έτσι το ΣΣ να αξιολογηθεί με περισσότερα κριτήρια. Ένα πρόβλημα με τα user studies είναι πως είναι πιθανό να υπάρχει ένας βαθμός μεροληψίας δεδομένου του μικρού μεγέθους χρηστών και της αντίληψης των χρηστών ότι συμμετέχουν σε μια τέτοια διαδικασία. Για αυτόν τον λόγο τα αποτελέσματα ενός user study θα πρέπει να ελέγχονται προσεκτικά για να θεωρηθούν αξιόπιστα.

Offline Αξιολόγηση

Η offline αξιολόγηση σε αντίθεση με την online συγκρίνει και βαθμολογεί ΣΣ αλγόριθμους αξιοποιώντας δεδομένα που έχουν είτε συλλεγεί ως οι πραγματικές αλληλεπιδράσεις χρηστών με κάποια υπηρεσία είτε σχεδιαστεί καταλλήλως μέσω κάποιας διαδικασίας προσομοίωσης. Η offline αξιολόγηση είναι η πιο δημοφιλής μέθοδος καθώς οι δοκιμαστικές μέθοδοι, οι μετρικές και τα δεδομένα μπορούν να κατασκευαστούν και να επαναχρησιμοποιηθούν. Καθώς δεν υπάρχει πραγματική αλληλεπίδραση ανάμεσα στο ΣΣ και τους χρήστες, γίνεται η υπόθεση ότι η συμπεριφορά των χρηστών βάσει των δεδομένων που έχουν συλλεχθεί θα παραμείνει όμοια για όταν το ΣΣ θα λειτουργήσει στην πραγματικότητα. Τα δεδομένα συνήθως διαχωρίζονται σε τρία μέρη: στο σύνολο εκπαίδευσης (training set), το οποίο αξιοποιείται για να χτιστεί το μοντέλο του ΣΣ, το σύνολο επικύρωσης (validation set) το οποίο αξιοποιείται για τη ρύθμιση των παραμέτρων του ΣΣ και στο σύνολο δοκιμής ή αξιολόγησης (testing or evaluation set) το οποίο αξιοποιείται για να αξιολογηθεί το τελικώς διαμορφωμένο ΣΣ. Ο διαχωρισμός αυτών των συνόλων γίνεται συνήθως με τυχαίο τρόπο βασισμένο σε ποσοστά (πχ το σύνολο εκπαίδευσης να είναι το 70%, της επικύρωσης το 10% και της δοκιμής το 20%). Με αυτόν τον τρόπο αποφεύγεται η εκπαίδευση και η αξιολόγηση του συστήματος πάνω στο ίδιο σύνολο δεδομένων, κάτι το οποίο δεν θα είχε αξία. Επίσης ένας διαφορετικός τρόπος διαχωρισμού των δεδομένων εκπαίδευσης και αξιολόγησης μπορεί να είναι η χρονική τους συσχέτιση, για παράδειγμα πριν από μια ημερομηνία να αξιοποιηθούν τα δεδομένα ως το σύνολο εκπαίδευσης και μετά από

αυτήν να αξιοποιηθούν ως το σύνολο αξιολόγησης.

Πολλά πειράματα κάνουν την αξιολόγηση σε πλαίσιο που αφορά τον χρήστη. Για κάθε χρήστη, επιλέγεται ένας αριθμός n αντικειμένων τα οποία θα απαρτίζουν το σύνολο εκπαίδευσης και τα υπόλοιπα θα απαρτίζουν το σύνολο αξιολόγησης. Σε αυτή την προσέγγιση για χαμηλές τιμές του n είναι δυνατή η αξιολόγηση του ΣΣ στα πλαίσια της αργής εκκίνησης όπου η πληροφορία για κάθε χρήστη είναι ελάχιστη.

Επιπλέον η αξιολόγηση ΣΣ μέσω MAB αλγορίθμων στο offline πλαίσιο εισήχθη από το *Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms* [69]

Η offline αξιολόγηση αν και η πιο δημοφιλής μέθοδος αξιολόγησης συστημάτων σύστασης, παρουσιάζει τα παρακάτω μειονεκτήματα:

1. Δεν υπάρχει βεβαιότητα ότι τα αποτελέσματα θα μπορούν να αντιστοιχηθούν με τα αποτελέσματα ενός ΣΣ εφαρμοσμένο σε ένα πραγματικό σύστημα. Ειδικά για την περίπτωση των προσομοιωμένων δεδομένων, η πραγματική συμπεριφορά των χρηστών είναι πιθανόν να διαφέρει αρκετά αφού κάθε διαδικασία προσομοίωσης βασίζεται σε υποθέσεις που εν τέλει μπορεί να μην ισχύουν.
2. Υπάρχουν πολλές μετρικές που αξιολογούν ένα ΣΣ offline βάσει της παλαιότερης εικόνας προτιμήσεων των χρηστών. Κάτι τέτοιο καθιστά αδύνατη την επιτυχημένη εκτίμηση την πραγματικής επιρροής του ΣΣ στο τωρινό πραγματικό σύστημα.
3. Η αντιμετώπιση των βαθμολογήσεων των χρηστών που λείπουν έχει πάντα ρίσκο. Δεν είναι ποτέ σαφές εάν οι χρήστες δεν έχουν δώσει κάποια βαθμολόγηση επειδή δεν ενδιαφέρθηκαν, δεν ήθελαν ή δεν γνώριζαν το εκάστοτε αντικείμενο.

Τα παραπάνω μειονεκτήματα μπορούν μερικώς να εξαλειφθούν εάν συνδυαστεί η offline αξιολόγηση με την online, δηλαδή ο αλγόριθμος να δοκιμαστεί online έτσι ώστε να απορριφθούν ή να επιβεβαιωθούν τα αποτελέσματα που έχουν αποκτηθεί μέσω της offline αξιολόγησης.

Αξιολόγηση βάσει Διάταξης

Τα μοντέλα που αντιμετωπίζουν το πρόβλημα της σύστασης χρησιμοποιώντας τεχνικές πρόβλεψης στοχεύουν στην ελαχιστοποίηση του σφάλματος (πχ τετραγωνικού) της εκτίμηση των κενών βαθμολογήσεων. Παρόλα αυτά, η προσέγγιση αυτή διαθέτει ένα σημαντικό ελάττωμα, καθώς σε ένα πραγματικό ΣΣ μόνο τα πρώτα k αντικείμενα παρουσιάζονται στον χρήστη και δεν έχει κάποια αξία το ΣΣ να διαθέτει μεγάλη ευστοχία εκτίμησης για αντικείμενα που δεν πρόκειται να παρουσιαστούν.

Ένα παράδειγμα που εξηγεί γιατί η ευστοχία εκτίμησης δεν ικανοποιεί πάντα τον στόχο ενός ΣΣ είναι να φανταστούμε ένα ΣΣ να κάνει πάρα πολύ καλές εκτιμήσεις για αντικείμενα με μέτρια ή χαμηλή βαθμολογία, αλλά κακές εκτιμήσεις για αντικείμενα που διαθέτουν υψηλές βαθμολογίες. Η μέση επίδοση ενός τέτοιου συστήματος δεν θα αξιολογείται χαμηλά καθώς μόνο τα αντικείμενα με υψηλές βαθμολογίες θα έχουν σημαντικό σφάλμα τα οποία δεν θα είναι και η πλειοψηφία. Παρόλα αυτά από την οπτική του χρήστη ένα τέτοιο ΣΣ θα δημιουργούσε μεγάλη ζημιά στην εμπειρία του με το σύστημα καθώς δεν θα του συστήνονταν τα καταλληλότερα αντικείμενα για τις προτιμήσεις του. Εξαιτίας αυτού υπάρχουν προσεγγίσεις όπου εστιάζουν στην βελτιστοποίηση της κατασκευής της λίστας των πρώτων k αντικειμένων η οποία και συστήνεται στον χρήστη. Αυτές οι προσεγγίσεις χωρίζονται στις παρακάτω κατηγορίες:

- *Βελτιστοποίηση κατά Σημεία:* Είναι η απλούστερη μορφή αυτών των μεθόδων όπου διαθέτουν τεχνικές που προσπαθούν να βελτιστοποιήσουν μετρικές ευστοχίας εκτίμησης και διατάσσουν τα αντικείμενα βάσει αυτών των προβλέψεων. Ως συνέπεια η βελτιστοποίηση κατά σημεία δεν στοχεύει άμεσα την βελτιστοποίηση της ταξινόμησης.
- *Βελτιστοποίηση κατά Ζεύγη:* Αντίστοιχα με την κατά σημεία αυτές οι μέθοδοι εστιάζουν σε ζεύγη αντικειμένων για την βελτιστοποίηση της ταξινόμησης αντικειμένων. Για κάθε χρήστη u ένα ζεύγος (i,j) αντικειμένων χρησιμοποιείται ως δεδομένο εκπαίδευσης μόνο εάν υπάρχουν βαθμολογήσεις

και για τα δύο αντικείμενα. Κάθε ζεύγος δίνει την πληροφορία για το ποιο από τα δύο αντικείμενα είναι πιο κοντά στις προτιμήσεις του χρήστη.

- *Βελτιστοποίηση κατά Λίστα*: Αντίστοιχα με την κατά ζεύγη βελτιστοποίηση, σε αυτήν την περίπτωση μεθόδων χρησιμοποιούνται ολόκληρες λίστες αντικειμένων όπου η βαθμολόγηση τους είναι γνωστή για όλα τα αντικείμενα.

Κατακλείδα

Στο κεφάλαιο αυτό δόθηκε μια περιγραφή των Συστημάτων Σύστασης, κάποιοι τυπικοί ορισμοί και αναφέρθηκαν τα πλεονεκτήματα και οι προκλήσεις τους. Επίσης έγινε μια ανασκόπηση των βασικών και σημαντικότερων τεχνικών σύστασης που παρουσιάζονται στην βιβλιογραφία καθώς και οι τρόποι που μπορεί να γίνει η αξιολόγηση ενός ολόκληρου συστήματος σύστασης όσον αφορά την ποιότητα του. Εντός των υποενοτήτων έγινε μια επιδερμική αναφορά στους MAB αλγορίθμους και πως ενδέχεται να εμπλέκονται με τα συστήματα σύστασης. Το επόμενο κεφάλαιο αφιερώνεται συγκεκριμένα στην έννοια των αλγορίθμων MAB και γενικότερα στο πρόβλημα αναζήτησης – αξιοποίησης χωρίς την εμπλοκή των συστημάτων σύστασης. Η επί του πρακτέου συσχέτιση μεταξύ των δύο εννοιών (Συστήματα Συστάσεων και MAB) θα ακολουθήσει στο τρίτο κεφάλαιο.

Αλγόριθμοι Multi-Armed Bandit

Εισαγωγή και Ορισμοί

Το πρόβλημα Multi-Armed Bandit (MAB) είναι ένα πρόβλημα της θεωρίας πιθανοτήτων και μηχανικής μάθησης και αφορά την αναζήτηση μιας βέλτιστης στρατηγικής αποφάσεων που μεγιστοποιεί μια συνάρτηση ωφέλειας, δεδομένης της έλλειψης πληροφόρησης για την ωφέλεια της κάθε απόφασης.

Η λήψη αποφάσεων υπό συνθήκες αβεβαιότητας εντάσσεται στα πλαίσια της ονομαζόμενης *Στοχαστικής Βελτιστοποίησης*. Η αβεβαιότητα μπορεί να αφορά όπως αναφέρθηκε σε οφέλη ή κόστη που μπορεί να επιφέρει η κάθε απόφαση.

Έστω ότι χρειάζεται να παρθεί μια απόφαση ανάμεσα σε δύο επιλογές. Η επιλογή A έχει p_A πιθανότητα να έχει όφελος K_A διαφορετικά μηδενικό ενώ η επιλογή B έχει πιθανότητα p_B να έχει όφελος K_B διαφορετικά μηδενικό. Η διαδικασία αυτή έχει την δυνατότητα να επαναλαμβάνεται και επομένως σε κάθε επανάληψη είναι εφικτό να λαμβάνεται διαφορετική απόφαση από την προηγούμενη. Η στρατηγική (ή πολιτική) αφορά την σειρά (ή το δέντρο) των αποφάσεων που πρέπει να παρθούν σε κάθε επανάληψη βάσει πληροφοριών που ενδέχεται να αποκαλύπτονται μετά από κάθε επιλογή. Εάν οι πιθανότητες p_A, p_B είναι γνωστές τότε το πρόβλημα απλοποιείται και η στρατηγική που μπορεί να ακολουθηθεί μπορεί να εντοπίσει άμεσα την βέλτιστη επιλογή βάσει κάποιου κριτηρίου βελτιστοποίησης. Για παράδειγμα, ένα κριτήριο είναι η μέση ωφέλεια $M_A = p_A \times K_A$, $M_B = p_B \times K_B$. Εάν $M_A > M_B$ προφανώς η βέλτιστη στρατηγική είναι η επιλογή A. Στην συνέχεια της παρούσας εργασίας οποτεδήποτε αναφερόμαστε σε μέγιστο κέρδος ή ελάχιστο κόστος θα υπονοείται πάντα ότι έχει οριστεί και κάποιο κριτήριο βελτιστοποίησης.

Το πρόβλημα MAB αφορά την έλλειψη της πληροφορίας των p_A , p_B και επομένως αναζητά έναν τρόπο για να εξαχθούν (ή προσεγγισθούν) αυτές οι πιθανότητες ενώ ταυτόχρονα να μεγιστοποιείται η συνολική ωφέλεια.

Το όνομα Multi-Armed Bandit έχει παρθεί από τα τυχερά παιχνίδια όπου ο παίκτης κάνει μια επιλογή χωρίς να γνωρίζει με ποια πιθανότητα μπορεί να κερδίσει (Κουλοχέρης - One-Armed Bandit). Η λέξη Multi-Armed (πολλαπλά χέρια) αφορά τις δυνατές επιλογές που μπορεί να έχει ο παίκτης, δηλαδή τις πιθανές αποφάσεις που μπορεί να πάρει εντός τους παιχνιδιού.

Εφόσον με κάποιο τρόπο οι πιθανότητες p_A , p_B μπορούν να προσεγγισθούν τότε γίνεται σαφές ποια από τις δύο επιλογές είναι εκείνη που με βάση κάποιο κριτήριο βελτιστότητας αποφέρει το μέγιστο κέρδος. Εάν για παράδειγμα δοκιμαζόταν η επιλογή A πολλές φορές θα μπορούσε να προσεγγιστεί η p_A και αντιστοίχως η p_B . Παρόλα αυτά, τα παρακάτω στοιχεία κάνουν το πρόβλημα ενδιαφέρον:

- Το πλήθος δοκιμών ή αποφάσεων που μπορούν να παρθούν είναι πεπερασμένο.
- Το κέρδος (ωφέλεια) της κάθε απόφασης εξαρτάται από μια κατανομή (άγνωστη) πιθανότητας.
- Εάν αντί για ωφέλεια αντιστοιχίσουμε το αποτέλεσμα της κάθε απόφασης σε κόστος τότε μπορεί να υπάρχει κατώφλι συνολικού κόστους όπου εάν ξεπεραστεί η διαδικασία πρέπει να σταματήσει.
- Η κατανομή της ωφέλειας (ή τους κόστους) της κάθε απόφασης μπορεί να μην είναι σταθερή στον χρόνο.

Τα παραπάνω μπορεί να υπάρχουν ως συνθήκες για το εκάστοτε πρόβλημα οπότε διαφοροποιούν και τον τρόπο μπορεί να αντιμετωπιστεί.

Στα πλαίσια της μηχανικής μάθησης, το πρόβλημα εμπίπτει στην κατηγορία της ενισχυτικής μάθησης (reinforcement learning) και αφορά το δίλημμα αναζήτησης-αξιοποίησης (exploration-exploitation dilemma). Αυτό σημαίνει ότι κάθε στιγμή

που πρόκειται να παρθεί μια απόφαση υπάρχει το εξής δίλημμα: Είτε η απόφαση που θα παρθεί θα αξιοποιήσει την επιλογή με την μέγιστη ωφέλεια που έχει εκτιμηθεί μέχρι τώρα (καλύτερη επιλογή), είτε θα διαλεχτεί μια άλλη επιλογή (όχι η καλύτερη) με στόχο την εξαγωγή επιπλέον πληροφορίας (αναζήτηση) όπου μπορεί να βοηθήσει να εντοπιστεί μια άλλη ενδεχόμενη καλύτερη επιλογή. Έτσι το πρόβλημα μπορεί να περιγραφεί με τον εξής τρόπο: Πως θα μπορούσαμε να εντοπίσουμε γρηγορότερα την επιλογή που δίνει την μέγιστη ωφέλεια;

Μαθηματικός Φορμαλισμός MAB

Έστω N ανεξάρτητοι στατιστικοί πληθυσμοί. Σε κάθε περίοδο επιλέγεται ένας πληθυσμός i όπου:

Όφελος: $X_i \sim f_i(\cdot|\theta_i)$ (Κατανομή Πιθανότητας με παράμετρο θ_i)

$\mu_i = E(X_i|\theta_i)$ (Μέσο Όφελος)

$\underline{\theta} = (\theta_1, \dots, \theta_N)$, όπου θ_i , $i=1, \dots, N$ άγνωστες παράμετροι

$\theta_i \in \Theta_i$, $\underline{\theta} \in \Theta = \prod_{i=1}^N \Theta_i$

Σκοπός: Να βρεθεί μια πολιτική επιλογής/δειγματοληψίας για την οποία μεγιστοποιείται το αναμενόμενο συνολικό όφελος (ασυμπτωτικά).

Στην συνέχεια οι όροι επιλογή και πληθυσμός όσον αφορά το πρόβλημα MAB θα χρησιμοποιούνται ως η ίδια έννοια που θα αφορά την απόφαση της εκάστοτε πολιτικής (choosing an arm).

Έστω A_t ο πληθυσμός που θα επιλεγεί την περίοδο t και R_t η ωφέλεια (κέρδος) που θα αποκτηθεί την ίδια περίοδο.

t -ιστορία: $H_t = (A_1, R_1, \dots, A_{t-1}, R_{t-1})$ όλες οι προηγούμενες επιλογές πληθυσμών μαζί με το κέρδος από την πρώτη έως και την ακριβώς προηγούμενη περίοδο

Προσαρμοστική πολιτική $\pi_t(h_t)$: πληθυσμός που επιλέγεται κατά την περίοδο t , δοθείσας της ιστορίας h_t .

$S_n^\pi(\underline{\theta}) = E_{\underline{\theta}}^\pi(\sum_{t=1}^n R_t)$: Μέσο συνολικό κέρδος

Κριτήριο βελτιστοποίησης: κάτω από κατάλληλο ασυμπτωτικό κριτήριο (πχ μέση τιμή ανά περίοδο), η ίδια πολιτική π πρέπει να μεγιστοποιεί το $S_n^\pi(\underline{\theta})$, ανεξάρτητα από τις τιμές των άγνωστων παραμέτρων $\underline{\theta} \in \Theta$

Συνέπεια

Η μεγιστοποίηση του κέρδους για όλες τις τιμές του $\underline{\theta}$ και για όλες τις πεπερασμένες περιόδους n γενικά δεν είναι δυνατή. Πιο ρεαλιστικά είναι τα ασυμπτωτικά κριτήρια.

Έστω $\mu^*(\underline{\theta}) = \max_i(\mu_i(\theta_i))$ δηλαδή η επιλογή με την μέγιστη μέση τιμή κέρδους. Προφανώς εάν γνωρίζαμε την πραγματική τιμή του διανύσματος $\underline{\theta}$, θα παίρναμε όλα τα δείγματα από τον πληθυσμό με $\mu_i(\theta_i) = \mu^*(\underline{\theta})$

Επομένως για οποιαδήποτε π και $\underline{\theta}$

$$S_n^\pi(\underline{\theta}) \leq n\mu^*(\underline{\theta}), \text{ δηλαδή } \frac{S_n^\pi(\underline{\theta})}{n} \leq \mu^*(\underline{\theta})$$

Ορισμός: η πολιτική π είναι συνεπής αν $\lim_{n \rightarrow \infty} \frac{S_n^\pi(\underline{\theta})}{n} = \mu^*(\underline{\theta})$, για κάθε $\underline{\theta} \in \Theta$

Υπάρχουν συνεπείς πολιτικές κάτω από αρκετά γενικές συνθήκες. Ένα παράδειγμα μιας τέτοιας πολιτικής είναι οι αναγκαστικές επιλογές σε αραιές ακολουθίες περιόδων[99].

Έστω $T_n(i) = \sum_{t=1}^n 1(A(t) = i)$ η συχνότητα επιλογής του πληθυσμού i στις πρώτες n περιόδους.

Τότε $S_n^\pi(\underline{\theta}) = \sum_{i=1}^N \mu_i(\theta_i) E_{\underline{\theta}}^\pi(T_n(i))$ είναι το άθροισμα του μέσου κέρδους της κάθε επιλογής επί την πιθανότητα επιλογής βάσει της πολιτικής π

Αναγκαία και ικανή συνθήκη για την συνέπεια της πολιτικής π :

$$\lim_{n \rightarrow \infty} \frac{E_{\underline{\theta}}^\pi(T_n(i))}{n} = 0, \text{ για κάθε } i \text{ τ.ω } \mu_i(\theta_i) < \mu^*$$

δηλαδή για τις επιλογές που δεν έχουν το μέγιστο αναμενόμενο κέρδος η πιθανότητα επιλογής τους να τείνει στο μηδέν καθώς το n μεγαλώνει.

Δεδομένου ότι υπάρχουν άπειρες συνεπείς πολιτικές, είναι χρήσιμο να βρεθούν συνεπείς πολιτικές με καλύτερες ιδιότητες όπως για παράδειγμα ο ρυθμός σύγκλισης του μέσου κέρδους να μεγιστοποιείται.

Συνάρτηση απώλειας λόγω έλλειψης πληροφορίας (Regret function)

$$R_n^\pi(\underline{\theta}) = n\mu^*(\underline{\theta}) - S_n^\pi(\underline{\theta})$$

δηλαδή η διαφορά του μέσου κέρδους που θα είχαμε εάν υπήρχε πλήρης πληροφόρηση για τα $\underline{\theta}$ από το μέσο κέρδος βάσει της πολιτικής π .

Από την σχέση αυτή φαίνεται ότι ο ρυθμός αύξησης της απώλειας καθορίζει τον ρυθμό σύγκλισης της μέσης τιμής στην βέλτιστη.

$$R_n^\pi(\underline{\theta}) = \sum_{i=1}^N (\mu^*(\underline{\theta}) - \mu_i(\theta_i)) E_{\underline{\theta}}^\pi(T_n(i)) \rightarrow R > 0$$

όταν παίρνουμε δείγμα από τον μη βέλτιστο πληθυσμό

Έχουμε λοιπόν ότι η πολιτική π είναι συνεπής αν $\lim_{n \rightarrow \infty} \frac{R_n^\pi(\underline{\theta})}{n} = 0$
 πχ αν $R_n^\pi(\underline{\theta}) = o(n^{\frac{1}{2}})$

Ορισμός: Μια πολιτική π είναι ομοιόμορφα καλή αν $R_n^\pi(\underline{\theta}) = o(n^\alpha)$, $\forall \alpha > 0$, $\underline{\theta} \in \Theta$

Αποδοτικότητα

Κάτω φράγμα καλής ομοιομορφίας[66]

$$\liminf_{n \rightarrow \infty} \frac{T_n(i)}{\log n} \geq \frac{1}{I(\theta_i, \theta^*)}$$

με πιθανότητα 1 για κάθε $\underline{\theta} \in \Theta$. Επομένως,

$$\liminf_{n \rightarrow \infty} \frac{R_n^\pi(\underline{\theta})}{\log n} \geq M(\underline{\theta}) \equiv \sum_{i=1}^n \frac{(\mu^*(\underline{\theta}) - \mu_i(\theta_i))}{I(\theta_i, \theta^*)}$$

Για να έχει μια πολιτική την καλή ιδιότητα ότι για κάθε δυνατή τιμή των άγνωστων παραμέτρων $\underline{\theta} \in \Theta$, οι μη βέλτιστοι πληθυσμοί χρησιμοποιούνται σπάνια

$$E_{\underline{\theta}}^{\pi}(T_n(i)) = o(n^{\alpha}), \forall \alpha > 0$$

θα πρέπει όλοι οι πληθυσμοί να χρησιμοποιούνται αρκετά συχνά $T_n(i) \geq \frac{\log n}{I(\theta_i, \theta^*)}$ ασυμπτωτικά.

Στο παράδειγμα με τις άγνωστες πιθανότητες p_A, p_B θέλουμε μια πολιτική όπου:

- Όταν $p_A < p_B$ να παίρνει δείγματα από τον πληθυσμό A πολύ σπάνια (δηλαδή η συχνότητα $T_n(A)/n$ να αυξάνει πιο αργά από n^{α} για κάθε $\alpha > 0$)
- Αντίστοιχα αν $p_A > p_B$ η συχνότητα $T_n(B)/n$ να αυξάνει πιο αργά από n^{α} για κάθε $\alpha > 0$

Για να γίνει αυτό η πολιτική πρέπει να μπορεί να «μαθαίνει» γρήγορα ποιος είναι ο καλύτερος πληθυσμός σε κάθε περίπτωση και επομένως να παίρνει «αρκετά» δείγματα και από τους δύο πληθυσμούς (τουλάχιστον $\log n / I$ όπου I η πληροφορία Kullback-Leibler : $I(\theta_i, \theta^*) = E_{\theta_i} \ln \frac{f(x|\theta_i)}{f(x|\theta^*)}$)

Είδαμε ότι για μια ομοιόμορφα καλή πολιτική η συνάρτηση απώλειας δε μπορεί να είναι πολύ μικρή:

$$\liminf_{n \rightarrow \infty} \frac{R_n^{\pi}(\underline{\theta})}{\log n} \geq M(\underline{\theta})$$

Αν τώρα βρούμε μια πολιτική π^0 τέτοια ώστε

$$\limsup_{n \rightarrow \infty} \frac{R_n^{\pi}(\underline{\theta})}{\log n} \leq M(\underline{\theta})$$

αυτή θα είναι ασυμπτωτικά αποδοτική (βέλτιστη) διότι είναι ομοιόμορφα καλή και έχει τον χαμηλότερο δυνατό ρυθμό αύξησης της απώλειας ανάμεσα σε όλες τις ομοιόμορφα καλές πολιτικές.

Επομένως ο μεγαλύτερος δυνατός ρυθμός σύγκλισης του μέσου κέρδους στη βέλτιστη τιμή υπό πλήρη πληροφόρηση.

Παραλλαγές του Προβλήματος

Η συνηθισμένη μορφή του προβλήματος αφορά το δυαδικό MAB ή το Bernoulli MAB, στα οποία το κέρδος είναι μονάδα με πιθανότητα p διαφορετικά 0.

