Online User Purchasing Behavior Modeling

by

Xiaohui Gong

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Digital Signals and Image Processing

Department of Electrical and Computer Engineering

University of Alberta

Abstract

With the proliferation of e-commerce business, the study of online user purchasing behavior plays an important role in improving purchasing experiences of users as well as providing valuable intelligence to sellers. While most previous research efforts focused on explicit user behavior modeling, implicit user behavior modeling provides a greater amount of information and is more feasible and reliable for online purchasing scenario. In addition, the recommendation based on similarities in previous research results in biased, delayed or incorrect recommendation due to the absence of explicit multi-attribute modeling. Although some works have used multi-criteria decision making to solve this problem, the cardinal functions used have the attributes independence restriction that causes convex hull problem in online purchasing scenario. This thesis proposes a probabilistic multi-criteria item ranking framework that predicts the probability of an item being a user's best choice and ranks items accordingly. It uses indifference curve in microeconomics to ordinarily model implicit user behavior by using users' purchasing history directly. The newly designed ordinal model offers a flexible way to model any kind of user behavior with explicit multi-attribute modeling and without information bias/loss or convex hull problem. The model also considers inter-item competition globally. In addition, different from all prior works in which users are assumed to be able to compare all items simultaneously, the proposed prediction framework considers the fact that a user can only compare a few items at the same time, and models the user's decision process as a two-step selection process, where the user first selects a few candidates, and then makes detailed comparison. Furthermore, according to the comprehensive simulation and real user test results, the proposed algorithm significantly outperforms existing multi-criteria ranking algorithms by achieving higher ranking accuracy with short learning curve. Besides making recommendations, the proposed framework in this thesis can further benefit online sellers to improve their marketing strategies.

Acknowledgments

Firstly, I would like to acknowledge my deepest thanks to my supervisor Dr. Vicky H. Zhao, who has provided a great deal of guidance throughout the entire project from the direction of the research to the details of conducting a real user test.

Then I would like to express my most grateful thanks to my co-supervisor Dr. Hai Jiang for all the support, patience and expert guidance. Dr. Jiang's support and help were enormous and pivotal that this thesis could not be finalized and be decent without his excellent advises and detailed suggestions.

Furthermore, I would like to give my special thanks to my family. They always provide support and trust to me. Also their advises and love made it possible for me to overcome lots of difficulties.

Finally, I would also like to thank all my lab colleges and friends, Yuan Yuan, Lu Qian, Hongming Xu, Huiquan Wang, Yue Li, Gencheng Guo, and Tao Xu for all their help in my research, especially the real user tests.

Table of Contents

List of Tables

List of Figures

List of Symbols

P	the real price of the item
R	the real reputation of the item
p	the normalized price
r	the normalized reputation
P_{MAX}	predefined maximum prize for normalization
P_{MIN}	predefined minimum prize for normalization
R_{MAX}	predefined maximum reputation for normalization
R_{MIN}	predefined minimum reputation for normalization
S	the skyline item set
s_i	the skyline item i
p_i	the normalized price of skyline item i
r_i	the normalized reputation of skyline item i
s_b	the best choice among skyline items S
$U(s_i)$	the utility of skyline item s_i
$U(p_i, r_i)$	the utility of skyline item s_i
$h = \{s_b, S\}$	the historical record for one complete transaction
rq	ranking quality
sr	success rate
k_s	personalized MRS of item s

List of Abbreviations

Acronyms	Definition
CB	Content Based
CF	Collaborative Filtering
EMR	Estimated MRS range
IC	Indifference Curve
IM	Indifference Map
IS	Interested Set
MCDM	Muilti-Criteria Decision Making
MCRS	Muilti-Criteria Recommender Systems
MRS	Marginal Rate of Substitution
NBC	non-best-choice items
RS	Recommender Systems

Chapter 1

Introduction

1.1 Online User Behavior Learning

The advance of Internet technologies and the easy access to rich information resources online have shifted people's lifestyles in the past decade by integrating social networks and online communities to their daily lives. Successful online applications provide users with massive information and choices within a few clicks, but overwhelm users more easily, refer to as the information overload problem.

The prevalent solution for this problem is to learn and predict personalized behavior. Examples include web searching [2], database queries [3], and recommender systems [4], [5], etc. Among them, the recommender system (RS) became an important research area with the growth of online activities. RS provides meaningful suggestions to a collection of users for items that might interest them [6]. Since the start of RS in mid-1990's, RS has been focusing on problems that rely on explicit ratings [4]. Explicit rating is the main component of the explicit user behavior, which is the result of translation from intuitive users' feelings to scaled numbers or other explicit profiles that the RS can utilize. The translations are always done by users themselves to represent the personalized quality of interaction between users and items [6]. As the client of explicit user behaviors, traditional RS utilize them to estimate ratings for future items that the user has never encountered. Once the estimated ratings are acquired, we can recommend to the user the items with the highest ratings. The growing e-commerce and e-marketing opportunities trigger the integration of explicit user behavior learning in different RS systems. Two classical RS examples are Netflix's movie recommendation and Amazon's book recommendation. In the two example, the RS first reads items rated by users, and then recommends new items to users.

1.2 Implicit Online User Purchasing Scenario

In this work, we mimic online user purchasing procedure in the real e-commercial platform and change the algorithm in the background to process data. We will use the following example to make explanations and discussions in the rest of the thesis. Consider an online user who queries for a 'Cuisinart Coffee Maker' and makes decisions among the list of available items. He/she may or may not have a history of purchasing coffer makers before this query. The proposed personalized ranking algorithm will in the background learn the user's historical purchasing behavior of coffee maker[1] and predict the possibility of each