Μια διαφορετική μορφή είναι η κάθε επιλογή (arm) να αναπαρίσταται από μια ανεξάρτητη Μαρκοβιανή διαδικασία (Markov Machine). Κάθε φορά που γίνεται μια επιλογή πληθυσμού, η μηχανή μεταβαίνει σε μια νέα κατάσταση, καθορισμένη από τις πιθανότητες μετάβασης. Λαμβάνεται ένα κέρδος βάσει της κατάστασης που βρίσκεται η μηχανή. Έχουν προταθεί πάρα πολλές γενικεύσεις αυτού του προβλήματος. Δύο σημαντικές είναι τα «restless bandit problem» όπου οι καταστάσεις όλων των μηχανών εξελίσσονται στην διάρκεια του χρόνου ανεξάρτητα από το αν επιλέγονται ή όχι[129] και τα «arm-acquiring bandits», όπου ο αριθμός επιλογών αυξάνεται κατά την διάρκεια του χρόνου[128]

Μια άλλη παραλλαγή του προβλήματος είναι η ενσωμάτωση του κόστους c_i που λαμβάνεται κατά την παρατήρηση του πληθυσμού i και ενός ανώτατου επιτρεπόμενου κόστους C_0 ανά περίοδο παρατήρησης[17]. Έτσι ο στόχος είναι η εύρεση μιας πολιτικής τέτοιας ώστε το μέσο κόστος ανά περίοδο να μην ξεπερνά το C_0 και το μέσο κέρδος ανά περίοδο να είναι μέγιστο.

Στρατηγικές MAB και Αλγόριθμοι

Βέλτιστες Στρατηγικές

Στην δημοσίευση "*Asymptotically efficient adaptive allocation rules*" [66] κατασκευάζονται πολιτικές επιλογής πληθυσμού με τον μέγιστο δυνατό ρυθμό σύγκλισης (στην εύρεση του πληθυσμού με το μέγιστο κέρδος) για την περίπτωση όπου οι κατανομές ανήκουν σε εκθετική οικογένεια μιας παραμέτρου.

Στην συνέχεια στο "*Sequential choice from several populations*" [56] δίνονται απλοποιήσεις της παραπάνω πολιτικής καθώς και αποδείξεις για την περίπτωση κανονικών κατανομών πληθυσμού με γνωστές διασπορές.

Επόμενη προσθήκη στην μελέτη είναι το "*Optimal adaptive policies for sequential allocation problems*" [15] όπου κατασκευάστηκαν πολιτικές κάτω από γενικότερες συνθήκες οι οποίες περιλαμβάνουν και την περίπτωση όπου κάθε πληθυσμός εξαρτάται από ένα διάνυσμα άγνωστων παραμέτρων. Επίσης δόθηκε και λύση για την περίπτωση όπου οι κατανομές κέρδους είναι διακριτές, μονο-παραμετρικές κατανομές.

Αργότερα στο "*Optimal adaptive policies for Markov decision processes*" [16] μελετήθηκε το πρόβλημα υπό το γενικότερο πρίσμα των διαδικασιών αποφάσεων Markov υπό ελλιπή πληροφόρηση, όπου οι κανόνες μετάβασης και το αναμενόμενο κέρδος μιας περιόδου να εξαρτώνται από άγνωστες παραμέτρους.

Για Bernoulli MAB έχουν μελετηθεί υπολογιστικές μέθοδοι για τον υπολογισμό βέλτιστων πολιτικών (όχι ασυπτωτικών) χρησιμοποιώντας δυναμικό προγραμματισμό [89]. Μέσω αναζήτησης πινάκων και άλλων τεχνικών, η δημοσίευση παρέχει εφαρμόσιμες βέλτιστες λύσεις για Bernoulli Bandits δεδομένου ότι ο χρονικός ορίζοντας και ο αριθμός επιλογών (arms) δεν είναι ακραία μεγάλος. Επέκταση της συγκεκριμένης δουλειάς είναι το "*Delayed Reward Bernoulli Bandits: Optimal Policy and Predictive Meta-Algorithm PARDI*" [90] όπου παρουσιάζεται μια μέθοδος που καθορίζει την βέλτιστη πολιτική όταν το κέρδος δεν

αποκαλύπτεται άμεσα μετά από μια απόφαση αλλά σε επόμενο χρόνο. Αυτή η μέθοδος βασίζεται στον υπολογισμό του αναμενόμενου κέρδους των αποφάσεων που δεν έχουν αποκαλυφθεί ακόμα και της ενημέρωσης της posterior πιθανότητας αφού αποκαλυφθούν.

Ασυμπτωτικές Στρατηγικές

Οι στρατηγικές που έχουν επινοηθεί μέχρι τώρα και παρέχουν βέλτιστες ή ασυμπτωτικά προσεγγιστικές λύσεις για το πρόβλημα MAB μπορούν να διαχωριστούν σε τέσσερις γενικές κατηγορίες:

Ημί-κανονικές Στρατηγικές

Αυτές οι στρατηγικές ήταν οι πρώτες στρατηγικές και οι πιο απλές στην μορφή που ανακαλύφθηκαν για την επίλυση του προβλήματος MAB προσεγγιστικά. Όλες οι πολιτικές που ανήκουν σε αυτήν κατηγορία διαθέτουν μια κοινή «άπληστη» συμπεριφορά όπου η καλύτερη επιλογή (βάσει της μέχρι τώρα παρατήρησης) επιλέγεται πάντα εκτός από κάποιες φορές όπου επιλέγεται τυχαία μια διαφορετική επιλογή. Παρουσιάζονται κάποιες βασικές στρατηγικές αυτής της κατηγορίας:

- *Epsilon-greedy*: Ορίζεται μια σταθερά ϵ (τυπικά $\epsilon = 0.1$). Σε κάθε επανάληψη επιλέγεται η καλύτερη μέχρι τώρα επιλογή με πιθανότητα $1-\epsilon$ ενώ με πιθανότητα ϵ επιλέγεται τυχαία (ομοιόμορφα) μια άλλη επιλογή.
- *Epsilon-first*: Παρόμοια με την *Epsilon-greedy* με την διαφορά ότι για ένα ορισμένο σύνολο επαναλήψεων N έχουμε τις πρώτες ϵN επιλογές εντός της φάσης αναζήτησης (exploration) όπου επιλέγονται ομοιόμορφα τυχαία επιλογές και στην συνέχεια τις επόμενες $(1-\epsilon)N$ επιλογές εντός της φάσης αξιοποίησης (exploitation) όπου επιλέγεται πάντα η καλύτερη επιλογή μέχρι τώρα.
- *Epsilon-decreasing*: Όμοια με την *Epsilon-greedy* αλλά το ϵ μειώνεται καθώς ο αριθμός επαναλήψεων αυξάνει. Αυτό βοηθάει την μεγάλη πιθανότητα για

αναζήτηση στην αρχή όπου δεν υπάρχει πληροφορία και την μεγάλη πιθανότητα για αξιοποίηση στην συνέχεια[108].

- *Adaptive epsilon-greedy strategy based on value differences (VDBE)*: Παρόμοια με την Epsilon-decreasing αλλά η μείωση του ϵ βασίζεται στον ρυθμό εκμάθησης (εύρεσης της καλύτερη επιλογής). Εάν για παράδειγμα καθυστερεί να βρεθεί ξεκάθαρη καλύτερη επιλογή καθυστερεί και η μείωση του ϵ [115]
- *Adaptive epsilon-greedy strategy based on Bayesian ensembles (Epsilon-BMC)*: Παρόμοια με την VDBE αλλά με εγγύηση σύγκλισης στην καλύτερη επιλογή. Εδώ η παράμετρος ϵ καθορίζεται μέσω της εκτίμησης μιας posterior κατανομής όπου ζυγίζει έναν «άπληστο» αλγόριθμο εκμάθησης που εμπιστεύεται το κέρδος που λαμβάνει από κάθε επιλογή και από έναν ομοιόμορφο αλγόριθμο εκμάθησης που δεν το εμπιστεύεται. Αυτή η posterior εκτιμάται χρησιμοποιώντας μια Beta κατανομή κάτω από την υπόθεση της κανονικότητας για την κατανομή του κέρδους που παρατηρείται από τις επιλογές. Για να αποφευχθεί το ρίσκο της ταχείας μείωσης του ϵ , μοντελοποιείται και η αβεβαιότητα της διασποράς του κέρδους χρησιμοποιώντας ένα normal-gamma μοντέλο[40]
- *Contextual-Epsilon-greedy*: Παρόμοια με την Epsilon-greedy αλλά το ϵ υπολογίζεται βάσει μια πειραματικής διαδικασίας όπου ο αλγόριθμος υπολογισμού αντιλαμβάνεται κάποιο περιεχόμενο (Context-Aware). Βασίζεται στην δυναμική εναλλαγή αναζήτησης-αξιοποίησης και μπορεί να προσαρμόσει την ισορροπία ανάμεσα τους αποφασίζοντας ποια κατάσταση είναι η πιο κατάλληλη για την εκάστοτε κατάσταση. Αυτό σημαίνει ότι η πολιτική αυτή παρέχει μεγάλες περιόδους αναζήτησης όταν η κατάσταση δεν είναι επιτακτική ενώ όταν είναι παρέχει περιόδους αξιοποίησης[10]

Στρατηγικές Πιθανοτικής Συσχέτισης

Οι στρατηγικές αυτές αντιπροσωπεύουν την ιδέα ότι ο αριθμός των φορών που θα επιλεγεί ένας πληθυσμός θα πρέπει να ταιριάζει με την πιθανότητα του να είναι αυτός ο πληθυσμός βέλτιστος. Αυτές οι πολιτικές είναι επίσης γνωστές ως Δειγματοληψία (Sampling) Thompson ή Bayesian Bandits και είναι αρκετά εύκολες στην υλοποίηση τους εάν είναι εφικτή η δειγματοληψία από την posterior κατανομή του μέσου κέρδους για κάθε πληθυσμό.

Η Δειγματοληψία Thompson είναι μια από τις πιο παλιές ευριστικές (heuristics) μεθόδους ξεκινώντας από το 1933 και έχει λάβει μεγάλη προσοχή τελευταία σε εργασίες που αφορούν το MAB πρόβλημα. Η πολιτική αυτή όπως λέει και το όνομα της δειγματοληπεί τιμές από μια Beta κατανομή με διαφορετικούς παραμέτρους για κάθε πληθυσμό. Ο πληθυσμός με την μέγιστη τιμή θα είναι η επιλογή του αλγορίθμου για την εκάστοτε επανάληψη. Η Beta κατανομή διαθέτει δύο παραμέτρους a και b και λαμβάνουν ως τιμές τον αριθμό των επιτυχιών και των αποτυχιών για τον εκάστοτε πληθυσμό. Επιτυχία για έναν πληθυσμό είναι να επιφέρει κέρδος που είναι το μέγιστο σε σχέση με τα κέρδη που έχουν επιφέρει οι υπόλοιποι πληθυσμοί στο παρελθόν. Παρακάτω η τυπική σχέση για την επιλογή του αλγορίθμου:

$$arm = \operatorname{argmax}_i \beta(success(i), failure(i))$$

Όσο πιο μεγάλο είναι το a (επιτυχίες) σε σχέση με το b (αποτυχίες) τόσο πιο πολλές φορές θα επιλεγεί ο εκάστοτε πληθυσμός. Οι πληθυσμοί με μη βέλτιστο λόγο επιτυχίες/αποτυχίες θα έχουν πάντα πιθανότητα να επιλεγθούν και έτσι θα επιτυγχάνεται η επιθυμητή αναζήτηση.

Οι πολιτικές αυτές επίσης εφαρμόζονται και στα προβλήματα περιεχομένου (Contextual MAB)[102, 24].

Άλλες διάσημες πολιτικές με παρόμοια χαρακτηριστικά είναι η Upper Confidence Bounds (UCB) και η Softmax (Boltzmann Exploration) για τις οποίες λεπτομέρειες θα αναφερθούν στην υποενότητα "Μια ανάλυση και αξιολόγηση Αλγορίθμων MAB" που κάνει σύγκριση μεταξύ διάφορων MAB πολιτικών.

Στρατηγικές Κοστολόγησης

Οι πολιτικές αυτές εφαρμόζουν μια τιμή κόστους για κάθε επιλογή. Για παράδειγμα το κόστος μιας επιλογής μπορεί να είναι το άθροισμα του αναμενόμενου κέρδους από την επιλογή και του εκτιμώμενου μελλοντικού κέρδους βάσει και της επιπλέον πληροφορίας που θα αποκτηθεί από την επιλογή αυτή. Στην συνέχεια επιλέγεται πάντα η επιλογή με το μέγιστο κέρδος (ή το ελάχιστο κόστος)[120]

Στρατηγικές με Ηθικούς Περιορισμούς

Αυτές οι πολιτικές στοχεύουν να εφαρμόζονται σε περιστάσεις όπου το κόστος των επιλογών είναι πολύ μεγάλο. Ένα βασικό παράδειγμα είναι το πρόβλημα του εντοπισμού της κατάλληλης θεραπείας για κάποιον ασθενή. Βάσει των προηγούμενων αλγορίθμων η εκτίμηση του κέρδους από την κάθε επιλογή και κατ' επέκταση η μεγιστοποίηση του συνολικού κέρδους παίζει τον κυρίαρχο ρόλο. Αντιθέτως σε αυτήν την περίπτωση η ελαχιστοποίηση του κόστους παίζει τον κυρίαρχο ρόλο ακόμα και εάν το συνολικό κέρδος δεν καταλήξει να είναι το μέγιστο. Στο παράδειγμα της θεραπείας στόχος είναι η ελαχιστοποίηση της αναμενόμενης ζημιάς που μπορεί να υποστεί ο ασθενής. Η ζημιά αυτή μπορεί να προκληθεί για παράδειγμα από την αξιοποίηση μια θεραπείας (επιλογή) που θα αποδειχθεί μελλοντικά ότι δεν είναι η βέλτιστη[92].

Στρατηγικές MAB βάσει Περιεχομένου (Contextual bandit)

Η χρήσιμη γενίκευση του προβλήματος MAB είναι η εφαρμογή ενός πλαισίου περιεχομένου (Context) μέσα στο πρόβλημα. Σε κάθε επανάληψη όπου τίθεται το ζήτημα της επιλογής πληθυσμού ο αλγόριθμός λαμβάνει υπόψιν για την εκτίμηση του κέρδους και έναν πίνακα X όπου περιέχει χαρακτηριστικά (features) που σχετίζονται με τις διαθέσιμες επιλογές και ονομάζεται διάνυσμα περιεχομένου ή διάνυσμα χαρακτηριστικών ή κατάσταση περιβάλλοντος (state of environment). Ο πίνακας αυτός μπορεί να φέρει και την ιστορικότητα του κέρδους της κάθε επιλογής επομένως ενημερώνεται μετά από κάθε επανάληψη μέσω μιας διαδικασίας

εκμάθησης. Στόχος είναι μετά από ένα διάστημα ο πίνακας X να διαθέτει επαρκή πληροφορία για την συσχέτιση των χαρακτηριστικών με το κέρδος της κάθε επιλογής και έτσι να παίζει τον κυρίαρχο ρόλο για την επιλογή του εκάστοτε πληθυσμού σε κάθε επανάληψη. Η συσχέτιση περιεχομένου με την ιστορικότητα κέρδους θα μπορούσε να υπάρχει και σε ξεχωριστή δομή δεδομένων χωρίς να αλλάζει η λογική του αλγορίθμου[67].

Αλγόριθμος MAB Περιεχομένου Σε κάθε επανάληψη $t = 1, 2, \dots, T$ γίνονται τα εξής βήματα σε σειρά:

1. Το περιβάλλον αποκαλύπτει το περιεχόμενο $x_t \in X$
2. Λαμβάνεται μία απόφαση $a_{t,k} \in A$ βάσει του x_t και της μεθόδου MAB για αναζήτηση ή αξιοποίηση.
3. Λαμβάνεται το κέρδος $R_{t,k}$ της επιλογής k
4. Ενημερώνεται ο X βάσει του $R_{t,k}$

Οι αλγόριθμοι αυτοί αναφέρονται στην βιβλιογραφία και ως bandits with side observations, associative reinforcement learning, reinforcement learning with immediate reward, bandit problems with covariates αλλά ο όρος Contextual Bandit (CMAB) εισήχθη από τον Langford και Zhang[67] και έχει καθιερωθεί

Προσεγγιστικές Λύσεις για Contextual MAB

Online Linear Bandits

- *Upper Confidence Bound (LinUCB)*: Ο αλγόριθμος λειτουργεί υπό την υπόθεση της γραμμικής εξάρτησης του κέρδους μιας επιλογής α με το περιεχόμενο της x ($E[R_{t,\alpha}|x_{t,\alpha}] = x_{t,\alpha}^T \theta_\alpha^*$). Η μοντελοποίηση του περιεχομένου γίνεται μέσω ενός συνόλου από γραμμικούς εκτιμητές[68, 28].
- *Linear Associative Reinforcement Learning (LinRel)*: Παρόμοιο με το LinUCB αλλά χρησιμοποιεί το Singular-value decomposition (SVD) αντί του Ridge

Regression ώστε να εκτιμηθεί το περιθώριο λάθους για την συσχέτιση του περιεχομένου με το κέρδος[3, 50].

- *Historic LinUCB with Clusters (HLINUCBC)*: Επέκταση του LinUCB αξιοποιώντας ταυτόχρονα την ιστορικότητα και την ταξινόμηση της πληροφορίας[12, 54].

Online non-linear bandits

- *UCBogram*: Η μη γραμμική εκτίμηση του κέρδους γίνεται μέσω ενός εκτιμητή ονόματι *regressogram* εντός της μη παραμετρικής παλινδρόμησης. Τότε ο UCB αλγόριθμος εφαρμόζεται σε κάθε κομμάτι. Ενημέρωση του κομματιού του χώρου περιεχομένων προγραμματίζεται ή επιλέγεται προσαρμοστικά[98, 106, 88].
- *Generalized Linear Algorithms*: Η κατανομή κέρδους ακολουθεί ένα generalized linear model[34, 70, 54, 63]
- *KenrelUCB*: Μια kernel εκδοχή του LinUCB αλγορίθμου με αποδοτική εφαρμογή του προβλήματος δεδομένου του πεπερασμένου ορίζοντα[118].
- *Bandit Forest*: Αξιοποίηση του τυχαίου περιπάτου (random forest) στην εύρεση της από κοινού κατανομής του περιεχομένου και του κέρδους[33]
- *Oracle-based*: Ο αλγόριθμος αυτός αποδομεί το πρόβλημα του Contextual MAB σε μια σειρά από προβλήματα εκμάθησης υπό παρακολούθηση (supervised learning). Δεν βασίζεται στην τυπική υπόθεση της συνάρτησης κέρδους[1].

Bootstrapped bandits

Η μέθοδος Bootstrapping αναφέρεται ως ένας τρόπος να τυχαιοποιούνται τα ιστορικά δεδομένα[46]. Η λογική είναι να γίνεται μια επαναλαμβανόμενη δειγματοληψία ίδιου μεγέθους από τα δεδομένα. Ο Cortes έχει προτείνει δύο UCB αλγόριθμους περιεχομένου[29] (online & offline) όπου ως είσοδο λαμβάνουν έναν

αριθμό από δείγματα, εκτιμητές και ένα ποσοστό p με στόχο να κάνουν την επιλογή με το μέγιστο κέρδος. Η offline παραλλαγή είναι μια ανακατασκευή bootstrapping ενώ η online λαμβάνει τα δείγματα βάρη w και ενημερώνει τους εκτιμητές με τις νέες παρατηρήσεις του περιεχομένου και του κέρδους.

Νευρωνικές Προσεγγίσεις

- *NeuralUCB*: Χρησιμοποιεί ένα διάνυσμα χαρακτηριστικών βασισμένο σε νευρωνικό δίκτυο ώστε να κατασκευάσει το UCB. Η συνάρτηση κέρδους εκπαιδεύεται χρησιμοποιώντας ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο βάθους $L \geq 2$. Η βασική ιδέα του NeuralUCB είναι η αξιοποίηση ενός νευρωνικού δικτύου $f(x; \theta)$ ώστε να προβλέπει το κέρδος του περιεχομένου x και τα άνω όρια εμπιστοσύνης (UCB)[140].
- *NeuralBandit*: Σε αυτόν το αλγόριθμο πολλά νευρωνικά δίκτυα εκπαιδεύονται ώστε να μοντελοποιήσουν το κέρδος βάσει του περιεχομένου. Αλγόριθμος αξιοποιεί μία multi-experts προσέγγιση ώστε να επιλέξει online παραμέτρους για τα *perceptrons* πολλαπλών επιπέδων[100].

Constrained Contextual bandit

Στην πράξη υπάρχει ένα κόστος για την αξιοποίηση της πληροφορίας περιεχομένου το οποίο θα μπορούσε να μεταφραστεί στο ότι σε κάθε επανάληψη μόνο ένα υποσύνολο του πίνακα χαρακτηριστικών μπορεί να αξιοποιηθεί (περιορισμένων αριθμός χαρακτηριστικών). Οι αλγόριθμοι που αντιμετωπίζουν αυτό πρόβλημα υπό αυτήν την μορφή (CCB) θεωρούν υπαρκτούς τους περιορισμούς χρόνου και κόστους (budget) για το πρόβλημα MAB. Στο "*Resourceful contextual bandits*"[6] μελετήθηκε πρώτα το πρόβλημα bandit περιεχομένου με περιορισμούς όπου αναφέρθηκε επίσης και ως πρόβλημα Resourceful CB και έδειξε ότι η $O(\sqrt{T})$ απώλεια είναι εφικτή. Ένας απλός αλγόριθμος με λογαριθμική απώλεια προτάθηκε στο "*Algorithms with Logarithmic or Sublinear Regret for Constrained Contextual Bandits*"[131] (*UCB-ALP*) όπου συνδυάζεται η μέθοδος UCB με την *Adaptive*

Linear Programming (ALP) και μπορεί να εφαρμοστεί εύκολα σε συστήματα. Αν και εστιάζει σε μια ειδική περίπτωση, όπου υπάρχει ένας μοναδικός περιορισμός κόστους και το κόστος είναι σταθερό, τα αποτελέσματα της ανάλυσης ανοίγουν τον δρόμο για νέους αλγόριθμους που εφαρμόζονται σε πιο γενικά προβλήματα CCB.

Στρατηγικές MAB σε Ανταγωνιστικό Περιβάλλον (Adversarial Bandit)

Μια επέκταση του MAB προβλήματος ονομάζεται Adversarial και θεωρείται η ισχυρότερη γενίκευση του προβλήματος. Στο Adversarial πρόβλημα αφού γίνει μια επιλογή σε μια επανάληψη θεωρείται ότι το κέρδος που επιστρέφεται καθορίζεται από κάποιον εξωτερικό μηχανισμό με αποτέλεσμα η κατανομή κέρδους όχι μόνο να θεωρείται μη σταθερή αλλά και κατά κάποιον τρόπο να παίρνει μορφή που να εναντιώνεται στις προθέσεις του παίκτη-παρατηρητή.

Ένα γνωστό παράδειγμα τέτοιου προβλήματος είναι το Επαναλαμβανόμενο Δίλημμα του Κρατούμενου. Στο πρόβλημα αυτό δύο παίκτες έχουν στην διάθεση τους δύο επιλογές. Η μία επιλογή είναι η Άρνηση και η άλλη η Αποδοχή. Σε κάθε επανάληψη οι δύο παίκτες επιλέγουν ανεξάρτητα ο ένας από τον άλλον και ταυτόχρονα. Εάν επιλέξουν και οι δύο Αποδοχή θα έχουν ο καθένας κέρδος K_α ενώ εάν επιλέξουν και οι δύο Άρνηση κέρδος K_β . Στην περίπτωση που ο ένας επιλέξει Αποδοχή και ο άλλος Άρνηση τότε αυτός που επέλεξε Αποδοχή θα λάβει κέρδος K_γ ενώ ο άλλος K_δ . Το Δίλημμα του Κρατούμενου δίνει τιμές $K_\alpha, K_\beta, K_\gamma$ και K_δ τέτοιες ώστε εάν το παιχνίδι παιζόταν για μια και μόνο φορά κανένας παίκτης δεν θα είχε λογικό κίνητρο να επιλέξει Άρνηση με αποτέλεσμα το κέρδος να ήταν πάντα $K_\alpha \ll K_\beta$. Μέσω της επαναληπτικής διαδικασίας οι παίκτες έχουν την δυνατότητα να εξερευνήσουν τις δύο επιλογές τους με τέτοιο τρόπο έτσι ώστε τελικά να μεγιστοποιήσουν το κέρδος τους. Οι κλασικοί αλγόριθμοι MAB που έχουν αναφερθεί και πιο πάνω δεν λειτουργούν πολύ καλά σε αυτό το σενάριο. Για παράδειγμα εάν ο αντίπαλος παίκτης επιλέγει Αποδοχή στις 10 πρώτες επαναλήψεις, μετά επιλέγει Άρνηση στις επόμενες 20, μετά επιλέγει Αποδοχή στις

επόμενες 30 κ.ο.κ , αλγόριθμοι όπως ο UCB δεν θα μπορέσουν να ανταπεξέλθουν γρήγορα σε αυτές τις αλλαγές. Αυτό συμβαίνει διότι μετά από κάποιο σημείο οι μη-βέλτιστες επιλογές σπάνια επιλέγονται ώστε να περιορίσουν την αναζήτηση και έτσι γίνεται εστίαση μόνο στην αξιοποίηση. Όταν το περιβάλλον αλλάζει, ο αλγόριθμος είναι ανίκανος να προσαρμοστεί εγκαίρως ή ακόμα χειρότερα δεν μπορεί να εντοπίσει τις αλλαγές αυτές.

Προσεγγιστικές Λύσεις για *Adversarial Bandits*

- *Exp3*: Ο αλγόριθμος Exp3 διαλέγει έναν πληθυσμό i από K επιλογές βάσει της πιθανότητας $p_{i,t}$ προτιμώντας πληθυσμούς με υψηλό βάρος w_i (αξιοποίηση) ενώ υπάρχει πάντα και θετική πιθανότητα να επιλέξει κάποιον άλλον πληθυσμό (αναζήτηση).

$$p_{i,t} = (1 - \gamma) \frac{w_i}{\sum_{j=1}^K w_j} + \frac{\gamma}{K}$$

Το γ καθορίζει την τυχαιότητα της επιλογής παίρνοντας τιμές από 0 (επιλογή με προτεραιότητα το βάρος) έως 1 (ομοιόμορφη επιλογή). Το εκτιμώμενο κέρδος υπολογίζεται ως $\widehat{R}_{i,t} = \frac{R_{i,t}}{p_{i,t}}$ όπου $R_{i,t}$ το πραγματικό κέρδος της επιλογής i την επανάληψη t . Μετά την δειγματοληψία το κέρδος που λαμβάνεται ενημερώνονται τα βάρη των πληθυσμών ως $w_i(t+1) = w_i(t) \exp \frac{\gamma \widehat{R}_{i,t}}{K}$. Η εκθετική αύξηση της διαδικασίας βοηθάει στην άμεση αύξηση των «καλών» πληθυσμών. Η απώλεια του αλγορίθμου είναι $o(\sqrt{KT \log(K)})$ [103]

- *Follow the Perturbed Leader (FPL)*: Ο αλγόριθμος επιλέγει τον πληθυσμό με το μεγαλύτερο μέσο κέρδος μέχρι τώρα προσθέτοντας εκθετικό θόρυβο σε αυτόν έτσι ώστε να προκύπτει και αναζήτηση[52].

Άλλες επεκτάσεις του προβλήματος MAB

Infinite-armed Bandit

Στα παραπάνω προβλήματα MAB και τις παραλλαγές τους είναι σαφές ότι οι επιλογές (πληθυσμοί) είναι διακριτές και πεπερασμένες, καθορισμένες από την παράμετρο K . Μια παραλλαγή που πρωτοαναφέρθηκε στο "*The Continuum-Armed Bandit Problem*" [2] είναι οι επιλογές να είναι μια συνεχής μεταβλητή στον χώρο K -διαστάσεων.

Dependent Arm Bandit

Το πρόβλημα MAB μπορεί να επεκταθεί στο επίπεδο που οι πληθυσμοί δεν είναι ανεξάρτητοι μεταξύ τους. Αυτό σημαίνει ότι η πληροφορία που εξάγεται από μια επιλογή μπορεί να υποδηλώνει πληροφορία και για άλλους πληθυσμούς που συσχετίζονται μεταξύ τους. Το πρόβλημα μπορεί να μοντελοποιηθεί μέσω της ταξινόμησης (clustering) των πληθυσμών έτσι ώστε πληθυσμοί που εξαρτώνται στενά μεταξύ τους να ανήκουν στο ίδιο σύνολο [87].

Non-stationary Bandit

Σε αυτή την περίπτωση το πρόβλημα MAB είναι μη στατικό, δηλαδή η παράμετρος η οποία προσπαθεί να εκτιμηθεί (πχ μέσο κέρδος) αλλάζει μέσα στον χρόνο. Σε αντίθεση με την Adversarial εκδοχή εδώ η μεταβλητότητα των μεγεθών δεν υποτίθεται ότι εξελίσσεται ενάντια στις προθέσεις βελτιστοποίησης. Υποθέτουμε ότι το μέσο εκτιμώμενο κέρδος του πληθυσμού i μπορεί να αλλάζει σε κάθε επανάληψη $t \in T : R_{i,t-1} \neq R_{i,t}$. Επομένως το $R_{i,t}$ δεν μπορεί να αντιπροσωπεύσει πια όλη την ακολουθία του εκτιμώμενου κέρδους για τον πληθυσμό i . Αντ' αυτού το $R_k = \{R_{i,t}\}_{t=1}^T$ [42].

Διάφοροι αλγόριθμοι έχουν προταθεί που αντιμετωπίζουν το πρόβλημα της αλλαγής των κατανομών στο χρόνο όπως το DiscountedUCB [58] και το Sliding-Window UCB [37].

Παρόμοια προσέγγιση υπάρχει και για τους αλγόριθμους Thompson Sampling όπως ο f-Discounted-Sliding Window TS (f-dsw TS)[20]. Ο f-dsw TS αλγόριθμος αξιοποιεί έναν συντελεστή έκπτωσης για το μέσο κέρδος βάσει της ιστορικότητας του για έναν πληθυσμό και επίσης ένα χρονικό παράθυρο συσχετισμένο με τον κάθε πληθυσμό ώστε να εξισορροπήσει το concept drift στο μη-στατικό περιβάλλον.

Μια άλλη δημοσίευση εισάγει το weighted least squares TS (WLS-TS) το οποίο παρέχει αξία σε στατικά και μη στατικά περιβάλλοντα [18].

Το Adjusted Upper Confidence Bound (A-UCB), υπό την υπόθεση ύπαρξης στοχαστικού μοντέλου, παρέχει άνω όρια για την συνάρτηση απώλειας.

Dueling Bandit

Το πρόβλημα Dueling Bandit στοχεύει να μοντελοποιήσει την ισορροπία αξιοποίησης-αναζήτησης βάσει της βαθμολόγησης (feedback). Σε αυτήν την παραλλαγή ο παίκτης έχει την δυνατότητα να κάνει δύο επιλογές ταυτόχρονα αλλά θα λάβει ως feedback μια δυαδική τιμή η οποία θα δείξει ποια από τις δύο επιλογές έχει μεγαλύτερο κέρδος. Η δυσκολία του προβλήματος αφορά την αδυναμία του παίκτη να παρατηρήσει άμεσα το κέρδος της κάθε επιλογής. Ο πρώτοι αλγόριθμοι που κλήθηκαν να αντιμετωπίσουν το συγκεκριμένο σενάριο είναι ο InterleavingFiltering[135] και ο Beat-The-Mean[136]. Το σχετικό feedback μπορεί επίσης να οδηγήσει στο παράδοξο της ψηφοφορίας

$$(A > B \rightarrow B > \Gamma \rightarrow \Gamma > A)$$

. Άλλοι πιο πρόσφατοι αλγόριθμοι που αντιμετωπίζουν το συγκεκριμένο σενάριο είναι:

- Relative Upper Confidence Bounds (RUCB)[146]
- Relative EXponential weighing (REX3)[35]
- Copeland Confidence Bounds (CCB)[145]
- Relative Minimum Empirical Divergence (RMED)[61]
- Double Thompson Sampling (DTS)[130]

Collaborative Bandit

Το πρόβλημα του MAB Συνεργατικό Φιλτράρισμα (Collaboration Filtering Bandit) (πχ COFIBA) έχει πρωτοαναφερθεί στο "*Collaborative Filtering Bandits*" [71] όπου οι μέθοδοι Συνεργατικού Φιλτραρίσματος και Φιλτραρίσματος βάσει Περιεχομένου εφαρμόζονται πάνω στα συστήματα συστάσεων. Αυτές οι προσεγγίσεις δεν είναι ιδανικές για τα πολύ δυναμικά πεδία του προβλήματος σύστασης (πχ άρθρα, διαφήμιση) όπου το σύνολο των αντικειμένων και των χρηστών αλλάζει ταχέως. Μελετώνται τεχνικές προσαρμοστικής ταξινόμησης για την σύσταση βάσει πολιτικών αξιοποίησης-αναζήτησης σε MAB περιεχομένου (Contextual MAB) [39]. Για παράδειγμα ο αλγόριθμος COFIBA λαμβάνει υπόψιν το φαινόμενο συνεργατικού φιλταρίσματος που προκύπτει από την αλληλεπίδραση των χρηστών με τα αντικείμενα, ταξινομώντας δυναμικά τους χρήστες και τα αντικείμενα. Ο αλγόριθμος επομένως αξιοποιεί τις ακολουθίες προτιμήσεων που προκύπτουν μέσα από τα δεδομένα στα οποία εφαρμόζονται μέθοδοι συνεργατικού φιλτραρίσματος.

Combinatorial Bandit

Στο πρόβλημα Συνδυαστικού MAB (Combinatorial MAB) ο παίκτης δεν καλείται να κάνει μια επιλογή από ένα σύνολο επιλογών αλλά να συνδυάσει επιλογές από πολλά σύνολα. Σχηματικά θα μπορούσαμε να πούμε ότι στην πρώτη περίπτωση ο παίκτης επιλέγει μια τιμή μιας διακριτής μεταβλητής ενώ στην δεύτερη πολλές τιμές όπου η κάθε μια είναι τιμή μιας διαφορετικής διακριτής μεταβλητής. Επομένως ο αριθμός των πιθανών επιλογών σε κάθε επανάληψη είναι εκθετικός βάσει του συνόλου των μεταβλητών. Έχουν μελετηθεί σενάρια όπου οι μεταβλητές λαμβάνουν δυαδικές τιμές [26] αλλά και τιμές εντός ενός αυθαίρετου συνόλου [86].

Μια ανάλυση και αξιολόγηση Αλγορίθμων MAB

Στην υπο-ενότητα αυτή παρουσιάζεται η ανάλυση και τα αποτελέσματα δοκιμών της εργασίας "Algorithms for the multi-armed bandit problem" [62] πάνω σε έξι γνωστούς αλγορίθμους.

ϵ -greedy

Σε κάθε επανάληψη $t = 1, 2, \dots$ ο αλγόριθμος επιλέγει τον πληθυσμό με το μεγαλύτερο μέσο κέρδος μέχρι τώρα με πιθανότητα $1 - \epsilon$ ή έναν τυχαίο πληθυσμό (ομοιόμορφη $1/K$) με πιθανότητα ϵ . Έστω $\hat{\mu}_1(0), \dots, \hat{\mu}_K(0)$ τα αρχικά μέση κέρδη για τον κάθε πληθυσμό $1, \dots, K$ έχουμε:

$$p_i(t+1) = \begin{cases} 1 - \epsilon + \epsilon/K & , \text{αν } i = \arg \max_{j=1 \dots K} \hat{\mu}_j(t) \\ \epsilon/K & , \text{διαφορετικά} \end{cases}$$

Αναζήτηση Boltzmann (Softmax)

Οι Softmax μέθοδοι βασίζονται στο *Αξίωμα της Επιλογής του Luce (1959)* και επιλέγουν τον πληθυσμό με πιθανότητα ανάλογη του μέσου κέρδους τους. Αυτό σημαίνει ότι οι πληθυσμοί με μεγαλύτερο μέσο κέρδος έχουν μεγαλύτερη πιθανότητα να επιλεγθούν. Η εργασία χρησιμοποιεί την αναζήτηση Boltzmann όπου είναι μια Softmax μέθοδος που επιλέγει έναν πληθυσμό χρησιμοποιώντας μια κατανομή Boltzmann:

$$p_i(t+1) = \frac{e^{\frac{\hat{\mu}_i(t)}{\tau}}}{\sum_{j=1}^K e^{\frac{\hat{\mu}_j(t)}{\tau}}}, i = 1, \dots, K$$

όπου τ είναι μια παράμετρος που καθορίζει την τυχαιότητα της επιλογής. Για $\tau = 0$ η μέθοδος συμπεριφέρεται όπως μια ϵ -greedy μέθοδος ενώ για $\tau \rightarrow \infty$ ο αλγόριθμός επιλέγει εντελώς τυχαία και ομοιόμορφα.

Αλγόριθμοι Καταδίωξης (*Pursuit*)

Οι αλγόριθμοι καταδίωξης σε αντίθεση με τους προηγούμενους διατηρούν μια καθορισμένη πολιτική για την επιλογή πληθυσμών η οποία ενημερώνεται σε κάθε επανάληψη από τις πληροφορίες που εξάγονται. Στην εργασία χρησιμοποιείται μια απλή μορφή αλγορίθμου καταδίωξης[108] όπου η επιλογή ξεκινάει μια ομοιόμορφη κατανομή για την επιλογή πληθυσμού $p_i(0) = 1/K$ και σε κάθε επανάληψη γίνεται η εξής ενημέρωση:

$$p_i(t+1) = \begin{cases} p_i(t) + \beta(1 - p_i(t)) , & \text{αν } i \text{ βέλτιστο} \\ p_i(t) + \beta(0 - p_i(t)) , & \text{διαφορετικά.} \end{cases}$$

$\beta \in (0, 1)$ και είναι ο ρυθμός εκμάθησης.

Αλγόριθμοι Ενισχυτικής Σύγκρισης (*Reinforcement Comparison*)

Οι μέθοδοι Ενισχυτικής Σύγκρισης μοιάζουν με τις μεθόδους Καταδίωξης καθώς διατηρούν μια κατανομή πιθανότητας για τον τρόπο επιλογής των πληθυσμών που δεν υπολογίζεται άμεσα από τις εμπειρικές πληροφορίες. Οι μέθοδοι αυτές επίσης διατηρούν ένα μέσο εκτιμώμενο κέρδος $\hat{r}(t)$. Η πιθανότητα να επιλεγεί ένας πληθυσμός υπολογίζεται από την σύγκριση των εμπειρικών μέσων κερδών με το $\hat{r}(t)$. Η πιθανότητα αυξάνεται εάν το μέσο κέρδος είναι πάνω από το $r(t)$ ή μειώνεται ανάλογα.

Τυπικά ο αλγόριθμος διατηρεί ένα σύνολο από «προτιμήσεις» $\pi_i(t)$ για κάθε πληθυσμό i . Σε κάθε επανάληψη t η πιθανότητα $p_i(t)$ υπολογίζεται βάσει της κατανομής Boltzmann και αυτών των «προτιμήσεων»:

$$p_i(t) = \frac{e^{\pi_i(t)}}{\sum_{j=1}^K e^{\pi_j(t)}}$$

Εάν ο πληθυσμός $j(t)$ επιλεγεί στην επανάληψη t και ληφθεί κέρδος $r(t)$, τότε η «προτίμηση» $\pi_{j(t)}$ και το μέσο κέρδος $\hat{r}(t+1)$ ενημερώνονται ως:

$$\pi_{j(t)}(t+1) = \pi_{j(t)}(t) + \beta(r(t) - \hat{r}(t))$$

$$\hat{r}(t+1) = (1 - \alpha)\hat{r}(t) + \alpha r(t)$$

Τα α, β είναι οι ρυθμοί εκμάθησης εντός του $(0,1)$

Upper Confidence Bounds (UCB)

Η οικογένεια των αλγορίθμων UCB έχει προταθεί από τους *Lai & Robbins (1985)* και έχει εφαρμοστεί σε εργασίες[4, 15, 16] ως μια πιο απλή εφαρμογή της ιδέα της *Αισιόδοξης εκτίμησης της Αβεβαιότητας*[66]. Επεκτάσεις αυτών των μεθόδων έχουν αποδειχθεί πολύ αποτελεσματικές απέναντι σε συγκεκριμένα προβλήματα.

Ο απλούστερος αλγόριθμος, ο UCB1, διατηρεί τον αριθμό που έχει επιλεγεί ο κάθε πληθυσμός, έστω $n_i(t)$, επιπλέον του μέσου κέρδους. Αρχικά όλοι οι πληθυσμοί επιλέγονται από μία φορά. Στην συνέχεια σε κάθε επανάληψη t ο αλγόριθμος «άπληστα» επιλέγει πάντα τον πληθυσμό $j(t)$ ως:

$$j(t) = \arg \max_{i=1 \dots K} \left(\hat{\mu}_i + \sqrt{\frac{2 \ln t}{n_i}} \right)$$

Η φόρμουλα υποδεικνύει την εξής λογική: Κάθε πληθυσμός επιλέγεται βάσει του μέσου κέρδους του, του αριθμού των φορών που έχει επιλεγεί σε σχέση με το χρονικό διάστημα που έχει παρέλθει από την αρχή του αλγορίθμου. Αυτό σημαίνει ότι η αναζήτηση προκύπτει ως φυσικό επόμενο όταν έχει παρέλθει μεγάλο διάστημα και ένας πληθυσμός έχει επιλεγεί πολύ λίγες φορές.

Λόγω της κλειστής μορφής της παραπάνω φόρμουλας αποδεικνύεται το αναμενόμενο ανώτατο όριο της απώλειας

$$8 \sum_{i: \mu_i < \mu^*} \frac{\ln t}{\Delta_i} + \left(1 + \frac{\pi^2}{3} \right) \sum_{i=1}^K \Delta_i$$

όπου $\Delta_i = \mu^* - \mu_i$ που δίνει $o(\log n)$ και λέγεται ότι επιλύει το πρόβλημα MAB. Στην εργασία μελετάται επίσης και μια προέκταση του UCB1 αλγορίθμου, ο UCB-Tuned, όπου αναφέρεται ότι παρουσιάζει καλύτερα αποτελέσματα καθώς λαμβάνει υπόψιν και την διασπορά του κάθε πληθυσμού

$$j(t) = \arg \max_{i=1 \dots K} \left(\hat{\mu}_i + \sqrt{\frac{\ln t}{n_i} \min\left(\frac{1}{4}, V_i(n_i)\right)} \right)$$

ΟΠΟΥ

$$V_i(t) = \hat{\sigma}_i^2(t) + \sqrt{\frac{2 \ln t}{n_i(t)}}$$

Περιγραφή Πειράματος και Αποτελέσματα

Στην εργασία στήθηκε ένα πείραμα έτσι ώστε να μελετηθεί η απόδοση των παραπάνω αλγορίθμων. Το πείραμα στήθηκε ως εξής:

Κάθε πείραμα εκτελεί τον εκάστοτε αλγόριθμο για 1000 επαναλήψεις όπου έχει φανεί ότι υπάρχει μια σύγκλιση για όλους τους αλγορίθμους. Στην αναφορά αξιολογούνται τρία κριτήρια:

1. Η συνολική απώλεια για όλο το πείραμα
2. Η απώλεια σε συνάρτηση με τον χρόνο
3. Το ποσοστό των επιλογών όπου επιλέχθηκε ο βέλτιστος πληθυσμός

Διεξάχθηκαν πειράματα με διαφορετικές παραμέτρους έτσι ώστε η σύγκριση μεταξύ των αλγορίθμων να γίνει σε διαφορετικά σενάρια. Η μια παράμετρος ήταν ο αριθμός των διαφορετικών πληθυσμών K και η άλλη η διασπορά της κατανομής των πληθυσμών σ^2 . Δεδομένου ότι τα αποτελέσματα ήταν πολύ παρόμοια για όλες τις κατανομές επιλέχθηκε η κατανομή για όλους τους πληθυσμούς να είναι η κανονική. Για αλγορίθμους που χρησιμοποιούν παραμέτρους (πχ learning rates α, β, τ κτλ) έχουν επιλεγεί οι κατάλληλες παράμετροι που βελτιστοποιούν τον αλγόριθμο για την εκάστοτε συνθήκη.

Τα αποτελέσματα του πειράματος δείχνουν τα εξής: Οι απλούστεροι αλγόριθμοι (*ϵ -greedy*, *Softmax*) υπερσχύουν απέναντι στους υπόλοιπους αλγορίθμους σχεδόν σε όλες τις συνθήκες εκτός από την περίπτωση που **η διασπορά είναι πολύ υψηλή και το πλήθος των επιλογών είναι μικρό**. Στις τελευταίες περιπτώσεις ο αλγόριθμος *UCB-tuned* φαίνεται να λειτουργεί πιο αποδοτικά, το οποίο ήταν αναμενόμενο καθώς ο συγκεκριμένος είναι σχεδιασμένος για να ανταποκρίνεται καλά σε υψηλές διασπορές.

Οι αλγόριθμοι φαίνεται να επηρεάζονται σημαντικά από τις παραμέτρους των συνθηκών. Συγκεκριμένα οι αλγόριθμοι UCB φάνηκε να αντιμετωπίζουν καλά τις

περιπτώσεις με μικρό K και μεγάλο σ , αλλά η απόδοση τους πέφτει κατακόρυφα όταν το K αυξάνεται.

Ο αλγόριθμος καταδίωξης φάνηκε να έχει την μικρότερη απόδοση σε όλες τις περιπτώσεις.

Τέλος το πείραμα έδειξε ότι οι μοναδικές παράμετροι που επηρεάζουν σημαντικά την απόδοση όλων των αλγορίθμων είναι το K και το σ . Οι παράμετροι όπως ο τύπος της κατανομής των πληθυσμών και οι υπερπαράμετροι δεν έπαιξαν καθόλου σημαντικό ρόλο στην διάταξη της απόδοσης μεταξύ των αλγορίθμων.

Εφαρμογές Αλγορίθμων MAB

Το πρόβλημα της ακολουθιακής λήψης αποφάσεων κάτω από συνθήκες αβεβαιότητας είναι ένα πρόβλημα που εμφανίζεται σε πάρα πολλές πτυχές της πραγματικότητας. Οι παρακάτω τομείς που θα παρουσιαστούν περιληπτικά είναι οι βασικότεροι τομείς στους οποίους το δίλημμα αναζήτησης-αξιοποίησης παρουσιάζεται ως σημαντική πρόκληση απέναντι στην λήψη αποφάσεων.[13]

Σύστημα Υγείας

Κλινικές Δοκιμές: Το πρόβλημα εμφανίζεται από την ανάγκη εύρεσης μιας θεραπείας ή αποτελεσματικότητας ενός φαρμάκου. Για παράδειγμα η δοκιμαστική περίοδος για ένα φάρμακο πάνω σε ασθενείς που έχουν συμφωνήσει να συμμετέχουν στην διαδικασία, αποτελείται από μια σειρά αποφάσεων που αφορά τον ρυθμό λήψης και την δόση του φαρμάκου συναρτήσει των συμπτωμάτων και της ανταπόκρισης του οργανισμού.

Μοντελοποίηση εγκεφαλικής δραστηριότητας και συμπεριφοράς: Με στόχο την μελέτη εγκεφαλικών διαταραχών και συμπεριφοράς μελετώνται αλγόριθμοι MAB περιεχομένου που ενσωματώνουν την κατά προτίμηση επεξεργασία κέρδους (bias) η οποία σχετίζεται με νευρολογικές και ψυχιατρικές διαταραχές όπως το Parkinson και το Alzheimer.

Οικονομία

Το πρόβλημα ορίζεται στο πλαίσιο της επιλογής χαρτοφυλακίου ενσωματώνοντας μεθόδους μηχανικής μάθησης. Οι μέθοδοι MAB που μπορούν να μελετηθούν και να εφαρμοστούν σε αυτό το πεδίο στοχεύουν στην επιλογή του βέλτιστου χαρτοφυλακίου το οποίο διαθέτει έναν συνδυασμό από παθητικές και ενεργητικές επενδύσεις βάσει συναρτήσεων διαχείρισης ρίσκου.

Δυναμική Τιμολόγηση

Η ανάγκη προκύπτει από τις εταιρίες λιανικής μέσω διαδικτύου όπου χρειάζονται έναν τρόπο δυναμικής και άμεσης τιμολόγησης που θα προσαρμόζεται στις εκάστοτε απαιτήσεις της αγοράς και με στόχευση την μεγιστοποίηση του μακροπρόθεσμου κέρδους. Στην κατηγορία αυτή εφαρμόζονται μη-στατικοί αλγόριθμοι MAB.

Συστήματα Συστάσεων

Το πεδίο αυτό, που θα αναλυθεί περισσότερο και στο επόμενο κεφάλαιο της παρούσας διπλωματικής εργασίας, αντιμετωπίζει το δίλλημα αναζήτησης-αξιοποίησης στα πλαίσια της αλληλεπίδρασης των συστημάτων σύστασης με τους χρήστες τους. Από την μία πλευρά η αναζήτηση προσδίδει στο σύστημα τις πληροφορίες που χρειάζεται και από την άλλη η αξιοποίηση ικανοποιεί τους χρήστες του μέσω των κατάλληλων συστάσεων τους. Οι ανάγκες των συστημάτων συστάσεις είναι σε πολλά επίπεδα και οι αλγόριθμοι που μπορούν να αξιοποιηθούν είναι όλων των κατηγοριών στατικοί, μη-στατικοί, περιεχομένου κτλ.

Βελτιστοποίηση επιρροής

Με εφαρμογή στα κοινωνικά δίκτυα και στόχο την επέκταση της δημοτικότητας ενός προϊόντος, το πρόβλημα αφορά την βελτιστοποίηση του τρόπου επιλογής χρηστών που διαθέτουν επιρροή εντός του δικτύου (influencers) έτσι ώστε να γίνεται διαφήμιση.

Εξαγωγή Πληροφοριών

Αξιοποιώντας αλγορίθμους MAB περιεχομένου το πρόβλημα ανάκτησης πληροφοριών μοντελοποιείται ως πρόβλημα της ισορροπίας αναζήτησης-αξιοποίησης των μετα-πληροφοριών.

Συστήματα Διαλόγου

Οι αλγόριθμοι MAB περιεχομένου μπορούν να χρησιμοποιηθούν στην κατασκευή συστημάτων αυτόματου διαλόγου για την βελτιστοποίηση του τρόπου επιλογής απαντήσεων καθώς επίσης και στην κατάλληλη έναρξη τέτοιων διαλόγων από αυτά τα συστήματα.

Εντοπισμός Σφαλμάτων

Ο εντοπισμός σφαλμάτων αφορά συστήματα που στόχο έχουν να εντοπίζουν στοιχεία δικτύων και συστημάτων όπου υπάρχει απόκλιση από την συνήθη συμπεριφορά τους. Η ενσωμάτωση αλγορίθμων MAB έχει στόχο την διαχείριση των περιπτώσεων όπου η αναζήτηση στοιχείων προς ανάλυση βοηθάει στην βελτίωση του μοντέλου πρόβλεψης σφαλμάτων και η αξιοποίηση στη ανάλυση στοιχείων που είναι σαφώς ύποπτα για ύπαρξη σφαλμάτων.

Τηλεπικοινωνίες

Εδώ η μοντελοποίηση MAB αφορά την βέλτιστη επιλογή ασύρματων δικτύων ή κυψελών επικοινωνίας από μία κινούμενη συσκευή με στόχο την μεγιστοποίηση της ποιότητας του σήματος.

Επισημαίνεται πως δεν έχει ουσία η εφαρμογή της κάθε κατηγορίας αλγορίθμων MAB σε όλα τα παραπάνω πεδία. Για παράδειγμα οι μη-στατικοί αλγόριθμοι δεν αξιοποιούνται σε εφαρμογές υγείας δεδομένου ότι δεν παρουσιάζονται σημαντικές αλλαγές στο περιβάλλον που γίνεται η διαδικασία λήψης αποφάσεων όπως η κατάσταση ενός ασθενή όπου είναι σαφές, σταθερό και μετρήσιμο εάν ανταποκρίνεται θετικά ή όχι σε μια θεραπεία. Από την άλλη στο πεδίο του εντοπισμού σφαλμάτων έχει μεγαλύτερη ουσία να αξιοποιηθούν μη στατικοί αλγόριθμοι MAB περιεχομένου καθώς μια αποκλίνουσα συμπεριφορά μπορεί να συμπεριφέρεται «ανταγωνιστικά» απέναντι στην εκπαίδευση του συστήματος εντοπισμού (προσαρμογή).

Μια επιπλέον επισημάνση είναι ότι φαίνεται πως είναι προς το παρόν αδύνατη η κατασκευή ενός καθολικού αλγορίθμου που μπορεί να αντιμετωπίζει όλο το σύνολο των προβλημάτων σε κάθε πεδίο ή να μαθαίνει από το ένα πεδίο και να μπορεί να

εφαρμόζεται σε ένα άλλο. Το γεγονός αυτό δίνει μια κατεύθυνση για την αξιοποίηση και μελέτη της μετάδοσης γνώσης (transfer-learning) και του multi-tasking στο πλαίσιο των MAB προβλημάτων.

Τέλος δεδομένης της online φύσης των προβλημάτων MAB στον πραγματικό κόσμο, το φυσικό επόμενο βήμα είναι η συνεχής και δια-βίου εκπαίδευσης των συστημάτων αποφάσεων έτσι ώστε μοντέλα που κατασκευάζονται σε κάποιο προηγούμενο βήμα να αξιοποιούνται σε κάποιο νέο χωρίς να υπάρχει το φαινόμενο της καταστροφικής αμνησίας (catastrophic forgetting).

Στην συνέχεια παρουσιάζονται οι βασικοί τομείς στους οποίους οι αλγόριθμοι MAB μπορούν να αξιοποιηθούν τεχνικά στα πλαίσια της Μηχανικής Μάθησης:

Επιλογή Αλγορίθμων

Η επιλογή αλγορίθμων τυπικά βασίζεται σε μοντέλα απόδοσης των αλγορίθμων τα οποία κατασκευάζονται μέσω offline εκπαίδευσης. Δεδομένου ότι κάτι τέτοιο έχει κόστος σε χρόνο και πόρους, η ενσωμάτωση αλγορίθμων MAB βοηθάνε σε ένα online πλαίσιο εκπαίδευσης των μοντέλων αυτών όπου γίνεται η επαναληπτική ενημέρωσή τους. Έχοντας δεδομένη μια επίλυση ενός προβλήματος, το δίλημμα αναζήτησης-αξιοποίησης αναπαριστάται μέσω της επιλογής αλγορίθμων υπό δοκιμή (δηλαδή αξιοποιώντας τους για να λύνουν το πρόβλημα) συγκρίνοντας τις απόδοσεις τους.

Βελτιστοποίηση Υπερ-παραμέτρων

Η απόδοση αλγορίθμων μηχανικής μάθησης βασίζεται σημαντικά στην εύρεση ενός συνόλου βέλτιστων υπερπαραμέτρων. Ενώ σύγχρονες προσεγγίσεις χρησιμοποιούν Bayesian βελτιστοποίηση ώστε να επιλέγονται οι βέλτιστες υπερπαραμέτροι, η εφαρμογή MAB αλγορίθμων επιταχύνει την τυχαία αναζήτηση μέσω της προσαρμοστικής κατανομής πόρων και της πρώιμης παύσης (*early-stopping*).

Επιλογή Χαρακτηριστικών

Στην κλασική επιβλεπόμενη μάθηση (supervised learning) η πραγματική τιμή ενός δείγματος αποκαλύπτεται πάντα στον ταξινομητή, σε αντίθεση με το πλαίσιο MAB όπου οποιαδήποτε λάθος ταξινόμηση δίνει μηδενικό κέρδος και μόνο η πραγματικά σωστή ταξινόμηση δίνει κέρδος μονάδα. Το πρόβλημα της Online επιλογής

χαρακτηριστικών (features) αφορά τις εύστοχες προβλέψεις μέσω της κατάλληλης επιλογής ενός μικρού συνόλου χαρακτηριστικών και επομένως η επιλογή αυτή μπορεί γίνει μέσω MAB αλγορίθμων.

Active Learning

Στην επιβλεπόμενη ταξινόμηση το να μαρκάρονται (labeled) όλα τα παραδείγματα εκπαίδευσης για γίνει η εκπαίδευση του μοντέλου έχει τεράστιο κόστος. Οι στρατηγικές active learning αντιμετωπίζουν το πρόβλημα αυτό μέσω της επιλογής των πιο χρήσιμων μη μαρκαρισμένων παραδειγμάτων έτσι ώστε να εκπαιδεύσουν το μοντέλο εκτίμησης. Η επιλογή των παραδειγμάτων αυτών μπορεί να μοντελοποιηθεί μέσω του διλήμματος αναζήτησης-αξιοποίησης και επομένως να αξιοποιηθούν αλγόριθμοι MAB για να το αντιμετωπίσουν.

Ταξινόμηση

Η συνεργατική ταξινόμηση (collaborative clustering) είναι μια μέθοδος μηχανικής μάθησης όπου αντιμετωπίζει την μη επιβλεπόμενη ανάλυση πολύπλοκων δεδομένων χρησιμοποιώντας πολλούς αλγορίθμους που συνεργάζονται μεταξύ τους. Σε γνωστές εφαρμογές συνεργατικής ταξινόμησης οι αλγόριθμοι που συνεργάζονται ανταλλάσσουν πληροφορίες με στόχο να βελτιώνονται αμφότεροι. Ένα σημαντικό σημείο κλειδί σε αυτήν την μέθοδο είναι η αξιολόγηση του ποια συνεργασία μεταξύ αλγορίθμων είναι ευεργετική ή επιβλαβής. Το τρόπος με τον οποίον μπορεί να γίνει αυτή η αξιολόγηση αφορά το πρόβλημα αναζήτησης-αξιοποίησης και μπορεί να αντιμετωπιστεί μέσω MAB αλγορίθμων.

Ενισχυτική Μάθηση

Αυτόνομα συστήματα που αλληλεπιδρούν με το φυσικό περιβάλλον παίζουν σημαντικό ρόλο στην καθημερινότητα (συστήματα ασφαλείας, ρομποτικά μηχανήματα κτλ). Για να εξασφαλιστεί ότι η συμπεριφορά τέτοιων συστημάτων είναι ευθυγραμμισμένη με τις αξίες της κοινωνίας εντός της οποίας λειτουργούν, είναι σημαντικό να αναπτυχθούν τεχνικές όπου επιτρέπουν σε αυτούς του αυτοματισμούς όχι μόνο να μεγιστοποιούν το κέρδος στο τεχνικό πλαίσιο στο οποίο λειτουργούν αλλά και να μαθαίνουν να λαμβάνουν υπόψιν τους έμμεσους περιορισμούς της κοινωνίας. Ένας τρόπος να γίνει αυτό είναι να επιτραπεί η

παρατήρηση της συμπεριφοράς μελών της κοινωνίας από το σύστημα χωρίς να επηρεάζεται από αυστηρά ορισμένους περιορισμούς. Αξιοποιώντας έτσι την αντίστροφη ενισχυτική μάθηση πάνω στα δεδομένα αυτής της συμπεριφοράς, είναι εφικτό το σύστημα να ενσωματώσει τέτοιου είδους περιορισμούς μέσω ενός πλαισίου περιεχομένου (contextual-constraints). Με αυτόν τον τρόπο το σύστημα αποφάσεων θα μπορέσει να διαλέγει ανάμεσα σε δύο πολιτικές, την πολιτική βάσει περιορισμών και την πολιτική βάσει μεγιστοποίησης κέρδους αξιοποιώντας αλγορίθμους MAB περιεχομένου.

Συνοψίζοντας τα παραπάνω περί συνεργασίας τεχνικών μηχανικής μάθησης και αλγορίθμων MAB έχουμε και εδώ μια κατηγοριοποίηση βάσει των τύπων των αλγορίθμων που εφαρμόζονται καταλλήλως στην κάθε τεχνική. Σαφώς οι απλοί αλγόριθμοι MAB μπορούν να εφαρμοστούν σε όλες τις τεχνικές με εξαίρεση την Επιλογή Αλγορίθμων όπου λόγω του δυναμικού περιβάλλοντος αξία έχουν μη-στατικοί αλγόριθμοι MAB. Η εφαρμογή των μη-στατικών MAB έχει αξία επίσης και στην Επιλογή Χαρακτηριστικών αλλά και στην Ενισχυτική Μάθηση. Τέλος η εφαρμογή αλγορίθμων MAB περιεχομένου φαίνεται χρήσιμη για τις περιπτώσεις του Active Learning και της Ενισχυτικής Μάθησης.

Κατακλείδα

Στην ενότητα αυτή περιγράφηκε το πρόβλημα του διλλήματος αναζήτησης-αξιοποίησης ή αλλιώς Multi-Armed-Bandit Problem. Δόθηκαν οι τυπικοί ορισμοί του προβλήματος και περιγράφηκαν οι προσεγγίσεις που έχουν κατασκευαστεί μέχρι τώρα για να αντιμετωπίσουν την κάθε εκδοχή του. Έγινε μια ανάλυση μεθοδολογίας σε κάποιους δημοφιλείς αλγορίθμους και παρουσιάστηκαν περιληπτικά τα αποτελέσματα ενός δημοσιευμένου πειράματος που έκανε μια σύγκριση μεταξύ τους σε ένα γενικευμένο πλαίσιο αξιολόγησης της απόδοσης τους. Τέλος αναφέρθηκαν οι τομείς στους οποίους το πρόβλημα MAB παρουσιάζεται και καλείται να επιλυθεί καθώς και οι εφαρμογές μηχανικής μάθησης που αξιοποιούν αντίστοιχους αλγορίθμους. Στην επόμενη ενότητα η παρούσα εργασία θα στοχεύσει στην συσχέτιση των διάφορων κατηγοριών των αλγορίθμων MAB με το πεδίο των Συστημάτων Σύστασης και τις διαφορές υποπεριοχές του.

ΜΑΒ σε Συστήματα Συστάσεων

Αλγόριθμοι ΜΑΒ στα Συστήματα Συστάσεων

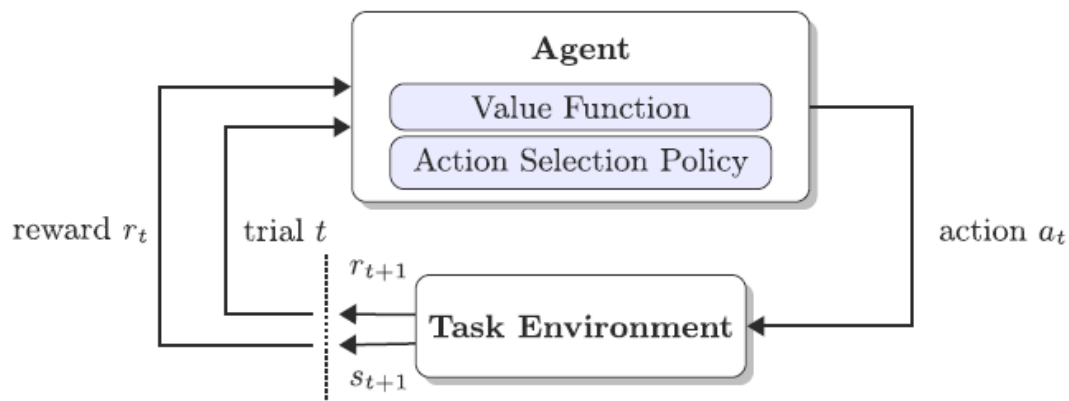
Υπάρχουν διάφορες εργασίες οι οποίες μοντελοποιούν την Online σύσταση αντικειμένων ως ένα ΜΑΒ πρόβλημα[32, 123, 124]. Στις περισσότερες αναπαραστάσεις ΜΑΒ τα αντικείμενα που συστήνονται παίζουν το ρόλο των πληθυσμών προς δειγματοληψία, δηλαδή τις διαθέσιμες επιλογές που έχει ο ΜΑΒ αλγόριθμος. Η επιλογή ενός πληθυσμού a είναι ισοδύναμο με την σύσταση ενός αντικειμένου i και το κέρδος είναι η βαθμολόγηση του χρήστη πάνω στην σύσταση αυτή (κλικ, βαθμός κτλ). Για τον λόγο αυτό ο κύριος στόχος είναι η μεγιστοποίηση του αναμενόμενου κέρδους μετά από T επαναλήψεις.

$$i^*(t) = \arg \max_{i=1 \dots K} \sum_{t=1}^T E[R_{u,i_t} | t]$$

Σε αυτό το πλαίσιο λοιπόν η μεγιστοποίηση του κέρδους συσχετίζεται με την μεγιστοποίηση της ικανοποίησης των χρηστών εντός του συστήματος. Αυτό απαιτεί πολιτικές εξατομικευμένων επιλογών βάσει των προτιμήσεων των χρηστών και των παρελθοντικών ενεργειών τους (ιστορία h). Το αντικείμενο $i^*(t)$ επομένως θα επιλεγεί βάσει μιας πολιτικής π η οποία ορίζεται ως μια συνάρτηση αναζήτησης – αξιοποίησης βάσει την υπάρχουσας γνώσης για τον χρήστη μέχρι τώρα. Στην βιβλιογραφία υπάρχουν πολλές τέτοιες πολιτικές οι οποίες βελτιστοποιούν την συνάρτηση αυτή κάτω από συγκεκριμένες συνθήκες.

Οι κατηγορίες MAB αλγορίθμων στα Συστήματα Συστάσεων

Με στόχο την προσαρμογή του MAB πλαισίου στα συστήματα συστάσεων, πολλοί ερευνητές κατανοούν πως το πλαίσιο αυτό καλείται να εφαρμόσει διάφορες στρατηγικές που ανταποκρίνονται σε πολλά διαφορετικά σενάρια του πραγματικού κόσμου. Για τον λόγο αυτό έχει αναπτυχθεί μια πληθώρα τεχνικών που προσαρμόζουν MAB αλγορίθμους καταλλήλως στις διάφορες κατηγορίες συστημάτων σύστασης. Στην συνέχεια θα περιγραφούν αυτές τις διάφορες κατηγορίες τεχνικών που επεκτείνουν το MAB πλαίσιο, και θα παρουσιαστούν τρόποι με τους οποίους διάφοροι γνωστοί αλγόριθμοι MAB μπορούν να βελτιωθούν μέσα από το πλαίσιο των συστημάτων σύστασης.[36]



Εικόνα 5: Η διαδικασία MAB σχηματικά

Αλγόριθμοι Πολλαπλών Επιλογών

Συνήθως οι αλγόριθμοι MAB εστιάζουν στην επιλογή ενός πληθυσμού μέσα από το σύνολο των διαθέσιμων αλλά σε πολλές εφαρμογές αυτό δεν είναι αρκετό. Είναι λοιπόν σημαντικό το σύστημα να έχει την δυνατότητα να μπορεί να κάνει πολλαπλές επιλογές ως τις βέλτιστες μέσα από ένα σύνολο. Υπάρχουν κάποιοι αλγόριθμοι

όπου επεκτείνουν το παραδοσιακό πρόβλημα με τον παραπάνω τρόπο επιτρέποντας την επιλογή περισσότερων πληθυσμών:

- **EXP3.M**: Αυτός ο ανταγωνιστικός (adversarial) αλγόριθμος μελετά το πρόβλημα των πολλαπλών επιλογών[117]. Αντίστοιχα με την μονή επιλογή, ο αλγόριθμος επιλέγει μια ενέργεια i με πιθανότητα $p_i(t)$ και υπολογίζει μια εκτίμηση του κέρδους $r_i(t)$ ως $\hat{r}_i(t) = \frac{r_i(t)}{p_i(t)}$ εάν η ενέργεια i είναι η όντως επιλεγμένη και $\hat{r}_i(t) = 0$ διαφορετικά. Επομένως το πρόβλημα εξομοιώνεται στην επιλογή των καλύτερων m πληθυσμών από τους K . Ο αλγόριθμος βασίζεται στο *online mirror descent principle* για την ελαχιστοποίηση της απώλειας στην online κυρτή βελτιστοποίηση. Στο “*A Multipleplay Bandit Algorithm Applied to Recommender Systems*”[75] προτείνεται μια υπολογιστικά αποδοτική υλοποίηση του αλγορίθμου η οποία έχει δοκιμαστεί σε μεγάλα σύνολα δεδομένων ώστε να αποδειχθεί εμπειρικά η απόδοση του.
- **CBMA**: Ο αλγόριθμος πολλαπλών επιλογών με περιεχόμενο έχει προταθεί στο “*Pairwise Regression with Upper Confidence Bound for Contextual Bandit with Multiple Actions*”[23] και χρησιμοποιεί ridge παλινδρόμηση ανά ζεύγη με ανώτατο όριο εμπιστοσύνης για πολλαπλή επιλογή. Η βασική ιδέα είναι η ενέργεια που επιλέγει ο αλγόριθμος να αφορά το πρόβλημα ιεράρχησης ανά ζεύγη το οποίο όμως λύνεται μέσω της μετατροπής του σε πρόβλημα ridge παλινδρόμησης ανά ζεύγη. Η προσέγγιση αυτή υποθέτει T επαναλήψεις όπου σε κάθε μία παρέχεται ένας πίνακας περιεχομένου X_t και επιλέγονται m ενέργειες από τις διαθέσιμες K αποκαλύπτοντας ένα διάνυσμα κέρδους r_t . Ο αλγόριθμος ενημερώνεται βάσει του διανύσματος r_t σε κάθε επανάληψη. Οι ενέργειες βαθμολογούνται και επιλέγεται η ενέργεια με την υψηλότερη βαθμολογία. Ο CBMA χρησιμοποιεί τον αλγόριθμο PairUCB ώστε να αντιμετωπίσει τον υπολογισμό της άγνωστης παραμέτρου θ (εκτιμητής κέρδους ανά ζεύγη) μέσω της ridge παλινδρόμησης ανάμεσα στον πίνακα περιεχομένου και τον διάνυσμα κέρδους. Ο αλγόριθμος έχει δοκιμαστεί σε πραγματικά και συνθετικά δεδομένα και έχει παρουσιάσει καλύτερα αποτελέσματα σε σχέση με τον LinUCB.
- **MP-TS**: Ο αλγόριθμος Thompson Sampling πολλαπλών επιλογών

παρουσιάζεται στο “*Optimal regret analysis of thompson sampling in stochastic multiarmed bandit problem with multiple plays*,” [60]. Ανακατασκευάζει τον κλασικό αλγόριθμο TS με την διαφορά ότι αντί να επιλέγει τον πληθυσμό με καλύτερη posterior $\theta_i(t)$, επιλέγει τους πρώτους L πληθυσμούς βάσει της posterior. Με στόχο ένα χαμηλό κάτω όριο απώλειας, ο αλγόριθμος καταπιέζει την ταυτόχρονη επιλογή 2 μη-βέλτιστων επιλογών. Ο MPTS αποδίδει καλύτερα σε σχέση με τους προηγούμενους αλγορίθμους EXP3.M και CBMA.

Αλγόριθμοι Βελτιστοποίησης Πολλαπλών Στόχων

Πολλές φορές υπάρχει η ανάγκη της βελτιστοποίησης ενός συστήματος προς παραπάνω από μια κατεύθυνση. Ο πιο διαδεδομένος στόχος βελτιστοποίησης για τα συστήματα συστάσεων είναι ο ρυθμός κλικ (*CTR – Click-Through-Rate*) αλλά δεν είναι πάντα και ο μοναδικός. Εάν για παράδειγμα έχουμε στόχο την βελτιστοποίηση της ικανοποίησης ενός αναγνώστη και ταυτόχρονα την βελτιστοποίηση της διαφήμισης του εκδότη τότε είναι αναγκαία η εφαρμογή ενός αλγορίθμου πολλαπλών στόχων. Στην συνέχεια παρουσιάζονται διάφορες προσεγγίσεις στο πλαίσιο αυτό:

- **MAB πολλαπλών στόχων:** Βελτιστοποίηση του Γενικευμένου *Gini Index*: Η τυποποίηση του προβλήματος της βελτιστοποίησης πολλαπλών στόχων γίνεται μέσω ενός διανύσματος D διαστάσεων το οποίο κωδικοποιεί τους πολλαπλούς στόχους που πρόκειται να βελτιστοποιηθούν [19]. Στην εργασία ο κυρίως στόχος είναι η εύρεση της τομής ανάμεσα στην βελτιστοποίηση όλων των στόχων του διανύσματος και της ισορροπίας μεταξύ αυτών. Αντί να χρησιμοποιηθεί *Pareto Front*, οι συγγραφείς αξιοποιούν τον Γενικευμένο Δείκτη *Gini* [127]. Σε αυτό το πλαίσιο, το κόστος κάθε επιλογής δεν έχει κλιμακωτή τιμή αλλά είναι διάνυσμα. Επομένως με στόχο την εύρεση των καλύτερων επιλογών, γίνεται σύγκριση των γενικευμένων δεικτών *Gini* όπου η βέλτιστη επιλογή ελαχιστοποιεί τον δείκτη αυτόν. Ο αλγόριθμος βελτιστοποίησης βασίζεται στο *Online Gradient Descent*.

- **Αλγόριθμος Βελτιστοποίησης Πολλαπλών Στόχων σε πλατφόρμα μουσικής ροής:** Στο παράδειγμα αυτό παρουσιάζεται πως το Spotify προσπαθεί να βελτιστοποιήσει τις συστάσεις του ώστε να ικανοποιήσει τον χρήστη, τον καλλιτέχνη και την ίδια την πλατφόρμα, κάτι το ποίο μεταφράζεται ως βελτιστοποίηση πολλαπλών στόχων [R. Mehrotra, N. Xue, and M. Lalmas, “Bandit based Optimization of Multiple Objectives on a Music Streaming Platform,”]. Για να το πετύχει αυτό οι συγγραφείς της εργασίας τυποποιούν το πρόβλημα ως MAB περιεχομένου πολλαπλών στόχων, όπου σε κάθε επανάληψη t ένα σύνολο από χαρακτηριστικά F_t μήκους M συσχετίζεται με κάθε επιλογή. Υποθέτοντας γραμμική σχέση, το κέρδος r_t από την επιλογή k είναι $r_t = F_{t,k}^T \theta^* + \zeta_t$ όπου $\theta^* \in R^{(D \times M)}$ είναι μια άγνωστη σταθερά παράμετρος και $\zeta_t \in R^D$ είναι ο ανεξάρτητος τυχαίος θόρυβος για κάθε στόχο. Για την επιλογή του βέλτιστου πληθυσμού χρησιμοποιείται και εδώ ο *Γενικευμένος Δείκτης Gini*. Ο προτεινόμενος αλγόριθμος ονομάζεται *MOLinCB* και βελτιστοποιεί την συνάρτηση απώλειας. Έχει δοκιμαστεί σε συνθετικά δεδομένα και σε πραγματικά από το Spotify, συγκρινόμενος με προσεγγίσεις άπληστων αλγορίθμων, MO και στρατηγικές απλής επιλογής. Ο αλγόριθμος αποδίδει καλύτερα στο πλαίσιο της βελτιστοποίησης όλων των στόχων και επίσης εντός χρονικών περιορισμών. Μια σημαντική ανακάλυψη στην παραπάνω εργασία είναι πως τα μοντέλα πολλαπλών στόχων αποδίδουν καλύτερα σε σχέση με τα μοντέλα ενός στόχου για όλες τις μετρικές.
- Άλλες εργασίες[134] προτείνουν MAB πολλαπλών στόχων που χρησιμοποιούν TS για την επιλογή συντελεστών βαρύτητας που φανερώνουν νέες επιλογές μέσω του Pareto Front με υψηλότερη συχνότητα. Μελετώνται προβλήματα δύο στόχων με κάποιον ως κυρίαρχο[112] ή περιγράφουν ένα πρόβλημα όπου χρησιμοποιείται μια γραμμική κλιμακωτή συνάρτηση για τον μετασχηματισμό του προβλήματος πολλαπλών στόχων σε ένα κλασικό πρόβλημα μέσω του αθροίσματος των συντελεστών βαρύτητας πάνω στους στόχους[116].

Ομαδοποίηση μέσω MAB Αλγορίθμων

Οι αλγόριθμοι MAB μελετώνται και στα πλαίσια της ομαδοποίησης (clustering). Οι ερευνητές έχουν ανακαλύψει ότι είναι χρήσιμη η ομαδοποίηση των αλγορίθμων, των πληθυσμών ή των χαρακτηριστικών έτσι ώστε να μειώνεται η διάσταση του προβλήματος ενώ ταυτόχρονα να διατηρείται ένα καλό όριο απώλειας. Στην συνέχεια παρουσιάζονται κάποιες προσεγγίσεις στα πλαίσια της ομαδοποίησης:

- **MAB Αλγόριθμοι με αλληλοεξαρτώμενες επιλογές:** Η προσέγγιση αυτή εστιάζει στην ομαδοποίηση των επιλογών (πληθυσμών) βάσει της εξάρτησης μεταξύ τους. Η υπόθεση είναι ότι οι επιλογές είναι εξαρτημένες[87]. Το πρόβλημα τυποποιείται ως ομαδοποίηση N πληθυσμών σε K γνωστές ομάδες με στόχο την μεγιστοποίηση του κέρδους σε πεπερασμένο χρόνο ή την ελαχιστοποίηση της απώλειας αντίστοιχα. Θεωρώντας π_i την παράμετρο της ομάδας C_i κάθε φορά που επιλέγεται ο πληθυσμός i ενημερώνει την $\pi_i(t)$ ταυτόχρονα για όλους τους πληθυσμούς εντός της ίδιας ομάδας. Σε κάθε βήμα η πολιτική αυτή υπολογίζει ένα ζεύγος (δείκτης, πληθυσμός) για κάθε ομάδα. Υπολογίζει το κέρδος και την διασπορά για κάθε ομάδα και στην συνέχεια χρησιμοποιεί μια δεύτερη πολιτική που επιλέγει μια υποομάδα εντός της ομάδας και χρησιμοποιεί την ίδια πολιτική για να επιλέξει τον πληθυσμό εντός της υποομάδας αυτής. Η προσέγγιση αυτή ονομάζεται Πολιτική Δύο Επιπέδων (Two Level Policy – TLP). Τα πειράματα που έχουν γίνει αφορούν συνθετικά δεδομένα και δεδομένα μιας διαφημιστικής καμπάνιας χρησιμοποιώντας TLP και UCB αλγορίθμους ξεχωριστά. Ο TLP έδωσε καλύτερα εμπειρικά αποτελέσματα και οι συγγραφείς συμπεραίνουν πως η απώλεια εξαρτάται από τα χαρακτηριστικά των ομάδων.
- **CLUB – Online Ομαδοποίηση:** Η ιδέα της Online Ομαδοποίησης είναι να παρέχονται συστάσεις σε N χρήστες οι οποίοι ομαδοποιούνται σε C σύνολα, βάσει των αξιολογήσεων τους¹. Σχετίζεται με την σύσταση σε ομάδες και κάθε ομάδα C_i παραμετροποιείται μέσω κάποιες άγνωστου διανύσματος $c_j \in R^D$,

¹C. Gentile, S. Li, and G. Zappella, “Online Clustering of Bandits,”

κοινή για κάθε χρήστη που ανήκει στο C_j . Οι συγγραφείς υποθέτουν έναν γραμμικό αλγόριθμο για κάθε ομάδα και έναν γραμμικό αλγόριθμο για κάθε χρήστη εντός μιας ομάδας. Κάθε χρήστης i διαθέτει ένα διάνυσμα w_i , ενώ μια ομάδα j διαθέτει ένα διάνυσμα z_j όπου z_j είναι το άθροισμα των w_i . Οι χρήστες λαμβάνουν συστάσεις σε ακολουθία. Ξεκινούν με μια μεγάλη ομάδα με n στοιχεία και διαγράφουν πλευρές εάν τα διανύσματα είναι πολύ διαφορετικά. Αφού ο χρήστης i λάβει μια σύσταση εντός της ομάδας j , τα διανύσματα του χρήστη και της ομάδας αυτής ενημερώνονται. Ο αλγόριθμος έχει δοκιμαστεί σε συνθετικά και πραγματικά δεδομένα, ανταγωνιζόμενος δύο εκδοχές *LinUCB* και απέδειξε πως δίνει καλύτερα αποτελέσματα από τους ανταγωνιστές του.

- ***DynUCB – Δυναμική Ομαδοποίηση MAB Αλγορίθμων***

Περιεχόμενου: Η βασική ιδέα είναι ο διαχωρισμός των χρηστών σε ομάδες και η κατάλληλη παραμετροποίηση του αλγορίθμου που θα εκτελείται για κάθε ομάδα. Επίσης η ομαδοποίηση γίνεται δυναμικά που σημαίνει ότι οι χρήστες μεταφέρονται ανάμεσα στις ομάδες από βήμα σε βήμα, κάτι το οποίο προσομοιώνει και την πραγματική συμπεριφορά χρηστών σε ένα σύστημα. Ο *DynUCB* βασίζεται στον *LinUCB* δεχόμενος ως είσοδο τον επιθυμητό αριθμό ομάδων K . Για κάθε χρήστη u εντός μια ομάδας C_k την χρονική στιγμή t , ο αλγόριθμος μαθαίνει τον συντελεστή θ_k από τις παραμέτρους στο επίπεδο της ομάδας του αλγορίθμου b_k και του πίνακα περιεχομένου M_k . Για την ενέργεια a το κέρδος θα είναι $r_{\alpha,t} = \hat{\theta}_k^T \chi_{t,\alpha} + \alpha \sqrt{\chi_{t,\alpha}^T \hat{M}_\alpha^{-1} \chi_{t,\alpha} \log(t+1)}$ όπου το θ_k είναι μια παράμετρος στο επίπεδο της ομάδας για έναν χρήστη u της ομάδας C_k και x_t είναι το διάνυσμα περιεχομένου για την ενέργεια a . Το κέρδος χρησιμοποιείται ώστε να ενημερώσει τον συντελεστή του χρήστη $\theta_{u,t}$ και στην συνέχεια ο χρήστης επανατοποθετείται σε κάποια ομάδα C_k με τον συντελεστή $\hat{\theta}_{k'}$ που είναι πιο κοντά στον δικό του συντελεστή θ_k . Τα πειράματα έχουν διεξαχθεί πάνω σε πραγματικά δεδομένα με τον *DynUCB* να ανταγωνίζεται τον *CLUB* που περιγράφηκε προηγουμένως και δύο εκδοχές *LinUCB*, με τα αποτελέσματα να δικαιώνουν σημαντικά τον *DynUCB*[84].

Προσεγγίσεις Συνεργατικού Φιλτραρίσματος

Το Συνεργατικό Φιλτράρισμα είναι μια από τις πιο γνωστές μεθόδους εξατομικευμένων συστάσεων, η οποία προσπαθεί να συσχετίσει τον ενεργό χρήστη με άλλους χρήστες έτσι ώστε να γίνονται επιτυχημένες συστάσεις. Η μέθοδος αυτή μπορεί να συνδυαστεί και με το πλαίσιο των MAB αλγορίθμων:

- **COFIBA – MAB Συνεργατικού Φιλτραρίσματος:** Ο αλγόριθμος αυτός είναι η τομή μεταξύ της μεθόδου ομαδοποίησης και του συνεργατικού φιλτραρίσματος, και προσπαθεί να εκμεταλλευτεί ακολουθίες δεδομένων έτσι ώστε να εκτιμήσει προτιμήσεις χρηστών. Ο COFIBA είναι μια ευρετική (heuristic) μέθοδος με άνω όριο συνδυασμένη με διαδικασίες προσαρμοστικής ομαδοποίησης μεταξύ χρηστών και αντικειμένων[71]. Αξιοποιεί ένα συνεργατικό φαινόμενο που προκαλείται από την μεταλλαγή μεταξύ των αλληλεπιδράσεων χρηστών και αντικειμένων. Το συνεργατικό φαινόμενο αυτό σημαίνει ότι οι χρήστες ομαδοποιούνται δυναμικά βάσει των αντικειμένων που τους προτείνονται ενώ ταυτόχρονα τα αντικείμενα ομαδοποιούνται βάσει της συσχέτισης τους με τους χρήστες. Η ομαδοποίηση αυτή είναι το πρώτο βήμα του αλγορίθμου. Στην συνέχεια τα αντικείμενα (που διακρίνονται από το διάνυσμα χαρακτηριστικών τους $I = \{x_1, x_2, \dots, x_N\}$ υφίστανται πολλαπλές ομαδοποιήσεις πάνω στο σύνολο U των χρηστών και μια ομαδοποίηση πάνω στο σύνολο I των αντικειμένων. Ο αλγόριθμος μετά υπολογίζει τα γειτονικά σύνολα $N_{i,t}(x_{i,t})$ σε αντιστοιχία με τα αντικείμενα στο $C_{i,t}$ και τα αποθηκεύει σε ομάδες από την πλευρά του χρήστη. Οι ομάδες αυτές ενημερώνονται από την πλευρά του χρήστη και του αντικειμένου.
- **FactorUCB:** Η προσέγγιση αυτή υποθέτει την ομοιότητα των χρηστών χρησιμοποιώντας έναν σταθμισμένο γράφο μέσω την συμπλήρωσης ενός πίνακα μικρής διάστασης, που διαθέτει τις αλληλεπιδράσεις των χρηστών-αντικειμένων, και εφαρμόζοντας μια UCB στρατηγική επιλογής[122]. Από την πλευρά του κάθε χρήστη χρησιμοποιείται ένας MAB αλγόριθμος που λειτουργεί με την μέθοδο Matrix Factorization. Ένα από τα βασικά

ερωτήματα της Matrix Factorization μεθόδου είναι πως επιλέγεται το επόμενο ζεύγος χρήστη-αντικειμένου για την ενημέρωση του μοντέλου έτσι ώστε να είναι όσο το δυνατόν λιγότερο επιρρεπές στο συστηματικό σφάλμα (bias). Η αναζήτηση των βέλτιστων αντικειμένων γίνεται επομένως μια «καλή επιλογή» ώστε να εξασφαλίσει βελτιστότητα. Για να γίνει αυτό αξιοποιείται μια UCB πολιτική και υποθέτει ότι το παρατηρούμενο κέρδος από κάθε χρήστη καθορίζεται από τους γειτονικούς του χρήστες. Το καλύτερο αντικείμενο επιλέγεται βάσει των χαρακτηριστικών περιεχομένου (γνωστών και παρατηρούμενων) και των κρυμμένων συντελεστών του (latent factors) καθώς επίσης και από έναν πίνακα Θ που αποθηκεύει τους κρυμμένων συντελεστές των χρηστών. Με άλλα λόγια το καλύτερο αντικείμενο επιλέγεται σύμφωνα με το εκτιμώμενο αναμενόμενο κέρδος του, που βασίζεται στην τωρινή εκτίμηση των κρυμμένων συντελεστών του χρήστη και του αντικειμένου. Ο αλγόριθμος αυτός ανταγωνίστηκε σε συνθετικά και πραγματικά δεδομένα διάφορες άλλες μεθόδους συνεργατικού φιλτραρίσματος και απέδειξε πως η αξιοποίηση των χαρακτηριστικών περιεχομένου και των αλληλεπιδράσεων των χρηστών βοηθά σημαντικά στην αντιμετώπιση της Αργής Εκκίνησης.

- ***PTS – Particle Thompson Sampling για Matrix Factorization:***

Στην εργασία αυτή το σύστημα σύστασης προτείνει αντικείμενα στον χρήστη χρησιμοποιώντας Δειγματοληψία Thompson. Ο χρήστης στην συνέχεια βαθμολογεί το αντικείμενο και το σύστημα ενημερώνει τις στατιστικές πληροφορίες του. Η πρωτοτυπία σε αυτόν τον αλγόριθμο είναι ότι τα συστήματα ενημερώνουν όχι μόνο τον latent πίνακα του χρήστη αλλά ενημερώνουν και την latent αναπαράσταση του αντικειμένου εντός του matrix factorization μοντέλου βάσει της βαθμολόγησης[57]. Για να γίνει αυτή η ενημέρωση σε πραγματικό χρόνο, το σύστημα χρησιμοποιεί το *RaoBlackwellized particle φίλτρο* για online matrix factorization. Η posterior πυκνότητα εκτιμάται από *particle filtering* (με ένα σύνολο από σταθμισμένα δείγματα). Το σύστημα αξιολογήθηκε σε συνθετικά και πραγματικά δεδομένα με πολύ καλά αποτελέσματα και έχει πατενταριστεί από την Adobe.

ΜΑΒ Πεπερασμένου Χρόνου

Κάποιες από τις προσεγγίσεις των συστημάτων σύστασης λαμβάνουν σημαντικά υπόψη το χρονικό πλαίσιο του περιεχομένου, των αντικειμένων και των αλληλεπιδράσεων τους με τους χρήστες (πχ διαφημίσεις με ημερομηνία λήξης). Θεωρώντας ότι το χρονικό πλαίσιο αυτό αλλάζει, όσον αφορά το λαμβανόμενο κέρδος (αλλαγές στις καταναλωτικές προτιμήσεις), κάποιες μέθοδοι προσπαθούν να αντιμετωπίσουν αυτό το δυναμικό χρονικό περιθώριο μέσω κατάλληλων ΜΑΒ αλγορίθμων.

- **ΜΑΒ με Θνητότητα:** Η ιδέα πίσω από αυτήν την προσέγγιση είναι ότι οι επιλογές έχουν διάρκεια ζωής όπου μετά το πέρας τους ο αλγόριθμος πρέπει να αναζητήσει ξανά[22]. Υπάρχουν δύο σενάρια που έχουν προταθεί. Το πρώτο και λιγότερο ρεαλιστικό ονομάζεται *stateaware*, όπου το κέρδος είναι επίσης και το μέγιστο δυνατό κέρδος. Το δεύτερο σενάριο ονομάζεται *stateoblivious* και διαπραγματεύεται με πιθανότητες κέρδους το οποίο είναι και το πιο ρεαλιστικό. Οι συγγραφείς μοντελοποιούν την θνητότητα των αλγορίθμων με δύο τρόπους. Ο ένας είναι η επιλογή να παραμένει «ζωντανή» αφού επιλεγεί και ο άλλος να μην παραμένει. Εξαιτίας της υπόθεσης σχετικά με το κέρδος, η απώλεια ενός ΜΑΒ με θνητότητα δεν μπορεί ποτέ να μηδενιστεί, κάτι που στους τυπικούς ΜΑΒ είναι εφικτό. Εντός του παραπάνω πλαισίου υπάρχουν διαφορετικοί αλγόριθμοι που αναπτύχθηκαν μεταξύ των *stateaware* και *stateoblivious* σεναρίων βάσει των ίδιων υποθέσεων για το κέρδος. Ο αλγόριθμός αποδεικνύεται βέλτιστος στην υπόθεση του ντετερμινιστικού κέρδους, Δεδομένης μιας κατανομής και ενός αναμενόμενου χρόνου ζωής, υπολογίζεται το αναμενόμενο κέρδος για την επιλογή r_i . Στην συνέχεια γίνεται μια τυχαία επιλογή και εάν το κέρδος της είναι μεγαλύτερο από το υπολογισμένο τότε ο αλγόριθμος κάνει την καινούρια επιλογή σε κάθε επανάληψη έως ότου λήξει. Η μόνη διαφορά σε σχέση με την υπόθεση στοχαστικού κέρδους είναι ότι στην δεύτερη περίπτωση αντί να γίνει μια επιλογή μια φορά για να καθοριστεί το κέρδος της, ο αλγόριθμος κάνει

την επιλογή n φορές και την εγκαταλείπει εάν δεν φαίνεται υποσχόμενη. Η εργασία χρησιμοποιεί εποχικό *UCB* και *e-greedy* (με ένα υποσύνολο των k/c επιλογών ανά εποχή). Είναι σημαντικό να σημειωθεί ότι οποιαδήποτε ευριστική αναζήτηση μπορεί να ενσωματωθεί στον MAB θνητότητας. Η εμπειρική αξιολόγηση έγινε με τρεις διαφορετικές κατανομές για τα κέρδη των επιλογών.

- **MAB Μεταβλητού Χρόνου:** Η εργασία “Online ContextAware Recommendation with Time Varying MultiArmed Bandit,”[137] εστιάζει στην αντιμετώπιση των χρονικά εξαρτώμενων αλλαγών του κέρδους σε προβλήματα MAB περιεχομένου. Οι συγγραφείς προτείνουν ένα δυναμικό πλαίσιο για το περιεχόμενο και μοντελοποιούν την δυναμική συμπεριφορά του κέρδους ως ένα σύνολο στοιχείων τυχαίου περιπάτου (*random walk*). Το κάθε στοιχείο αυτό αποθηκεύει για μια επιλογή την τρέχουσα κατάσταση της. Οι πληροφορίες των καταστάσεων των επιλογών ενσωματώνονται στην στρατηγική επιλογής του MAB έτσι ώστε να γίνονται οι συστάσεις. Το κέρδος μιας επιλογής a μοντελοποιείται ως ένας γραμμικός συνδυασμός των χαρακτηριστικών περιεχομένου x_t και του διανύσματος συντελεστών θ_t . Για να μπορεί να μετρηθεί η μεταβολή του περιεχομένου, οι συγγραφείς υποθέτουν ότι το x_t διαθέτει ένα σταθερό και ένα μεταβαλλόμενο μέρος όπου το σταθερό c_t δημιουργείται μέσω μιας *prior* κατανομής και το μεταβαλλόμενο είναι παράγωγο ενός κανονικού *Gaussian* τυχαίου περιπάτου $rand_{t,k}$ και μιας παραμέτρου μ_k . Αφού ληφθεί το κέρδος μια επιλογής a δεδομένου ενός περιεχομένου x_t , ο αλγόριθμος μαθαίνει τις παραμέτρους που βασίζονται στις εκτιμήσεις του μ_k από όλα τα στοιχεία και υπολογίζει το κέρδος βάσει αυτών των παραμέτρων. Στην συνέχεια οι εκτιμήσεις των στοιχείων ενημερώνονται και υπολογίζονται βάρη για κάθε στοιχείο. Γίνεται η δειγματοληψία των στοιχείων βάσει των συντελεστών βάρους τους και ενημερώνονται όλες οι στατιστικές πληροφορίες.

Συνδυαστικές Μέθοδοι

Στην προσέγγιση αυτή συνδυάζονται πολλαπλοί αλγόριθμοι MAB αντί να δημιουργείται ένα καθολικό μοντέλο πρόβλεψης. Οι συνδυαστικές μέθοδοι επομένως συλλέγουν τις προβλέψεις διαφορετικών πολιτικών ώστε να αποκτήσουν μια τελική πρόβλεψη[91].

- ***HyperTS* , *HyperTSFB***: Η προσέγγιση αυτή εφαρμόζει το παράδειγμα μετα-αλγορίθμων το οποίο πρακτικά κατασκευάζει έναν υπερ-αλγόριθμο ο οποίος συντονίζει την εργασία κάποιων άλλων αλγορίθμων σε βασικό επίπεδο και παρατηρεί την σημαντικότητα του καθενός στο πλαίσιο την αναζήτησης/αξιοποίησης βάσει της βαθμολόγησης του εκάστοτε χρήστη[110]. Εξαιτίας της φύσης της συνδυαστικής στρατηγικής που κάνει εκτίμηση της απόδοσης της κάθε βασικής πολιτικής, η στρατηγική αυτή έχει αποδειχθεί πολύ καλή για την αντιμετώπιση του προβλήματος της αργής εκκίνησης. Η ιδέα της συνδυαστικής πολιτικής είναι να έχει ένα σύνολο ενεργειών A και ένα σύνολο πολιτικών $\Pi = \{\pi_1, \dots, \pi_n\}$ έτσι ώστε να εντοπίζει την πολιτική που είναι πιο κοντά σε μια πολιτική $\pi^* = \arg \max_{\pi_i \in \Pi} E[r_{\pi_i}]$. Στην εργασία παρουσιάζονται δύο αλγόριθμοι, ο *HyperTS* και ο *HyperTSFB*. Ο *HyperTS* κάνει εκτίμηση του αναμενόμενου κέρδους για κάθε πολιτική π χρησιμοποιώντας την μέθοδο *Monte Carlo*. Το περιεχόμενο x_1, \dots, x_n λαμβάνεται από την κατανομή στην οποία η πολιτική π έχει επιλεγεί. Για είσοδο περιεχομένου x_i , η πολιτική π_i επιλέγει τον πληθυσμό α_j και λαμβάνει κέρδος r_j . Λαμβάνοντας υπόψη όλες τις κατανομές η εκτίμηση Monte Carlo είναι $\hat{E}[r_{\pi_i}] \sim \text{Beta}(1 + \alpha_{\pi_i}, 1 + \beta_{\pi_i})$, όπου $\alpha_{\pi_i} = \sum r_j$ and $\beta_{\pi_i} = n - \alpha_{\pi_i}$. Σε κάθε επανάληψη το r_i είναι ένα δείγμα από την $E[r_{\pi_i}]$ μέσω της *Beta* κατανομής. Η τελική πολιτική που θα επιλεγεί είναι αυτή με το μεγαλύτερο r_i . Συνεχίζοντας, ο *HyperTSFB* βελτιώνει την απόδοση της εκτίμησης του *HyperTS* μέσω της πλήρους αξιοποίησης της κάθε λαμβανόμενης βαθμολόγησης για την αναμενόμενη εκτίμηση του κέρδους. Οι δύο αλγόριθμοι έχουν καταφέρει να επιτύχουν συγκρίσιμη απόδοση απέναντι σε κορυφαίες βασικές πολιτικές.

- **BEER**: Η προσέγγιση αυτή επεκτείνει την δειγματοληψία Thompson με στόχο τον συντονισμό ενός συνόλου βασικών συστημάτων σύστασης για ηλεκτρονική διαφήμιση. Λαμβάνει υπόψη ρεαλιστικά σενάρια και υλοποιεί την TS όταν ούτε υπάρχει διαθέσιμη επιλογή ούτε όταν υπάρχει εγγύηση για σταθερό κέρδος[14]. Η μέθοδος αυτή είναι παρόμοια με την *HyperTS* στο πλαίσιο ότι κάθε επιλογή αναπαριστά μια πολιτική σύστασης που θα ταιριάζει στον τρόπο που αναζητά ο χρήστης αντικείμενα στην βάση, αλλά διαφέρει στο πως οι συγγραφείς μοντελοποιούν το κέρδος των επιλογών αυτών. Οι επιλογές, που είναι απλά συστάσεις περιεχομένου, διαχωρίζονται σε ομάδες μέσω IDF (*Inverse Document Frequency*). Ο προτεινόμενος αλγόριθμος BEER (Bandit Ensemble for Ecommerce Recommendations) χρησιμοποιείται μαζί με δειγματοληψία Thompson επιπλέον της μεθόδου διαχωρισμού του. Όπως προτείνουν και οι συγγραφείς, τα στοιχεία σύστασης του συνδυασμού (ensemble) μπορούν να δημιουργηθούν αυτόματα από τα υπάρχοντα στοιχεία αντικειμένων ή το ιστορικό αλληλεπιδράσεων των χρηστών με αντικείμενα. Αυτός ο τύπος αρχιτεκτονικής καταφέρνει να έχει δυνατότητα κλιμάκωσης καθώς ο αριθμός επιλογών δεν εξαρτάται από το πλήθος των αντικειμένων.
- **PDMS**: Μια άλλη προσέγγιση ονομάζεται Επιλογή Μοντέλου Πρόβλεψης (Prediction Model Selection) και αντιμετωπίζει το πρόβλημα αργής εκκίνησης χρηστών βασιζόμενη στην βαθμολόγηση όλων των χρηστών ώστε να επιλέξει την πιο κατάλληλη σύσταση. Οι επιλογές θεωρούνται ως ομάδες χρηστών που υπολογίζονται μέσω matrix factorization και εφαρμόζουν *UCB* και *e-greedy* αλγορίθμους για την επιλογή πολιτικής[32].

Υπάρχουν πάρα πολλές επιπλέον εφαρμογές MAB στα Συστήματα Συστάσεων με την μέθοδο Thomson Sampling να χρησιμοποιείται επί το πλείστον για την δημιουργία συστάσεων και μπορεί να βρεθεί στην βιβλιογραφία ως Active Thompson Sampling[11, 51, 74]

Αντιμετωπίζοντας τις προκλήσεις των Συστημάτων Σύστασης με αλγορίθμους MAB

Όπως φάνηκε και στην κατηγοριοποίηση των MAB αλγορίθμων υπάρχουν αρκετές προτάσεις στην βιβλιογραφία σχετικά με την εφαρμογή τους πάνω στις τεχνικές Συστημάτων Σύστασης. Παρακάτω γίνεται μια ομαδοποίηση τους αυτήν την φορά βάσει των κύριων προβλημάτων των συστημάτων σύστασης που καλούνται να επιλύσουν καθώς είναι ένα σημαντικό πρίσμα από το οποίο η έρευνα πάνω στους MAB αλγορίθμους εξελίσσεται:[104]

Αραιότητα και Κλιμάκωση Αλγορίθμων

Παραδοσιακά, ενώ αραιότητα αναφέρεται ως η απουσία δεδομένων εντός ενός συνόλου δεδομένων, η κλιμάκωση αφορά την δυσκολία να προταθούν αλγόριθμοι ικανοί να ανταποκρίνονται γρήγορα απέναντι σε ένα τεράστιο πλήθος δεδομένων. Ειδικά στο πεδίο των συστημάτων σύστασης, η αραιότητα και η κλιμάκωση, είναι συνέπειες του μεγάλου ρυθμού αύξησης των δεδομένων μέσω της ταχείας εισαγωγής χρηστών και αντικειμένων στα πραγματικά συστήματα. Όσο πιο πολλά είναι τα αντικείμενα εντός του συστήματος, τόσο πιο δύσκολο είναι οι χρήστες να έχουν δώσει βαθμολόγηση για όλα και επομένως το πρόβλημα της αραιότητας αυξάνεται. Αντίστοιχα όσο πιο πολλοί χρήστες εισέρχονται στο σύστημα τόσο πιο δύσκολο είναι να γίνεται η επεξεργασία όλων των δεδομένων έτσι ώστε να παρέχονται συστάσεις για όλους τους χρήστες. Στην βιβλιογραφία υπάρχουν αρκετές προτάσεις για την αντιμετώπιση της πρόκλησης αυτής αλλά ελάχιστες από αυτές συμπεριλαμβάνουν αλγορίθμους MAB. Κάποιες προτάσεις σχετικές με MAB αναφέρονται σε αλγορίθμους ταξινόμησης (cluster-based) με στόχο την μείωση του μεγάλου αριθμού διαστάσεων (πχ προτιμήσεις) σε έναν πιο διαχειρίσιμο αριθμό διαστάσεων. Οι πιο σχετικές εργασίες σε αυτό το επίπεδο είναι ο αλγόριθμος CLUB[30], DynUCB[84] και COFIBA[71]. Συγκεκριμένα στην εργασία "*Dynamic Clustering of Contextual Multi-Armed Bandits*"[84] παρουσιάζονται μελέτες που διαπραγματεύονται το πρόβλημα της αραιότητας μέσω της ταξινόμησης των

χρηστών. Άλλες εργασίες όπως "*Dual Graph Attention Networks for Deep Latent Representation of Multifaceted Social Effects in Recommender Systems*"[132] προτείνουν την αντιμετώπιση της αραιότητας μέσω της διαθέσιμης πληροφορίας κοινωνικών αλληλεπιδράσεων μεταξύ χρηστών και αντικειμένων. Πράγματι, όλες οι παραπάνω τεχνικές βοηθούν να βελτιωθεί το πρόβλημα της κλιμάκωσης αλγορίθμων MAB μέσω της εκπαίδευσης των μοντέλων τους για ένα σύνολο χρηστών. Από την άλλη, υπάρχουν εργασίες που εστιάζουν μεμονωμένα στο πρόβλημα της κλιμάκωσης. Για παράδειγμα στο "*Finite-time Analysis of the Multiarmed Bandit Problem*"[5] γίνεται πρόταση για μια παράλληλη εκδοχή του κλασικού UCB αλγορίθμου και στα [113, 83] παρουσιάζονται κατανομημένοι MAB αλγόριθμοι για ΣΣ.

Αργή Εκκίνηση

Το πρόβλημα της *Αργής Εκκίνησης* (Cold-start) είναι το πιο διάσημο πρόβλημα στην βιβλιογραφία των συστημάτων σύστασης και αναφέρεται στην απουσία πληροφορίας που καθιστά δύσκολη την εξατομικευμένη σύσταση. Υπάρχουν τρεις περιπτώσεις αργής εκκίνησης:

- Η εισαγωγή νέου χρήστη όπου δεν υπάρχει αρκετή πληροφορία για τον ίδιο ώστε να μπορέσει το σύστημα να δημιουργήσει κάποιο εξατομικευμένο προφίλ.
- Η εισαγωγή νέου αντικειμένου όπου δεν υπάρχουν βαθμολογίες από χρήστες για αυτό.
- Η εισαγωγή νέας κοινότητας όπου αναφέρεται στην έναρξη του συστήματος όπου μπορεί να υπάρχουν μεν τα αντικείμενα αλλά δεν υπάρχουν ακόμα χρήστες για να τα βαθμολογήσουν.

Στην βιβλιογραφία όλες οι παραπάνω περιπτώσεις έχουν αντιμετωπιστεί μέσω της αξιοποίησης MAB αλγορίθμων. Καθώς οι ερευνητές συνήθως αξιολογούν την ικανότητα εκμάθησης κάποιου MAB αλγόριθμου, προσομοιώνουν σενάρια έναρξης όπου δεν υπάρχει ακόμα πληροφορία για τους χρήστες και τα αντικείμενα με

αποτέλεσμα η μεγαλύτερη έρευνα να αφορά το κομμάτι της εισαγωγής νέας κοινότητας[32, 93, 105]. Παρόλα αυτά υπάρχουν εργασίες που κάνουν προτάσεις συγκεκριμένα πάνω στο κομμάτι της εισαγωγής νέου χρήστη ή αντικειμένου. Το πρόβλημα εισαγωγής νέου αντικειμένου συνήθως αντιμετωπίζεται μέσω υβριδικών προσεγγίσεων οι οποίες συνδυάζουν MAB αλγορίθμους περιεχομένου[109, 122]. Αυτές οι προσεγγίσεις χρησιμοποιούν τα χαρακτηριστικά των αντικειμένων ώστε να συσχετίζουν αντικείμενα με προτιμήσεις χρηστών όταν δεν υπάρχει καθόλου πληροφορία από βαθμολογίες χρηστών για τα αντικείμενα. Ανάλογες προσεγγίσεις υπάρχουν και για την περίπτωση εισαγωγής νέων χρηστών όπου χωρίζονται σε δύο κατηγορίες: τους MAB αλγορίθμους περιεχομένου και τις υβριδικές τεχνικές. Η πρώτη κατηγορία εφαρμόζει πληροφορίες περιεχομένου για τους χρήστες και τα αντικείμενα (πχ ηλικία, φύλο) ώστε να επιτρέψει στους χρήστες να αναζητήσουν τις επιλογές του συστήματος[84, 94, 137]. Η δεύτερη κατηγορία προτείνει την εφαρμογή επιλογής μοντέλου ώστε να αντιμετωπιστούν τα πρώτα δύο στάδια της εμπειρίας του χρήστη και στην συνέχεια την αλλαγή σε κάποιο άλλο μοντέλο όταν το σύστημα θα διαθέτει αρκετή πληροφορία για τον χρήστη[32, 43, 44, 64, 65, 76].

Ασφάλεια

Η αυξανόμενη ανησυχία για την ασφάλεια των χρηστών έχει επηρεάσει έντονα τα συστήματα ηλεκτρονικού εμπορίου και έχει ανοίξει αρκετές συζητήσεις και στον τομέα των συστημάτων σύστασης. Η επιτυχία των συστημάτων σύστασης συνδέεται άμεσα με την ικανότητα τους να δημιουργούν εξατομικευμένα προφίλ χρηστών με αποτέλεσμα πολλοί αλγόριθμοι να προσπαθούν να συλλέξουν όσο περισσότερα δεδομένα μπορούν για τους χρήστες. Αυτό είναι πολύ περισσότερο εμφανές στις περιπτώσεις των αλγορίθμων περιεχομένου. Με στόχο την βελτίωση της ποιότητας των συστάσεων, τέτοιοι αλγόριθμοι προσαρμόζονται δυναμικά στις προτιμήσεις των χρηστών και τις απαιτήσεις βάσει των πληροφοριών που έχουν συλλεχθεί από το ΣΣ. Παρόλα αυτά κάποιες από αυτές τις πληροφορίες μπορεί να είναι εμπιστευτικές καθώς μπορεί να σχετίζονται με παρελθοντικές επιλογές των χρηστών που θα θέλουν να κρατηθούν μυστικές λόγω του περιεχομένου είτε μπορεί

να αφορούν ευαίσθητες πληροφορίες τους όπως διευθύνσεις. Εξαιτίας των παραπάνω έχουν συζητηθεί αρκετές πολιτικές ιδιωτικότητας με στόχο την αλλαγή των αλγορίθμων σύστασης. Στο πλαίσιο τις συνεργασίας των MAB αλγορίθμων με τα ΣΣ υπάρχουν κάποιες σχετικές προτάσεις όπως στο *"Challenges in recommender systems: scalability, privacy, and structured recommendations"*[111] όπου οι συστάσεις πραγματοποιούνται στην συσκευή του χρήστη δίνοντας του τον πλήρη έλεγχο των πληροφοριών που θέλει να διαθέσει δημόσια. Στο *"Privacy-Preserving Bandits"*[77] προτείνεται η εφαρμογή ενός μοντέλου ιδιωτικότητας που επιτρέπει στον χρήστη να αναμειγνύεται με άλλους χρήστες με τρόπο που αντικαθιστά τα δεδομένα του με δεδομένα άλλων χρηστών όταν αυτό δεν αλλάζει το αποτέλεσμα της σύστασης. Με αυτόν τον τρόπο προστατεύεται η ιδιωτικότητα ενός ολόκληρου συνόλου χρηστών που έχουν παρόμοιες προτιμήσεις. Μια άλλη προσέγγιση έχει προταθεί στο *"A Privacy-Preserving Distributed Contextual Federated Online Learning Framework with Big Data Support in Social Recommender Systems"*[142] όπου αναφέρεται στην αποθήκευση του κεντρικού μοντέλου σύστασης σε κάποιον cloud server, στο οποίο γίνεται η ενημέρωση των παραμέτρων εκμάθησης από το κάθε τοπικό σύστημα σύστασης. Με αυτόν τον τρόπο τα δεδομένα εκπαίδευσης διατηρούνται στην τοπική συσκευή του εκάστοτε χρήστη. Άλλες εργασίες περί ιδιωτικότητας έχουν συζητηθεί ειδικώς πάνω σε MAB αλγορίθμους οι οποίες προτείνουν ένα σύστημα που ενημερώνει τα τοπικά συστήματα μέσω της συλλογής βαθμολογήσεων από άλλα τοπικά συστήματα με ασφάλεια[45, 96].

Διαφάνεια

Η διαφάνεια βοηθά τον χρήστη να κατανοήσει τον τρόπο με τον οποίο τα ΣΣ δημιουργούν τις συστάσεις τους. Όταν υπάρχει διαφάνεια, τα ΣΣ γίνονται πιο ελκυστικά και προπάντων πιο έμπιστα απέναντι στους χρήστες τους. Παραδοσιακά οι συστάσεις ακολουθούν δύο ορθογώνιες διαστάσεις. Η μία είναι ο τρόπος που γίνεται η επεξήγηση (πχ κείμενο, εικόνες), ο οποίος αντιπροσωπεύει την οπτική της αλληλεπίδρασης ανθρώπου – μηχανής και η άλλη είναι το μοντέλο που δημιουργεί

αυτές τις επεξηγήσεις, η οποία αντιπροσωπεύει την οπτική της μηχανικής μάθησης. Δύο εργασίες που σχετίζονται με την επεξήγηση των ΣΣ μέσω της εφαρμογής της ενισχυτικής μάθησης είναι οι *"Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits"*[80] και *"TEM: Tree-enhanced Embedding Model for Explainable Recommendation"*[125]. Η πρώτη προτείνει ότι οι χρήστες ανταποκρίνονται διαφορετικά και δυναμικά απέναντι στις επεξηγήσεις και επομένως μια προσέγγιση MAB θα βοηθούσε στην εύρεση του καταλληλότερου τρόπου για τον εκάστοτε χρήστη. Συγκεκριμένα προτείνονται μέθοδοι όπου αξιολογούν τους τρόπους επεξήγησης στους οποίους οι χρήστες ανταποκρίνονται καλύτερα, τα περιεχόμενα και την ισορροπία ανάμεσα στην αναζήτηση και την αξιοποίηση απέναντι στην αβεβαιότητα. Η εργασία αυτή δείχνει πως όσο ευεργετική είναι η αντιμετώπιση του διλήματος για τις συστάσεις τόσο είναι και για την επεξήγηση. Αντίστοιχα η δεύτερη εργασία προτείνει ένα *αγνωστικιστικό μοντέλο ενισχυτικής μάθησης* το οποίο δημιουργεί προτάσεις επεξήγησης για κάθε μοντέλο σύστασης. Σε αυτόν τον σχεδιασμό, το σύστημα είναι υπεύθυνο για την δημιουργία επεξηγήσεων και την πρόβλεψη βαθμολογήσεων βάσει των επεξηγήσεων. Με αυτόν τον τρόπο το σύστημα μαθαίνει να δημιουργεί επεξηγήσεις με καλή ποιότητα παρουσίασης μέσω της βελτίωσης του αναμενόμενου κέρδους.

Αξιολόγηση των MAB στο πεδίο των Συστημάτων Σύστασης

Η αξιολόγηση των MAB αλγορίθμων είναι ένα δύσκολο εγχείρημα[24]. Επιπλέον για τους αλγορίθμους που δημιουργούνται βάσει online διαδικασίας εκμάθησης είναι ακόμα πιο δύσκολο και είναι πολύ σημαντικό να μπορεί να γίνεται η ορθή αξιολόγηση τους προτού ξεκινήσουν να εφαρμόζονται στην παραγωγή. Υπάρχουν κάποιοι καθιερωμένοι τρόποι αξιολόγησης σε offline ή ψευδο-online εικονικά περιβάλλοντα.

Offline Αξιολόγηση

- *Επαναληπτική Μέθοδος*: Πολλές δημοσιεύσεις αναφέρονται στην *Επαναληπτική Μέθοδο αξιολόγησης*, μια από τις καθιερωμένες μεθόδους αξιολόγησης στην βιομηχανία, η οποία έχει προταθεί ως μια offline μέθοδος που βασίζεται στα δεδομένα και έχει αποδειχθεί ότι δεν διαθέτει συστηματικά σφάλματα (bias). Η τεχνική αυτή υποθέτει ότι οι πληθυσμοί είναι ανεξάρτητοι και ισόνομοι (i.i.d.) και κάθε επιλογή τους γίνεται ομοιόμορφα τυχαία από την πολιτική σε κάθε επανάληψη. Η βασική ιδέα αυτής της προσέγγισης είναι να υπάρχει πρόσβαση σε ιστορικά δεδομένα κατά την διάρκεια της αξιολόγησης και ο αλγόριθμος να λαμβάνει υπόψη μόνο τις ενέργειες που επιλέγουν το ίδιο αντικείμενο που υπάρχει και στα ιστορικά δεδομένα, αγνοώντας τα υπόλοιπα. Όπως έχει σημειωθεί στο “*Online ContextAware Recommendation with Time Varying MultiArmed Bandit*”[137] η βασική ιδέα της Επαναληπτικής Μεθόδου είναι να επαναλαμβάνει τον κάθε χρήστη που επισκέπτεται τον αλγόριθμο κατά την διάρκεια της αξιολόγησης. Τυπικά, αν η πολιτική A επιλέξει τον ίδιο πληθυσμό με τον πληθυσμό που έχει αποθηκευτεί στο παρελθόν από την πολιτική L , τότε ενημερώνονται τα στατιστικά στοιχεία, διαφορετικά δεν συμβαίνει τίποτα. Το μειονέκτημα της μεθόδου αυτής είναι ότι χρειάζεται πολύ μεγάλο πλήθος αρχικών ιστορικών δεδομένων, ειδικά εάν υπάρχουν πολλές επιλογές, διότι ο αριθμός των επιλογών που θα αγνοηθούν μπορεί να αυξηθεί ραγδαία. Εάν έχουμε k επιλογές και T δείγματα, τότε το μέγεθος το νέου συνόλου δεδομένων θα είναι T/k .

- *Bootstrapped Αξιολόγηση MAB Περιεχομένου*: Η προσέγγιση αυτή υποθέτει ένα πλαίσιο που ονομάζεται δυναμική σύσταση και στο οποίο είναι διαθέσιμες μόνο κάποιες δεκάδες αντικειμένων κάθε στιγμή. Βασίζεται στην Επαναληπτική Μέθοδο αλλά χρησιμοποιεί *bootstrap* έτσι ώστε να δημιουργεί σύνολα δεδομένων που χρησιμοποιούνται για την αξιολόγηση. Συγκεκριμένα, από ένα σύνολο δεδομένων μεγέθους T με K πιθανές επιλογές μιας ενέργειας σε κάθε επανάληψη, δημιουργούνται B σύνολα μεγέθους T/K χρησιμοποιώντας *δειγματοληψία με αντικατάσταση* με μια μη παραμετρική διαδικασία *bootstrap*[85]. Για τα νέα δημιουργούμενα υποσύνολα χρησιμοποιείται η κλασική Επαναληπτική Μέθοδος για να εκτιμήσει το κέρδος προ δοκιμών. Οι συγγραφείς συμπεραίνουν ότι εξαιτίας αυτού, ο αλγόριθμος A αξιολογείται σε T επίπεδα (records) κατά μέσο όρο. Συγκρίνοντας την μέθοδο αυτή με την Επαναληπτική, συμπεραίνουν ότι η πρώτη συγκλίνει γρηγορότερα και διαθέτει υψηλότερη ευστοχία για πραγματικά δεδομένα.
- *Άμεση Μέθοδος (Direct Method)*: Η μέθοδος αυτή έχει προταθεί από τους Beygelzimer και Langford και εφαρμόζει ένα μοντέλο παλινδρόμησης ώστε να μάθει την συνάρτηση κέρδους. Το κέρδος εξαρτάται από τις ενέργειες και το περιεχόμενο. Η μέθοδος επιτρέπει την επαναχρησιμοποίηση οποιουδήποτε υπάρχοντος αλγορίθμου επιβλεπόμενης δυαδικής ομαδοποίησης σε πλαίσιο ελλιπούς πληροφόρησης[8]. Η αξία της πολιτικής δίνεται από

$$\hat{V}_{DM}(\pi) = \frac{1}{n} \sum_{k=1}^n \sum_{\alpha \in A} \pi(\alpha|x_k) (\hat{r}(x_k, \alpha))$$

με $\hat{r}(x, \alpha)$ να είναι μια προσέγγιση του κέρδους και π η πολιτική υπό αξιολόγηση.

- *Doubly Robust Policy Evaluation*: Σύμφωνα με τους συγγραφείς η μέθοδος αυτή είναι ένας τύπος στατιστικής εκτίμησης από ελλιπή δεδομένα όπου εάν ένας από δύο εκτιμητές είναι ορθός, τότε η εκτίμηση δεν έχει συστηματικό σφάλμα. Η προσέγγιση αυτή προτείνει έναν μη στατικό εκτιμητή πολιτικής βασισμένο στην επαναληπτική μέθοδο, με βελτιωμένο ποσοστό αποδοχής (όπως αναφέρθηκε στην Επαναληπτική Μέθοδο, η αγνόηση μεγάλου πλήθους

επιλογών είναι ένας σημαντικός περιορισμός). Η εκτιμώμενη τιμή της πολιτικής π είναι

$$\hat{V}_{DR}(\pi) = \frac{1}{n} \sum_{k=1}^n \left[\hat{r}(x_k, v) + \frac{v(x_k | \alpha_k)}{\hat{\mu}(x_k | \alpha_k)} \times (r_k - \hat{r}(x_k, \alpha_k)) \right]$$

όπου

$$\hat{r}(x_k, v) = \sum_{\alpha \in A} v(\alpha | x) (\hat{r}(x_k, \alpha_k))$$

είναι η εκτιμώμενη τιμή της πολιτικής που προέρχεται από το \hat{r} [31]. Ο εκτιμητής DR χρησιμοποιεί μια σταθμισμένη εκδοχή των κερδών καθώς και την συνάρτηση εκτιμώμενου μέσου κέρδους ώστε να μειώσει την διασπορά[101]. Κατά κάποιο τρόπο ο DR συνδυάζει τις ιδέες των εκτιμητών *Direct Method* και *Inverse Probability Weighting*.

Περιβάλλοντα Προσομοίωσης

Ένας άλλος τρόπος για την δοκιμή MAB είναι η δημιουργία προσομοιώσεων online πλαισίων. Για παράδειγμα μπορεί να προσομοιωθούν τα κλικ που θα περιλαμβάνονται σε ένα διάνυσμα το οποίο θα δημιουργείται μέσω κάποιας κατανομής. Η μέθοδος προσομοίωσης θεωρείται ότι παρέχει καλύτερα αποτελέσματα σε σχέση με την επαναληπτική μέθοδο όταν τα διαθέσιμα δεδομένα αντικειμένων για συστάσεις είναι πολλά[137]. Οι προσομοιώσεις και τα συνθετικά δεδομένα χρησιμοποιούνται πάρα πολύ συχνά σε δημοσιεύσεις. Παρόλα αυτά η δημιουργία τους είναι αρκετά χρονοβόρα και δεν υπάρχει βεβαιότητα ότι αντικατοπτρίζουν τα πραγματικά δεδομένα και κατ' επέκταση την πραγματική απόδοση[69]. Υπάρχουν βιβλιοθήκες τεχνητής νοημοσύνης που παρέχουν διαφορετικά συνθετικά περιβάλλοντα για να δημιουργήσουν και να δοκιμάσουν αλγόριθμους ενισχυτικής μάθησης και επομένως εφαρμόζουν και στους MAB αλγόριθμους¹².

¹<https://gym.openai.com>

²http://github.com/tensorflow/agents/tree/master/tf_agents/bandit

Κριτήρια Αξιολόγησης

Στην βιβλιογραφία, τα κριτήρια αξιολόγησης είναι το ορισμένο πρωτόκολλο για να μπορεί να μετρηθεί το εάν μια τεχνική σύστασης είναι αποτελεσματική ή όχι. Καθ' όλη εξέλιξη της έρευνας πάνω στα συστήματα συστάσεων, τα κριτήρια αυτά ενημερώνονται και βελτιώνονται διαρκώς μέσω των δύο βασικών πλαισίων μέτρησης της ποιότητας των ΣΣ: την offline και την online δοκιμασία. Ενώ τα offline πειράματα εστιάζουν στην εύστοχη πρόβλεψη για τις επιλογές των χρηστών, τα online πειράματα στοχεύουν στην μέτρηση της επιχειρηματικής αξίας του συστήματος, δηλαδή την αξία που θα δώσει η τεχνική υπό δοκιμή στο ήδη υπάρχον σύστημα συστάσεων. Παρόλα αυτά, εξαιτίας των AB δοκιμών με πραγματικούς χρήστες, η online αξιολόγηση έχει μεγαλύτερο κόστος σε σχέση με την offline και συνήθως απαιτεί ένα πραγματικό σύστημα ώστε να συλλέξει τα δεδομένα των αλληλεπιδράσεων των χρηστών. Επομένως είναι συνηθισμένη πρακτική η αξιοποίηση των offline μεθόδων αξιολόγησης ειδικά για την περιπτώσεις που πρόκειται για την αξιολόγηση νέων τεχνικών σύστασης και εφαρμογών. Εντοπίζονται τρία βασικά βήματα για την offline αξιολόγηση:

1. Επιλογή ή δημιουργία ενός συνόλου δεδομένων που προσομοιώνει το σενάριο σύστασης
2. Επεξεργασία των δεδομένων αυτών εάν είναι απαραίτητο
3. Μέτρηση των κύριων μετρικών ποιότητας βάσει της μεθοδολογίας που χρησιμοποιήθηκε για την προσομοίωση της κατανάλωσης αντικειμένων.

Τα κριτήρια αυτά είναι βασικά ώστε να παρέχεται η σύγκριση μεταξύ διάφορων τεχνικών που αξιολογούνται

Σύνολο Δεδομένων

Στην εφαρμογή των MAB αλγορίθμων στα συστήματα σύστασης εντοπίζονται συνηθέστερα κάποια διακριτά σύνολα δεδομένων που επιλέγονται ή δημιουργούνται ώστε να μετρηθεί η απόδοση των αλγορίθμων. Η μία συνηθισμένη πρακτική είναι να δημιουργούνται συνθετικά σύνολα δεδομένων που εφαρμόζονται ειδικά για κάποιες πειραματικές αναλύσεις. Παρόλα αυτά κάτι τέτοιο δεν προτείνεται καθώς τα δεδομένα αυτά δεν είναι δυνατόν να αναπαραχθούν ξανά και έτσι δεν μπορούν να παρέχουν αποτελέσματα εμπιστοσύνης. Η εφαρμογή πραγματικών δεδομένων σαφώς είναι η καλύτερη πρακτική για την ανάλυση της απόδοσης οποιασδήποτε νέας τεχνικής σύστασης και κάτι τέτοιο συμβαίνει στις περισσότερες δημοσιεύσεις που αφορούν την αξιολόγηση MAB πάνω σε ΣΣ.

Επεξεργασία Δεδομένων

Στα περισσότερα δεδομένα που αναφέρθηκαν προηγουμένως δεν υπάρχει μόνο η πληροφορία των χρηστών και των αντικειμένων εντός του συστήματος αλλά επιπλέον και οι βαθμολογίες των χρηστών για κάποια από αυτά τα αντικείμενα. Αυτές οι βαθμολογίες παραδοσιακά είναι άμεσες (πχ επιλογές χρηστών, βαθμολογία κλίμακας κτλ) αλλά υπάρχουν και σενάρια όπου είναι έμμεσες και επομένως δεν είναι ξεκάθαρο εάν ένας χρήστης προτιμάει ή όχι κάποιο αντικείμενο. Για τον λόγο αυτό υπάρχει το βήμα της επεξεργασίας έτσι ώστε να γίνει καθαρίσμα δεδομένων, δηλαδή αφαίρεση ελλιπούς ή ασφαούς πληροφορίας, θορύβου και πλεονασματικών στοιχείων καθώς και η κανονικοποίηση βαθμολογήσεων μέσω της αντιστοίχισης τους με μια σταθερή κλίμακα. Επίσης μπορούν να παρθούν κάποιες αποφάσεις επεξεργασίας δεδομένων βάσει υποθέσεων (πχ σε μια κλίμακα 1-5 οι βαθμολογίες από 3 και κάτω θα θεωρηθούν πως ο χρήστης δεν προτιμά το αντικείμενο) ή να οριστούν συναρτήσεις που κανονικοποιούν τις βαθμολογίες μέσω του μετασχηματισμού τους στο $[0,1]$ σύνολο. Δεν υπάρχει σαφής απάντηση εάν υπάρχουν καλύτερες πρακτικές για την επεξεργασία δεδομένων στο πλαίσιο της κανονικοποίησης. Παρόλα αυτά γίνεται όλο και πιο δημοφιλές μετά την ανάπτυξης των UCB και TS μεθόδων λόγω των στατιστικών υποθέσεων των κατανομών που χρησιμοποιούνται όπως η Beta και η Bayesian.

Μεθοδολογίες και Μετρικές

Εφόσον ένας MAB αλγόριθμος είναι μια online μέθοδος εκμάθησης (βλέπε *εικόνα 5 στην σελίδα 80*), τα περισσότερα offline πειράματα πρέπει να καθορίζουν τον τρόπο με τον οποίο θα προσομοιώνουν τις αλληλεπιδράσεις των χρηστών με τα αντικείμενα όπως και σε ένα πραγματικό σύστημα. Στην συνέχεια αναφέρονται πέντε διαφορετικές προσεγγίσεις που εφαρμόζονται κατά την διάρκεια ενός πειράματος για την αξιολόγηση ενός MAB αλγορίθμου.

- **Δοκιμές (Trials):** Όταν γίνεται μια εκτίμηση της βαθμολόγησης για ένα ζεύγος χρήστη-αντικειμένου σε κάθε αλληλεπίδραση[25, 46, 55, 72, 79].
- **Αλληλεπιδράσεις (Interactions):** Όταν περισσότερα του ενός αντικείμενα συστήνονται σε έναν χρήστη[27, 73, 107, 114, 121].
- **Replayer:** Όταν το σύστημα προσπαθεί να προβλέψει τα αντικείμενα που καταναλώνονται με χρονολογική σειρά βάσει του ιστορικού του χρήστη[69, 109, 124, 137].
- **Leave-one-out:** Όταν το σύστημα αφήνει εκτός ένα αντικείμενο κατά την διάρκεια του σταδίου της εκπαίδευσης και προσπαθεί να κάνει εκτίμηση αξιολόγησης για αυτό το αντικείμενο για κάποιον χρήστη[32].
- **Cross-Validation:** Όταν το σύστημα πραγματοποιεί την ίδια αλληλεπίδραση χρήστη-αντικειμένου πολλές φορές μέσω της εφαρμογής διαφορετικών δεδομένων στο στάδιο εκπαίδευσης[47, 126].

Φαίνεται ότι η πιο δημοφιλής μέθοδος είναι οι Δοκιμές και είναι πιθανό αυτό οφείλεται σε μια παραδοσιακή μεροληψία που υπάρχει ακόμα στα κριτήρια αξιολόγησης διάφορων εργασιών. Στην πρώτη δεκαετία ερευνών οι περισσότεροι αλγόριθμοι αξιολογούνται βάσει της προβλεπτικής τους ικανότητας, δηλαδή πόσο εύστοχα μπορούν να προβλέψουν τις επιλογές των χρηστών. Παρόλα αυτά σήμερα θεωρείται πως οι εύστοχες προβλέψεις, αν και πολύ σημαντικές, δεν αρκούν για την

εφαρμογή ενός επιτυχημένου συστήματος συστάσεων. Σε πολλές εφαρμογές οι χρήστες χρησιμοποιούν το σύστημα σύστασης για πολλά περισσότερα από την απλή ανταπόκριση του στις προσδοκίες τους. Βέβαια, το επιπλέον κομμάτι αυτό δεν εμπλέκεται ακόμη με την έρευνα πάνω στους MAB και την εφαρμογή τους στα συστήματα συστάσεων. Προς το παρόν η έρευνα αξιοποιεί τις κλασικές μετρικές κέρδους και απώλειας στο πεδίο των συστημάτων σύστασης και εστιάζει την εφαρμογή των MAB για την βελτιστοποίηση αυτών. Πρακτικά οι μετρικές αυτές αντιπροσωπεύουν το σφάλμα στην εκτίμηση της βαθμολόγησης ενός ζεύγους χρήστη-αντικειμένου. Άλλες μετρικές, όπως *precision*, *recall* και *nDCG*, κλασικές στο πεδίο συστάσεων, χρησιμοποιούνται πολύ λιγότερο για την αξιολόγηση των MAB.

MAB σε Συστήματα Συστάσεων στην Βιομηχανία

Στην συνέχεια θα αναφερθούν κάποιες μεγάλες και διάσημες εταιρίες οι οποίες χρησιμοποιούν MAB αλγόριθμους για να αντιμετωπίζουν προκλήσεις που υπάρχουν στα συστήματα συστάσεων που εφαρμόζουν.

- **Amazon:** Η δημοσίευση "*An Efficient Bandit Algorithm for Realtime Multivariate Optimization*"[49] αναφέρεται στο πρόβλημα της βελτιστοποίησης της διάταξης των σελίδων της Amazon μέσω ενός παραμετρικού Μπεϋζιανού μοντέλου το οποίο ενσωματώνει τις αλληλεπιδράσεις μεταξύ στοιχείων της σελίδας εφαρμόζοντας έναν αλγόριθμο *hill climbing* βελτιστοποίησης ώστε να προσεγγίσει μια Δειγματοληψία Thompson περιεχομένου. Κάθε πιθανή διάταξη απαρτίζεται από D διαφορετικά κομμάτια. Εάν κάθε κομμάτι έχει N εκδοχές, τότε το στοχαστικό πρόβλημα MAB διαμορφώνεται ως MAB με $N \times D$ επιλογές σε $t = 1, 2, \dots, T$ επαναλήψεις με διάνυσμα χαρακτηριστικών b_t (πληροφορίες χρήστη) και διάνυσμα επιλογών a_k (που αντιπροσωπεύει την διάταξη της σελίδας). Το τελικό διάνυσμα χαρακτηριστικών $x_{\alpha,t}$ δημιουργείται από τα δύο

διανύσματα χαρακτηριστικών που αναφέρθηκαν. Το κέρδος δίνεται μέσω μια γραμμικής κλίμακας του $x_{a,t}$ πάνω σε ένα διάνυσμα άγνωστων συντελεστών βάρους θ . Ο αλγόριθμος επιλέγει το βέλτιστο περιεχόμενο, αλλά είναι ταυτόχρονα υπεύθυνος για την εξατομίκευση του περιεχομένου της διάταξης της σελίδας. Προσομοιώθηκε το πείραμα σε 512 διαφορετικές διατάξεις συγκρίνοντας διαφορετικές εκδοχές του αλγορίθμου εναντίον MAB χωρίς περιεχόμενο N^D επιλογών και MAB χωρίς περιεχόμενο N επιλογών για κάθε ένα από τα D κομμάτια. Η μέθοδος αποδείχθηκε κατάλληλη να προσομοιώσει τις αλληλεπιδράσεις μεταξύ των κομματιών που παρουσιάζονται μαζί στην σελίδα λαμβάνοντας υπόψη το περιεχόμενο.

- **Netflix:** Σύμφωνα με blog τεχνολογίας του Netflix, η εταιρεία χρησιμοποιεί MAB για την επιλογή εξατομικευμένης εικονογράφησης για κάθε ταινία που παρουσιάζεται στον χρήστη. Οι συγγραφείς υποστηρίζουν πως χρησιμοποιώντας MAB περιεχομένου ώστε να επιλεγεί η πιο κατάλληλη εικόνα από δεκάδες υποψήφιες ειδικά για έναν συγκεκριμένο χρήστη¹. Για να δημιουργηθεί το περιεχόμενο για τον MAB, οι συγγραφείς χρησιμοποιούν πληροφορίες όπως ο τίτλος, το είδος, το ιστορικό αλληλεπιδράσεων του χρήστη, δημογραφικά στοιχεία κτλ. Στην συνέχεια το μοντέλο ιεραρχεί τις εικόνες για κάθε περιεχόμενο προβλέποντας την πιθανότητα να συνδυαστεί η κάθε εικόνα με τον τίτλο που δόθηκε και την εικόνα που έχει την μεγαλύτερη πιθανότητα να επιλεγεί. Το σύστημα μπορεί να αναλυθεί στο offline κομμάτι, που αφορά την ανάθεση χαρακτηριστικών και την εκπαίδευση του μοντέλου, και στο online κομμάτι που αφορά την πολιτική αναζήτησης-αξιοποίησης, την καταγραφή πληροφοριών περιεχομένου και την δημιουργία συστάσεων. Η πληροφορία περιεχομένου για κάθε απόφαση αποθηκεύεται έτσι ώστε γίνει κατανοητό γιατί δημιουργήθηκε η κάθε σύσταση. Η βασική ιδέα για την εικονογράφηση είναι να μεγιστοποιηθεί η πιθανότητα επανάληψης της παρακολούθησης για ένα αντικείμενο, όπου η πιθανότητα αυτή υπολογίζεται ως η διαφορά της πιθανότητας παρακολούθησης όταν ο τίτλος εμφανίζεται στην εικονογράφηση σε σχέση με όταν απουσιάζει.

¹<https://netflixtechblog.com/artworkpersonalizationc589f074ad76>

- **Spotify:** Το Spotify διαθέτει έναν μηχανισμό στις συστάσεις του τον οποίο τον ονομάζουν «recsplanations» και χρησιμοποιείται ώστε να βοηθάει τους χρήστες να κατανοούν τους λόγους που λαμβάνουν συστάσεις. Ο μηχανισμός αυτός είναι ένας συνδυασμός επεξηγήσεων και MAB αλγορίθμων που ονομάζονται BART όπως επεξηγεί και η εργασία "*Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits*" [80]. Ο BART αλγόριθμος διαθέτει μοντέλο κέρδους και περιεχομένου το οποίο επιβάλλει την γραμμικότητα του κέρδους

$$r(j, e, x) = \sigma(\theta_{global} + \theta^T x')$$

όπου $x' = [1_j^T, 1_e^T, x^T]^T$ είναι το συνολικό περιεχόμενο με πληροφορία σχετική με το αντικείμενο j , την επεξήγηση e και το περιεχόμενο x με 1_j να είναι διάνυσμα με μονάδα στο j και μηδέν αλλού; $\theta = [\theta_j^T, \theta_e^T, \theta_x^T]^T$, με θ να είναι ο συντελεστής της λογιστικής παλινδρόμησης. Για να ξεπεραστεί το μειονέκτημα της λογιστικής παλινδρόμησης στις συστάσεις (η επιλογή του αντικειμένου και της επεξήγησης δεν λαμβάνει υπόψη τα χαρακτηριστικά του χρήστη) και να υπάρχει καλύτερη εξατομίκευση σε αυτές, οι συγγραφείς εισάγουν σταθμισμένα αθροίσματα αλληλεπιδράσεων ανώτερης τάξης μεταξύ των στοιχείων στο διάνυσμα συνολικού περιεχομένου. Στην συνέχεια χρησιμοποιούν ένα CRM (*Counterfactual Risk Minimization*) για την εκπαίδευση του MAB. Τέλος, χρησιμοποιούν έναν e-greedy αλγόριθμο. Ο BART χρησιμοποιεί αναζήτηση βάσει συνθηκών, που σημαίνει ότι αποφασίζει εάν αξίζει η αναζήτηση ή η αξιοποίηση των αντικειμένων ανεξάρτητα από τις επεξηγήσεις, ενώ ταυτόχρονα κρατά το μοντέλο κέρδους το ίδιο.

- **Expedia Group:** Η Expedia χρησιμοποιεί MAB στο παραγωγικό περιβάλλον όπου οι επιλογές είναι εκδοχές ενός συστήματος σύστασης. Το σύστημα λαμβάνει τα στατιστικά των προβολών και των κλικ για κάθε εκδοχή και αποθηκεύει την κινητικότητα για την κάθε μία σε μια προσωρινή μνήμη [48]. Χρησιμοποιείται Δειγματοληψία Thompson για τον καταμερισμό της κινητικότητας και συγκρίνεται η απόδοση της κάθε νέας εκδοχής (βάσει CTR) σε σύγκριση με την εκδοχή που βρίσκεται στην παραγωγή.

- **Zozotown:** Zozo είναι μια τεράστια Ιαπωνική εταιρεία ηλεκτρονικής διαφήμισης στον τομέα της μόδας. Στην εργασία “*A largescale open dataset for bandit algorithms*”[101] οι συγγραφείς αποκαλύπτουν πως χρησιμοποιείται Δειγματοληψία Thompson μαζί με μια πολιτική τυχαίας επιλογής έτσι ώστε να γίνονται συστάσεις προϊόντων στους πελάτες τους. Η πιο σημαντική πληροφορία της δημοσίευσης είναι η πρόταση των συγγραφέων για έναν αλγόριθμο OBP ο οποίος διαχειρίζεται την έρευνα πάνω στους MAB αλγορίθμους. Εφαρμόζουν κάποιους από τους πιο γνωστούς MAB αλγορίθμους μαζί με διάφορους εκτιμητές πολιτικών πάνω σε ένα ανοικτό σύνολο δεδομένων. Πέρα από τον κλασικό εκτιμητή Επανάληψης, οι συγγραφείς εφαρμόζουν και συγκρίνουν και κάποιους άλλους εκτιμητές από την βιβλιογραφία και συμπεραίνουν ότι η εύρεση ενός σταθερού και βέλτιστου συστήματος αξιολόγησης είναι πάρα πολύ σημαντικό για την επιλογή μιας βέλτιστης πολιτικής. Η εργασία είναι σημαντική διότι προσφέρει μια εφαρμογή των MAB και των εκτιμητών η οποία μπορεί να επεκταθεί και να εφαρμοστεί και σε άλλα σύνολα δεδομένων.
- **Yahoo:** Υπάρχουν πολλά πειράματα στην Yahoo με MAB αλγορίθμους[69, 68]. Εδώ παρουσιάζεται το “*A Batched MultiArmed Bandit Approach to News Headline Testing*”[78] το οποίο χρησιμοποιεί Δειγματοληψία Thompson ώστε να επιλέξει μεταξύ διαφορετικών εκδοχών τίτλων με στόχο την μεγιστοποίηση του CTR. Υπάρχουν K διαφορετικοί τίτλοι που σχετίζονται με μια επιλογή. Η ιδέα είναι να βρεθεί η βέλτιστη επιλογή που μεγιστοποιεί το CTR αλλά ταυτόχρονα να λάβει υπόψη την φθορά των άρθρων (παλαιότητα). Οι συγγραφείς εισάγουν μια ροή Thompson Sampling (bTS) που συγχρονίζεται με το βέλτιστη βαθμολόγηση των χρηστών. Η επιλογή της ροής έγινε λόγω της υψηλής κινητικότητας στην ιστοσελίδα Yahoo διότι θα ήταν έντονο πρόβλημα η ενημέρωση του αλγορίθμου μετά από κάθε αλληλεπίδραση κάποιου χρήστη. Η ιδέα πίσω από την ροή είναι ότι μεταξύ των ροών η κατανομή Beta για κάθε επιλογή παραμένει ίδια. Η κίνηση μεταξύ επιλογών κατανέμεται βάσει τυχαίου δείγματος της Beta κατανομής για κάθε εισερχόμενο γεγονός

(αλληλεπίδραση). Τα στατιστικά ενημερώνονται στο τέλος της κάθε ροής. Υπάρχουν δύο τρόποι ενημέρωσης: ενημέρωση βάσει αθροισμάτων (μέτρηση των κλικ) και η κανονικοποίηση (αριθμός κανονικοποιημένων κλικ και μη κλικ). Τρία σημεία θα πρέπει να συγχρονίζονται στον bTS, το σημείο σταματήματος του αλγορίθμου, οι μέθοδοι ενημέρωσης που αναφέρθηκαν και η συχνότητα ενημέρωσης. Τα πλεονεκτήματα του bTS είναι τα εξής: Εκθέτει πολύ λιγότερους μη βέλτιστους τίτλους, επομένως βελτιώνει την εμπειρία του χρήστη, βοηθάει στην εύρεση την βέλτιστης επιλογής πολύ πιο γρήγορα, διότι έχει περισσότερα δείγματα να συγκρίνει μεταξύ των «καλών» επιλογών, και έχει κέρδος 3.69% κλικ συνολικά.

- **Deezer:** Η εταιρία έχει παρουσιάσει την χρησιμοποίηση MAB στο σύστημα σύστασης των playlists (συστήνεται στον χρήστη μια λίστα από L αντικείμενα από τα οποία μόνο 3 είναι φανερά)[7]. Η προσέγγιση αυτή θεωρεί ένα ημι-εξατομικευμένο πλαίσιο συστάσεων, με ένα αριθμό Q ομάδων σε N χρήστες ($Q \ll N$). Σε κάθε επανάληψη, σε ένα τυχαίο υποσύνολο χρηστών παρουσιάζεται μια πολιτική που ενημερώνει το μοντέλο της βάσει των βαθμολογήσεων των χρηστών. Οι πολιτικές πρέπει να συστήνουν μια λίστα από $L = 12$ playlists. Οι πολιτικές αυτές που χρησιμοποιήθηκαν σε πειράματα είναι *e-greedy* με *τμηματοποίηση* (*segmentation*), *explore then commit*, *KLUCB* (βάσει της απόκλισης *KullbackLeibler*), *Thompson Sampling* και γραμμική *Thompson Sampling*, και συγκρίθηκαν ενάντια σε μια τυχαία στρατηγική βάσης. Το πείραμα διεξήχθη σε δείγμα 900 χιλιάδων χρηστών με 862 playlists και στην συνέχεια εφαρμόστηκε στην εφαρμογή Deezer μέσω ενός A/B test. Ένα από τα εντυπωσιακά αποτελέσματα ήταν ότι αυτό το ημι-εξατομικευμένο σύστημα συστάσεων απέδωσε τόσο καλά όσο και το πλήρως εξατομικευμένο στα πλαίσια της ιεράρχησης των playlists υποθέτοντας καλή αρχική ομαδοποίηση.

Συμπεράσματα, Κατευθύνσεις και Μελλοντική Έρευνα

Η εφαρμογή των MAB στο πεδίο των Συστημάτων Σύστασης είναι σχετικά πρόσφατη και επομένως υπάρχουν ακόμα αρκετές ερευνητικές ευκαιρίες και βελτιώσεις για το μέλλον. Στο τελευταίο αυτό κεφάλαιο θα γίνει μια σύντομη και συγκεντρωτική περιγραφή όσον συζητήθηκαν στα προηγούμενα κεφάλαια και κάποιες προτάσεις για πιθανές κατευθύνσεις που θα μπορούσε να ακολουθήσει η έρευνα μελλοντικά. Αρχικά, θα αναφερθούν οι βασικές προσεγγίσεις που θα έπρεπε να μελετηθούν ή να βελτιωθούν περισσότερο. Στην συνέχεια θα γίνει μια αναφορά σχετική με τα κριτήρια αξιολόγησης βάσει των μέχρι τώρα δημοσιεύσεων και τέλος θα αναφερθεί περιληπτικά το πως είναι πιθανόν να αντιμετωπισθούν οι υπάρχουσες προκλήσεις.

Σχεδιασμός και μελέτη αλγορίθμων στο μέλλον

Σε σχέση με όλους τους γνωστούς MAB αλγορίθμους αυτό που έχει αποδειχθεί ότι αξιοποιείται επί το πλείστον είναι ο UCB και οι παραλλαγές του. Αυτό φαίνεται να συμβαίνει για δύο λόγους: Ο ένας αφορά την κλειστή μορφή του ορίου εμπιστοσύνης του UCB το οποίο έχει μελετηθεί εκτενώς στην βιβλιογραφία καθώς και η αποτελεσματικότητά του να μετρά την αβεβαιότητα της κάθε επιλογής[47, 141]. Ο δεύτερος λόγος είναι γιατί ο UCB είναι περισσότερο επιδεκτικός σε σχέση με τους υπολοίπους στο να συνδυάζει νέες έννοιες και ιδέες λόγω της χαμηλής πολυπλοκότητάς του και της υπερπαραμετρικής ανεξαρτησίας

που διαθέτει. Σήμερα ο UCB είναι ακόμα ο βασικός αλγόριθμος που αξιοποιείται στο πεδίο των συστημάτων σύστασης και ειδικά με την δυνατότητα του να συνδυάζεται με την μέθοδο Συνεργατικού Φιλτραρίσματος καταφέρνει πολύ υψηλές αποδόσεις[124, 130, 139]. Επίσης μέθοδοι όπως η ομαδοποίηση και το matrix factorization εφαρμόζονται συνήθως με UCB αλγορίθμους. Παρόλα αυτά η συνήθης επιλογή του UCB δεν συνεπάγεται ότι είναι και η μόνη αποδοτική επιλογή για το πεδίο των ΣΣ. Φαίνεται ότι και ο αλγόριθμος *Δειγματοληψίας Thompson* αποκτά αναγνωρισιμότητα και εφαρμογή στο πεδίο. Υπάρχουν έρευνες που προσπαθούν να πετύχουν την ίδια TS απόδοση από τα παραδοσιακά σενάρια ενισχυτικής μάθησης για το πεδίο των ΣΣ. Η βασική πρόκληση βέβαια είναι η προσαρμογή του αλγορίθμου κάθε φορά λόγω της πολυπλοκότητας του. Παραδοσιακά η κλασική TS δεν αποδίδει καλά λόγω των ειδικών παραμέτρων περιβάλλοντος του πεδίου των ΣΣ. Μια πρόσφατη δημοσίευση[140] προτείνει την ενσωμάτωση συνεργατικών επιδράσεων στο TS μέσω της αξιοποίησης των βαθμολογήσεων όλων των χρηστών εντός του ίδιου δυναμικού συνόλου (ομάδας) ώστε να εκτιμάται το αναμενόμενο κέρδος για το τρέχον περιεχόμενο. Παρόλο βέβαια της υψηλής απόδοσης που έδειξε η μέθοδος, οι συγγραφείς δεν συγκρίνουν τον νέο αλγόριθμο με τα παραδοσιακά UCB. Γενικότερα τέτοιου είδους βελτιώσεις παραμένουν ανοιχτές για μελλοντική έρευνα.

Συνεχίζοντας, φαίνεται πως το μεγαλύτερο μέρος της εφαρμογής των MAB στο πεδίο των συστάσεων αφορά τα πεδία των ταινιών και των ειδησεογραφικών άρθρων. Αυτό είναι λογικό καθώς στα δύο πεδία αυτά υπάρχει συνεχής αλληλεπίδραση των χρηστών με το σύστημα και έτσι η μοντελοποίηση των συστημάτων χρειάζεται online προσεγγίσεις. Παρόλα αυτά η εφαρμογή των MAB αφορά και άλλα πεδία όπως έχουν δείξει διάσημες εταιρίες στο πεδίο της μουσικής και των διαφημίσεων.

Κριτήρια Αξιολόγησης

Τα κριτήρια, που φαίνεται να χρησιμοποιούνται στις έρευνες για την αξιολόγηση MAB αλγορίθμων πάνω σε συστήματα συστάσεων, είναι παρόμοια με τις μετρικές που χρησιμοποιούνται για την αξιολόγηση των ίδιων των συστημάτων σύστασης τα οποία μετράνε την προβλεπτική ικανότητα του συστήματος. Αυτές οι μετρικές είναι τα Hit Rate, MAE, MSE και RMSE[9].

Παρόλα αυτά, τώρα πια είναι γνωστό πως οι ακριβείς προβλέψεις είναι σημαντικές αλλά δεν αρκούν για την αξία που προσφέρει ένα σύστημα σύστασης. Σε πολλές εφαρμογές οι χρήστες χρησιμοποιούν το ΣΣ για περισσότερα πράγματα από την εύρεση των προτιμήσεων τους, όπως για παράδειγμα για την εύρεση νέων διαφορετικών επιλογών που θα θέλανε να δοκιμάσουν. Επομένως οι μετρικές για την αξιολόγηση της πρόβλεψης δεν αρκούν να αξιολογήσουν την συνολική ποιότητα ενός συστήματος. Μια πρόταση είναι να εφαρμόζονται μετρικές ικανοποίησης χρηστών όπως η πρωτοτυπία και η ποικιλομορφία.

Επιπλέον ίσως η τωρινή αξιολόγηση μέσω δοκιμών (trials) (πχ η εκτίμηση της βαθμολόγησης μεταξύ ζευγών αντικείμενου – χρήστη σε κάθε αλληλεπίδραση) να μην είναι ιδανική για την αξιολόγηση της συνολικής ποιότητας του συστήματος. Η μέθοδος αυτή εστιάζει στην ικανότητα εκμάθησης του μοντέλου αλλά μια σύσταση δεν αφορά μόνο το σύστημα.

Ο ορισμός ενός ποιοτικού συστήματος σύστασης αφορά την βοήθεια που προσφέρει στους χρήστες να εντοπίσουν τα πιο σχετικά αντικείμενα σε σχέση με τις ανάγκες και τις προτιμήσεις τους. Για τον λόγο αυτό οι αλγόριθμοι MAB θα έπρεπε να συμπορεύονται μαζί μετά την έρευνα σχετικά με τα συστήματα σύστασης μέσω της μελέτης διάφορων μεθόδων και τεχνικών που εστιάζουν στις αλληλεπιδράσεις των χρηστών. Στην πραγματικότητα η αλληλεπίδραση με το σύστημα δεν γίνεται με έναν χρήστη την φορά όπως προσομοιώνουν οι μέθοδοι δοκιμών (trials). Εξάλλου η αντίδραση ενός συστήματος θα πρέπει να είναι άμεση διότι οι χρήστες δεν πρόκειται να περιμένουν πολλές δοκιμές (trials) έτσι ώστε να

λάβουν μια μεγάλη ενημέρωση για την δημοτικότητα των αντικειμένων ή την ποιότητα τους.

Η έλλειψη αυστηρά καθορισμένων κριτηρίων αξιολόγησης για την ποιότητα των MAB αλγορίθμων είναι ένα σοβαρό πρόβλημα και ανοιχτό για έρευνα. Επιπλέον είναι αρκετά δύσκολο να αναπαραχθούν ακόμα και απλοί MAB αλγόριθμοι εξαιτίας της έλλειψης ειδικών πρωτοκόλλων αξιολόγησης για το εκάστοτε σύστημα που εφαρμόζονται. Τις περισσότερες φορές οι ερευνητές μπορούν να αναπαράξουν αλγορίθμους όταν τα κριτήρια αξιολόγησης είναι καλώς ορισμένα για κάποιο ειδικό σενάριο και έτσι είναι εφικτό να εφαρμόζεται ορθά.

Αντιμετώπιση των Προκλήσεων

Οι προκλήσεις που συζητήθηκαν στην προηγούμενη ενότητα (*Αργή Εκκίνηση, Αραιότητα, Κλιμάκωση, Ασφάλεια, Επεξήγηση*) είναι αδιαμφισβήτητα οι πιο σχετικές προκλήσεις που μπορούν να αντιμετωπιστούν μέσω του πλαισίου των MAB αλγορίθμων. Το ερώτημα που προκύπτει και παραμένει ανοιχτό είναι πως μπορούν οι λύσεις που έχουν προταθεί να βελτιωθούν καθώς και εάν υπάρχουν επιπλέον νέες λύσεις που μπορούν να δοκιμαστούν. Σύμφωνα με την βιβλιογραφία οι υπάρχουσες λύσεις αντιμετωπίζουν μερικώς τα προβλήματα και κάποιες φορές οδηγούν και σε νέα προβλήματα.

Για παράδειγμα για το πρόβλημα της αραιότητας των δεδομένων έχουν προταθεί δύο πιθανές λύσεις. Η μία συνδυάζει MAB με μεθόδους ομαδοποίησης ή matrix factorization ώστε να γίνονται προβλέψεις βάσει των ομάδων των χρηστών και η άλλη εισάγει έναν MAB αλγόριθμο περιεχομένου ώστε να μειώσει την έλλειψη πληροφορίας που αφορά τους χρήστες και τα αντικείμενα. Στην πρώτη λύση οι βασικές τεχνικές που συχνά εφαρμόζονται, είναι μη κλιμακωτές ώστε να κάνουν online συστάσεις μέσω ομαδοποίησης των προτιμήσεων των χρηστών σε πραγματικό χρόνο. Στην δεύτερη πέρα από το πρόβλημα της κλιμάκωσης, υπάρχουν και ζητήματα ιδιωτικότητας που αφορούν το περιεχόμενο.

Μέχρι τώρα, οι πιο συχνά εφαρμοσμένες λύσεις για το πρόβλημα της αραιότητας και της κλιμάκωσης, σχετίζονται με την πρώτη μέσω της εφαρμογής ενημερώσεων στις ομάδες (ή τους latent factors) κατά την διάρκεια του σταδίου εκπαίδευσης ώστε να αποφεύγεται η επανεκπαίδευση της διαδικασίας[84, 93, 124, 133]. Παρόλα αυτά ακόμα και αυτή η πρόταση χρειάζεται βελτίωση έτσι ώστε να μπορεί να συμπεριλαμβάνει συνεργατικές βαθμολογήσεις των άλλων χρηστών με κλιμακωτό τρόπο.

Παρόμοια, παρόλο που έχει μελετηθεί εκτενώς στο MAB πλαίσιο, το πρόβλημα της αργής εκκίνησης παραμένει πρόκληση για τους υπάρχοντες αλγόριθμους. Συγκεκριμένα αυτό συμβαίνει διότι οι περισσότερες δημοσιεύσεις φαίνεται να παρερμηνεύουν τον ορισμό του προβλήματος όταν πρόκειται να παραθέσουν τρόπους αντιμετώπισης. Αυτό ισχύει διότι όπως έχει περιγραφεί και στην αντίστοιχη ενότητα η αργή εκκίνηση διαφέρει ανάλογα με την κατηγορίας της. Για παράδειγμα είναι διαφορετικές οι παράμετροι που συντελούν το πρόβλημα της αργής εκκίνησης σε μια νέα κοινότητα σε σχέση με την αργή εκκίνηση νέου χρήστη. Όταν ένας αλγόριθμος προτείνεται ώστε να αντιμετωπίζει το πρόβλημα της νέας κοινότητας, προσομοιώνει ένα νέο σύστημα όπου δεν υπάρχει καθόλου πληροφορία για τους χρήστες και τα αντικείμενα και ενδεχομένως δεν θα αντιμετώπιζε βέλτιστα την περίπτωση ενός νέου χρήστη σε ένα ήδη υπάρχον σύστημα που υπάρχουν σχετικές πληροφορίες.

Στο πρώτο σενάριο το σύστημα δεν γνωρίζει καν πιο αντικείμενο είναι το πιο δημοφιλές ή με την καλύτερη βαθμολόγηση ενώ στο δεύτερο υπάρχει ήδη μια κοινότητα που έχει καθορίσει αυτές τις πληροφορίες και μπορεί να επηρεάσει τις συστάσεις. Έτσι λοιπόν ένας MAB αλγόριθμος δεν μπορεί να διαχειρίζεται με τον ίδιο τρόπο τις δύο αυτές περιπτώσεις. Φαίνεται ότι στην βιβλιογραφία οι εργασίες για MAB αγνοούν το σενάριο των νέων χρηστών ή αντικειμένων για το πρόβλημα της αργής εκκίνησης και επομένως είναι ανάγκη στο μέλλον η έρευνα να επεκταθεί και σε αυτό το πλαίσιο με περισσότερη λεπτομέρεια. Μέσω της ανάλυσης της ποιότητας συστάσεων από την πλευρά των χρηστών φαίνεται πως οι περισσότεροι MAB δημιουργούν «κακές» συστάσεις στις αρχικές αλληλεπιδράσεις.

Συνεχίζοντας, υπάρχουν λίγες δημοσιεύσεις που ασχολούνται με τα προβλήματα τις ιδιωτικότητας και της επεξήγησης. Επειδή όμως οι προκλήσεις αυτές είναι πολύ σχετικές με την βιομηχανία είναι πολύ πιθανόν να λάβουν περισσότερη προσοχή στο μέλλον. Ενδέχεται η έλλειψη αυτή να αφορά τον τρόπο που εφαρμόζονται οι MAB στα συστήματα συστάσεων, καθώς συνήθως συνδυάζονται με παραδοσιακές τεχνικές ΣΣ και ίσως οι ερευνητές να υποθέτουν ότι αυτό είναι αρκετό για την επεξήγηση των συστάσεων.

Προτάσεις για την προέκταση της παρούσας διπλωματικής εργασίας

Στην Διπλωματική εργασία αυτή επιχειρήθηκε να γίνει η συγκέντρωση και η ταξινόμηση των ερευνητικών εργασιών, που έχουν στόχο την μελέτη της εφαρμογής MAB αλγορίθμων σε Συστήματα Σύστασης. Η ευρύτητα του πεδίου όσον αφορά τους MAB αλγορίθμους αλλά και όσον αφορά τα ΣΣ δυστυχώς δεν επέτρεψε και μια επιπρόσθετη πειραματική προσέγγιση εντός της, καθώς το μέγεθος και ο χρόνος που θα απαιτούνταν ξεπερνούσε τα φυσιολογικά όρια μιας Διπλωματικής Εργασίας. Παρόλα αυτά, η ευρύτητα αυτή δημιουργεί μια πολύ μεγάλη γκάμα επιλογών που θα μπορούσαν να προεκτείνουν την παρούσα εργασία στα πλαίσια πειραμάτων και περαιτέρω μελέτης. Το πειραματικό κομμάτι θεωρώ ότι στόχο θα είχε την αξιολόγηση της επίδρασης που θα είχε η ενσωμάτωση ενός MAB αλγορίθμου σε ένα σύστημα σύστασης το οποίο έχει κατασκευαστεί μέσω προσομοίωσης. Θα απαιτούνταν λοιπόν ως πρώτο βήμα η δημιουργία των εξής:

1. Ένα σύνολο χρηστών με συγκεκριμένα χαρακτηριστικά και ένα σύνολο αντικειμένων επίσης με συγκεκριμένα χαρακτηριστικά. Το μέγεθος των συνόλων αλλά και το πλήθος των χαρακτηριστικών θα καθιστά την πολυπλοκότητα που υπό εξέταση ΣΣ.
2. Ένα σύνολο αξιολογήσεων των χρηστών για κάποια αντικείμενα. Εδώ το σύνολο μπορεί να κατασκευαστεί κατάλληλα ώστε να προσομοιώνει κάθε φορά

το σενάριο που ενδέχεται να ενδιαφέρει το πείραμα. Για παράδειγμα ένας κενός πίνακας αξιολογήσεων μπορεί να προσομοιώσει το σενάριο της νέας κοινότητας.

3. Στην συνέχεια απαιτείται η υλοποίηση των μεθόδων σύστασης (Συνεργατικού Φιλτραρίσματος ή Βάσει Περιεχομένου). Η υλοποίηση αυτή θα είναι συγκεκριμένη κάθε φορά διότι η μελέτη της επίδρασης του υπό εξέταση MAB θα πραγματοποιείται με την ίδια μέθοδο πριν και μετά την εφαρμογή του.
4. Καθορισμός των μετρικών αξιολόγησης του συστήματος σύστασης. Βάσει των μετρικών αυτών θα έχουμε την τελική αξιολόγηση της επίδρασης του MAB.
5. Μετά την εκτέλεση της προσομοίωσης και την αποθήκευση των μετρικών θα υλοποιείται η εφαρμογή του υπό εξέταση MAB στο μέρος του ΣΣ που μας ενδιαφέρει. Για παράδειγμα μπορεί να χρησιμοποιηθεί ένας MAB μόνο για το σενάριο νέου χρήστη ή αντικειμένου ώστε να αξιολογηθεί η αντιμετώπιση του στο πρόβλημα Αργής Εκκίνησης (Cold-Start).
6. Τέλος θα επαναλαμβάνεται η προσομοίωση ακριβώς με το ίδιο σενάριο που πραγματοποιήθηκε την πρώτη φορά και θα συγκεντρώνονται τα αποτελέσματα των καθορισμένων μετρικών, τα οποία θα συγκρίνονται με τα αρχικά.

Το πλήθος των υπό εξέταση MAB, η μέθοδος σύστασης του προσομοιωμένου ΣΣ, το μέγεθος και η πολυπλοκότητα του συνόλου χρηστών, αντικειμένων και αξιολογήσεων καθιστούν μια αρκετά μεγάλη γκάμα πιθανών επόμενων διπλωματικών εργασιών που θα έχουν μια πιο συγκεκριμένη πειραματική στόχευση σε σχέση με την παρούσα εργασία.

Μια άλλη προέκταση θα μπορούσε να είναι όχι η αξιολόγηση κάποιου συγκεκριμένου MAB αλλά η μελέτη της ποιότητας των κριτηρίων αξιολόγησης που υπάρχουν στην βιβλιογραφία και η πειραματική επιβεβαίωση των κενών που αναφέρθηκαν και στην υποενότητα Κριτήρια Αξιολόγησης του κεφαλαίου αυτού.

Βιβλιογραφία

- [1] Alekh Agarwal, Daniel Hsu, Satyen Kale, John Langford, Lihong Li, and Robert Schapire. Taming the monster: A fast and simple algorithm for contextual bandits. *31st International Conference on Machine Learning, ICML 2014*, 5, 02 2014.
- [2] Rajeev Agrawal. The continuum-armed bandit problem. *SIAM Journal on Control and Optimization*, 33(6):1926–1951, 1995.
- [3] P. Auer. Using upper confidence bounds for online learning. pages 270–279, 2000.
- [4] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2004.
- [5] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 05 2002.
- [6] Ashwinkumar Badanidiyuru, John Langford, and Aleksandrs Slivkins. Resourceful contextual bandits. 35:1109–1134, 13–15 Jun 2014.
- [7] Walid Bendada, Guillaume Salha, and Théo Bontempelli. Carousel personalization in music streaming apps with contextual bandits. *Fourteenth ACM Conference on Recommender Systems*, sep 2020.
- [8] Alina Beygelzimer and John Langford. The offset tree for learning with partial labels. 2008.

- [9] Jesus Bobadilla, Fernando Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132, 07 2013.
- [10] Djallel Bouneffouf, Amel Bouzeghoub, and Alda Gançarski. A contextual bandit algorithm for mobile context-aware recommender system. *Neural Information Processing*, pages 324–331, 11 2012.
- [11] Djallel Bouneffouf, Romain Laroche, Tanguy Urvoy, Raphaël Feraud, and Allesiaro Robin. Contextual bandit for active learning: Active thompson sampling. 11 2014.
- [12] Djallel Bouneffouf, Srinivasan Parthasarathy, Horst Samulowitz, and Martin Wistuba. Optimal exploitation of clustering and history information in multi-armed bandit. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019.
- [13] Djallel Bouneffouf and Irina Rish. A survey on practical applications of multi-armed and contextual bandits. *IBM Thomas J. Watson Research Center, Yorktown Heights, NY USA*, 2019.
- [14] Björn Brodén, Mikael Hammar, Bengt Nilsson, and Dimitris Paraschakis. Ensemble recommendations via thompson sampling: an experimental study within e-commerce. pages 19–29, 03 2018.
- [15] A.N. Burnetas and M.N. Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17:122–142, 1996.
- [16] A.N. Burnetas and M.N. Katehakis. Optimal adaptive policies for markov decision processes. *Mathematics of Operations Research*, 22:222–255, 1997.
- [17] Apostolos N. Burnetas, Odysseas Kanavetas, and Michael N. Katehakis. Optimal data driven resource allocation under multi-armed bandit observations. 2018.

- [18] Giuseppe Burtini, Jason Loeppky, and Ramon Lawrence. Improving online marketing experiments with drifting multi-armed bandits. *ICEIS 2015 - 17th International Conference on Enterprise Information Systems, Proceedings*, 1, 04 2015.
- [19] Róbert Busa-Fekete, Balazs Szorenyi, Paul Weng, and Shie Mannor. Multi-objective bandits: Optimizing the generalized gini index. 06 2017.
- [20] Emanuele Cavenaghi, Gabriele Sottocornola, Fabio Stella, and Markus Zanker. Non stationary multi-armed bandit: Empirical evaluation of a new concept drift-aware algorithm. *Entropy*, 23, 2021.
- [21] Celma and Herrera. A new approach to evaluating novel recommendations. *Proceedings of the 2008 ACM conference on Recommender systems*, pages 179–186, 2008.
- [22] Deepayan Chakrabarti, Ravi Kumar, Filip Radlinski, and Eli Upfal. Mortal multi-armed bandits. *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*, pages 273–280, 01 2008.
- [23] Ya-Hsuan Chang and Hsuan-Tien Lin. Pairwise regression with upper confidence bound for contextual bandit with multiple actions. *2013 Conference on Technologies and Applications of Artificial Intelligence*, pages 19–24, 2013.
- [24] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. *NIPS*, 01 2011.
- [25] Niladri S. Chatterji, Vidya Muthukumar, and Peter L. Bartlett. Osom: A simultaneously optimal algorithm for multi-armed and linear contextual bandits. *AISTATS*, 2020.
- [26] W. Chen, Y. Wang, and Y. Yuan. Combinatorial multi-armed bandit: General framework, results and applications. *30th International Conference on Machine Learning, ICML 2013*, pages 151–159, 01 2013.

- [27] Konstantina Christakopoulou and Arindam Banerjee. *Learning to Interact with Users: A Collaborative-Bandit Approach*, pages 612–620. 05 2018.
- [28] Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. *Journal of Machine Learning Research - Proceedings Track*, 15:208–214, 01 2011.
- [29] David Cortes. Adapting multi-armed bandits policies to contextual bandits scenarios. 11 2018.
- [30] Koby Crammer and Claudio Gentile. Multiclass classification with bandit feedback using adaptive regularization. *Machine Learning*, 90:273–280, 01 2011.
- [31] Miroslav Dudik, Dumitru Erhan, John Langford, and Lihong Li. Doubly robust policy evaluation and optimization. *Statistical Science*, 29(4), nov 2014.
- [32] Cricia Felicio, Klérisson Paixão, Célia Barcelos, and Philippe Preux. A multi-armed bandit model selection for cold-start user recommendation. *Proceedings of the 25th conference on user modeling, adaptation and personalization*, pages 32–40, 07 2017.
- [33] Raphaël Feraud, Allesiaro Robin, Tanguy Urvoy, and Fabrice Clerot. Random forest for the contextual bandit problem. 05 2016.
- [34] Sarah Filippi, Olivier Cappé, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. pages 586–594, 01 2010.
- [35] Pratik Gajane, Tanguy Urvoy, and Fabrice Clerot. A relative exponential weighing algorithm for adversarial utility-based dueling bandits (poster). 07 2015.
- [36] Elena Gangan, Milos Kudus, and Eugene Ilyushin. Survey of multiarmed bandit algorithms applied to recommendation systems. *International Journal of Open Information Technologies*, 9(4), 2021.

- [37] Aurélien Garivier and Eric Moulines. On upper-confidence bound policies for non-stationary bandit problems. 06 2008.
- [38] M. Ge, C. Delgado-Battenfeld, and D. Jannach. Beyond accuracy: evaluating recommender systems by coverage and serendipity. *Proceedings of the fourth ACM conference on Recommender systems*, pages 257–260, 2010.
- [39] Claudio Gentile, Shuai Li, and Giovanni Zappella. Online clustering of bandits. *31st International Conference on Machine Learning, ICML 2014*, 3, 01 2014.
- [40] Michael Gimelfarb, Scott Sanner, and Chi-Guhn Lee. ϵ -bmc: A bayesian ensemble approach to epsilon-greedy exploration in model-free reinforcement learning. 07 2019.
- [41] Frédéric Guillo. *On Recommendation Systems in a Sequential Context*. Université Lille 3, 2016.
- [42] Yonatan Gur, Omar Besbes, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in Neural Information Processing Systems*, 1, 12 2014.
- [43] Nicolas Gutowski, Tassadit Amghar, Olivier Camp, and Fabien Chhel. Global versus individual accuracy in contextual multi-armed bandit. 04 2019.
- [44] Nicolas Gutowski, Tassadit Amghar, Olivier Camp, and Fabien Chhel. Gorthaur: A portfolio approach for dynamic selection of multi-armed bandit algorithms for recommendation. pages 1164–1171, 2019.
- [45] Awni Y. Hannun, Brian Knott, Shubho Sengupta, and Laurens van der Maaten. Privacy-preserving contextual bandits. *ArXiv*, abs/1910.05299, 2019.
- [46] Botao Hao, Yasin Abbasi-Yadkori, Zheng Wen, and Guang Cheng. Bootstrapping upper confidence bound. 06 2019.

- [47] Negar Hariri, Bamshad Mobasher, and Robin Burke. Context adaptation in interactive recommender systems. *RecSys 2014 - Proceedings of the 8th ACM Conference on Recommender Systems*, pages 41–48, 10 2014.
- [48] Meisam Hejazi Nia, Kyler Eastman, Shuqin Ye, Abbas Amirabadi, and Ravi Divvela. Accelerated learning from recommender systems using multi-armed bandit. 08 2019.
- [49] Daniel N. Hill, Houssam Nassif, Yi Liu, Anand Iyer, and S.V.N. Vishwanathan. An efficient bandit algorithm for realtime multivariate optimization. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, aug 2017.
- [50] Tzung-Pei Hong, Wei-Ping Song, and Chu-Tien Chiu. Evolutionary composite attribute clustering. pages 305–308, 2011.
- [51] Chu-Cheng Hsieh, James Neufeld, Tracy King, and Junghoo Cho. Efficient approximate thompson sampling for search query recommendation. pages 740–746, 04 2015.
- [52] Marcus Hutter and Jan Poland. Adaptive online prediction by following the perturbed leader. *Journal of Machine Learning Research*, 6, 05 2005.
- [53] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich. *Recommender Systems An Introduction*. Cambridge University Press, 2010.
- [54] Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett. Scalable generalized linear bandits: Online computation and hashing. 05 2017.
- [55] Sumeet Katariya, Branislav Kveton, Zheng Wen, and Vamsi Potluru. Conservative exploration using interleaving. 06 2018.
- [56] M.N. Katehakis and H. Robbins. Sequential choice from several populations. *Proceedings of the National Academy of Sciences of the United States of America*, 92(19):8584–8585, 1995.

- [57] Jaya Kawale, Hung Bui, Branislav Kveton, Long Tran-Thanh, and Sanjay Chawla. Efficient thompson sampling for online matrix-factorization recommendation. 12 2015.
- [58] Levente Kocsis and Csaba Szepesvári. Bandit based monte-carlo planning. *Machine Learning: ECML*, 2006:282–293, 09 2006.
- [59] P. Kohli, M. Salek, and G. Stoddard. A fast bandit algorithm for recommendation to users with heterogenous tastes. *Proc. of the 27th AAAI Conference on Artificial Intelligence (AAAI 2013)*, 2013.
- [60] Junpei Komiyama, Junya Honda, and Hiroshi Nakagawa. Optimal regret analysis of thompson sampling in stochastic multi-armed bandit problem with multiple plays. *ICML*, 37:1152–1161, 2015.
- [61] Junpei Komiyama, Junya Honda, and Hiroshi Nakagawa. Copeland dueling bandit problem: Regret lower bound, optimal algorithm, and computationally efficient algorithm. *ICML*, 2016.
- [62] Volodymyr Kuleshov and Doina Precup. Algorithms for the multi-armed bandit problem. *Journal of Machine Learning Research* 1, pages 1–48, 2000.
- [63] Branislav Kveton, Manzil Zaheer, Csaba Szepesvári, Lihong Li, Mohammad Ghavamzadeh, and Craig Boutilier. Randomized exploration in generalized linear bandits. 06 2019.
- [64] Anisio Lacerda. Multi-objective ranked bandits for recommender systems. *Neurocomputing*, 246, 02 2017.
- [65] Anisio Lacerda, Adriano Veloso, and Nivio Ziviani. Adding value to daily-deals recommendation: Multi-armed bandits to match customers and deals. pages 216–221, 2015.
- [66] T.L. Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

- [67] John Langford and Tong Zhang. The epoch-greedy algorithm for contextual multi-armed bandits. *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*, 01 2007.
- [68] Lihong Li, Wei Chu, John Langford, and Robert Schapire. A contextual-bandit approach to personalized news article recommendation. *Computing Research Repository - CORR*, 02 2010.
- [69] Lihong Li, Wei Chu, John Langford, and Xuanhui Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011*, pages 297–306, 01 2011.
- [70] Lihong Li, Yu Lu, and Dengyong Zhou. Provable optimal algorithms for generalized linear contextual bandits. 02 2017.
- [71] Shuai Li, Alexandros Karatzoglou, and Claudio Gentile. Collaborative filtering bandits. pages 539–548, 07 2016.
- [72] Shuai Li, Baoxiang Wang, Shengyu Zhang, and Wei Chen. Contextual combinatorial cascading bandits. 06 2016.
- [73] Chang Liu, Qingpeng Cai, and Yukui Zhang. Multi-armed bandit mechanism with private histories. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, page 1607–1609, 2017.
- [74] Manuel Lopes, Benjamin Clement, Didier Roy, and Pierre-Yves Oudeyer. Multi-armed bandits for intelligent tutoring systems. *CoRR*, abs/1310.3174, 2013.
- [75] Jonathan Louedec, Max Chevalier, Josiane Mothe, Aurélien Garivier, and Sébastien Gerchinovitz. A multiple-play bandit algorithm applied to recommender systems. 05 2015.
- [76] Xiuyuan Lu, Zheng Wen, and Branislav Kveton. Efficient online recommendation via low-rank ensemble sampling. pages 460–464, 2018.

- [77] Mohammad Malekzadeh, Dimitrios Athanasakis, Hamed Haddadi, and Ben Livshits. Privacy-preserving bandits. 09 2019.
- [78] Yizhi Mao, Miao Chen, Abhinav Wagle, Junwei Pan, Michael Natkovich, and Don Matheson. A batched multi-armed bandit approach to news headline testing. *2018 IEEE International Conference on Big Data (Big Data)*, pages 1966–1973, 2018.
- [79] Benedict May, Nathan Korda, Anthony Lee, and David Leslie. Optimistic bayesian sampling in contextual-bandit problems. *The Journal of Machine Learning Research*, 13:2069–2106, 03 2012.
- [80] James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. Explore, exploit, and explain: Personalizing explainable recommendations with bandits. page 31–39, 2018.
- [81] Prem Melville and Vikas Sindhwani. *Recommender Systems*. IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, 2017.
- [82] T. Murakami, K. Mori, , and R. Orihara. Metrics for evaluating the serendipity of recommendation lists. *Annual Conference of the Japanese Society for Artificial Intelligence*, pages 40–46, 2007.
- [83] Minh N. H. Nguyen, Chuan Pham, Jaehyeok Son, and Choong Seon Hong. Online learning-based clustering approach for news recommendation systems. pages 1–4, 2016.
- [84] Trong T. Nguyen and Hady W. Lauw. *Dynamic Clustering of Contextual Multi-Armed Bandits*. CIKM '14. Association for Computing Machinery, New York, NY, USA, 2014.
- [85] Olivier Nicol, Jérémie Mary, and Philippe Preux. Improving offline evaluation of contextual bandit algorithms via bootstrapping techniques. 2014.

- [86] Santiago Ontanon. Combinatorial multi-armed bandits for real-time strategy games. *Journal of Artificial Intelligence Research*, 58:665–702, 03 2017.
- [87] Sandeep Pandey, Deepayan Chakrabarti, and Deepak Agarwal. Multi-armed bandit problems with dependent arms. pages 721–728, 01 2007.
- [88] Vianney Perchet and Philippe Rigollet. The multi-armed bandit problem with covariates. *The Annals of Statistics*, 41, 10 2011.
- [89] S. Pilarski, S. Pilarski, and D. Varró. Optimal policy for bernoulli bandits: Computation and algorithm gauge. *IEEE Transactions on Artificial Intelligence*, 2(1):2–17, 2021.
- [90] S. Pilarski, S. Pilarski, and D. Varró. Delayed reward bernoulli bandits: Optimal policy and predictive meta-algorithm. *IEEE Transactions on Artificial Intelligence*, 3(2):152–163, 2022.
- [91] R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45, 2006.
- [92] William Press. Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research. *Proceedings of the National Academy of Sciences of the United States of America*, 106:22387–92, 12 2009.
- [93] Runtao Qiao, Shuhan Yan, and Beijun Shen. A reinforcement learning solution to cold-start problem in software crowdsourcing recommendations. *2018 IEEE International Conference on Progress in Informatics and Computing (PIC)*, pages 8–14, 2018.
- [94] Lijing Qin, Shouyuan Chen, and Xiaoyan Zhu. *Contextual Combinatorial Bandit and its Application on Diversified Online Recommendation*, pages 461–469. 04 2014.

- [95] F. Radlinski, R. Kleinberg, and T. Joachims. Learning diverse rankings with multi-armed bandits. *Proceedings of the 25th international conference on Machine learning*, pages 784–791, 2008.
- [96] Wenbo Ren, Xingyu Zhou, Jia Liu, and Ness Shroff. Multi-armed bandits with local differential privacy. 07 2020.
- [97] Francesco Ricci, Lior Rokach, and Bracha Shapira. *Recommender Systems Handbook*. Springer, 2010.
- [98] Philippe Rigollet and Assaf Zeevi. Nonparametric bandits with covariates. 03 2010.
- [99] H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58:527–535, 1952.
- [100] Allesiardo Robin, Raphaël Feraud, and Djallel Bouneffouf. A neural networks committee for the contextual bandit problem. 11 2014.
- [101] Yuta Saito, Shunsuke Aihara, Megumi Matsutani, and Yusuke Narita. A large-scale open dataset for bandit algorithms. 08 2020.
- [102] Steven Scott. A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26:639 – 658, 11 2010.
- [103] Yevgeny Seldin, Csaba Szepesvari, Peter Auer, and Yasin Abbasi-Yadkori. Evaluation and analysis of the performance of the exp3 algorithm in stochastic environments. 2012.
- [104] Nícollas Silva, Heitor Werneck, Thiago Silva, Adriano C.M. Pereira, and Leonardo Rocha. Multi-armed bandits in recommendation systems: A survey of the state-of-the-art and future directions. *Expert Systems with Applications*, 197, 2022.
- [105] Thiago Silva, Nícollas Silva, Heitor Werneck, Adriano C. M. Pereira, and Leonardo Rocha. *The Impact of First Recommendations Based on Exploration*

or *Exploitation Approaches in Recommender Systems' Learning*. Association for Computing Machinery, New York, NY, USA, 2020.

- [106] Aleksandrs Slivkins. Contextual bandits with similarity information. *Computing Research Repository - CORR*, 15, 07 2009.
- [107] Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. 07 2015.
- [108] R.S. Sutton and A.G. Barto. Reinforcement learning: An introduction. *IEEE Transactions on Neural Networks*, 9(5):1054–1054, 1998.
- [109] Liang Tang, Yexi Jiang, Lei Li, and Tao Li. Ensemble contextual bandits for personalized recommendation. *RecSys 2014 - Proceedings of the 8th ACM Conference on Recommender Systems*, 06 2014.
- [110] Liang Tang, Yexi Jiang, Lei Li, Chunqiu Zeng, and Tao Li. Personalized recommendation via parameter-free contextual bandits. 04 2015.
- [111] Xin Technology. *Challenges in recommender systems : scalability, privacy, and structured recommendations*. PhD thesis, 01 2015.
- [112] Cem Tekin and Eralp Turgay. Multi-objective contextual bandits with a dominant objective. *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2017.
- [113] Cem Tekin, Simpson Zhang, and Mihaela van der Schaar. Distributed online learning in social recommender systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):638–652, 2014.
- [114] Choon Hui Teo, Houssam Nassif, Daniel Hill, Sriram Srinivasan, Mitchell Goodman, Vijai Mohan, and S.V.N. Vishwanathan. Adaptive, personalized diversity for visual discovery. *Proceedings of the 10th ACM Conference on Recommender Systems*, sep 2016.
- [115] Michel Tokic. Adaptive ϵ -greedy exploration in reinforcement learning based on value differences. pages 203–210, 09 2010.

- [116] Eralp Turgay, Doruk Öner, and Cem Tekin. Multi-objective contextual bandit problem with similarity information. 03 2018.
- [117] Taishi Uchiya, Atsuyoshi Nakamura, and Mineichi Kudo. Algorithms for adversarial bandit problems with multiple plays. 6331:375–389, 10 2010.
- [118] Michal Valko, Nathaniel Korda, Remi Munos, Ilias Flaounas, and Nelo Cristianini. Finite-time analysis of kernelised contextual bandits. *Uncertainty in Artificial Intelligence - Proceedings of the 29th Conference, UAI 2013*, 09 2013.
- [119] Vargas and Castells. Rank and relevance in novelty and diversity metrics for recommender systems. *Proceedings of the fifth ACM conference on Recommender systems*, pages 109–116, 2011.
- [120] Joannès Vermorel and Mehryar Mohri. Multi-armed bandit algorithms and empirical evaluation. 10 2005.
- [121] Aleksandr Vorobev, Damien Lefortier, Gleb Gusev, and Pavel Serdyukov. Gathering additional feedback on search results by multi-armed bandits with respect to production ranking. pages 1177–1187, 05 2015.
- [122] Huazheng Wang, Qingyun Wu, and Hongning Wang. Factorization bandits for interactive recommendation. 01 2017.
- [123] Lu Wang, Chengyu Wang, Keqiang Wang, and Xiaofeng He. Biucb: A contextual bandit algorithm for cold-start and diversified recommendation. *2017 IEEE International Conference on Big Knowledge (ICBK)*, pages 248–253, 2017.
- [124] Qing Wang, Chunqiu Zeng, Wubai Zhou, Tao Li, S. S. Iyengar, Larisa Shwartz, and Genady Ya. Grabarnik. Online interactive collaborative filtering using multi-armed bandit with dependent arms. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1569–1580, 2019.

- [125] Xiang Wang, Xiangnan He, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. *TEM: Tree-enhanced Embedding Model for Explainable Recommendation*. ACM, 2018.
- [126] Xinxi Wang, Yi Wang, David Hsu, and Ye Wang. Exploration in interactive personalized music recommendation: A reinforcement learning approach. 2013.
- [127] John Weymark. Generalized gini indices of equality of opportunity. *Journal of Economic Inequality*, 1:5–24, 02 2003.
- [128] P. Whittle. Arm-acquiring bandits. *The Annals of Probability*, 6(2):284–292, 1981.
- [129] P. Whittle. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, 25:287–298, 1988.
- [130] Huasen Wu and Xin Liu. Double thompson sampling for dueling bandits. *Annual Conference on Neural Information System Processing (NIPS)*, 12 2016.
- [131] Huasen Wu, R Srikant, Xin Liu, and Chong Jiang. Algorithms with logarithmic or sublinear regret for constrained contextual bandits. 12 2015.
- [132] Qitian Wu, Hengrui Zhang, Xiaofeng Gao, Peng He, Paul Weng, Han Gao, and Guihai Chen. Dual graph attention networks for deep latent representation of multifaceted social effects in recommender systems. 03 2019.
- [133] Xian Wu, Suleyman Cetintas, Deguang Kong, Miao Lu, Jian Yang, and Nitesh Chawla. Learning from cross-modal behavior dynamics with graph-regularized neural contextual bandit. pages 995–1005, 04 2020.
- [134] Saba Yahyaa, Madalina Drugan, and Bernard Manderick. Thompson sampling in the adaptive linear scalarized multi objective multi armed bandit. 2, 01 2015.

- [135] Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences - JCSS*, 78, 01 2009.
- [136] Yisong Yue and Thorsten Joachims. Beat the mean bandit. pages 241–248, 01 2011.
- [137] Chunqiu Zeng, Qing Wang, Shekoofeh Mokhtari, and Tao Li. Online context-aware recommendation with time varying multi-armed bandit. pages 2025–2034, 08 2016.
- [138] M. Zhang and N. Hurley. Avoiding monotony: improving the diversity of recommendation lists. *Proceedings of the 2008 ACM conference on Recommender systems*, pages 123–130, 2008.
- [139] Xiaoxue Zhao, Weinan Zhang, and Jun Wang. Interactive collaborative filtering. *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, page 1411–1420, 2013.
- [140] Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural contextual bandits with UCB-based exploration. 119:11492–11502, 13–18 Jul 2020.
- [141] Li Zhou and Emma Brunskill. Latent contextual bandits and their application to personalized recommendations for new users. 04 2016.
- [142] Pan Zhou, Kehao Wang, Linke Guo, Shimin Gong, and Bolong Zheng. A privacy-preserving distributed contextual federated online learning framework with big data support in social recommender systems. *IEEE Transactions on Knowledge and Data Engineering*, 33(3):824–838, 2021.
- [143] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511–4515, 2010.

- [144] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. *Proceedings of the 14th international conference on World Wide Web*, pages 22–32, 2005.
- [145] Masrour Zoghi, Zohar Karnin, Shimon Whiteson, and Maarten Rijke. Copeland dueling bandits. 05 2015.
- [146] Masrour Zoghi, Shimon Whiteson, Remi Munos, and Maarten Rijke. Relative upper confidence bound for the k-armed dueling bandit problem. *31st International Conference on Machine Learning, ICML 2014*, 2, 12 2013.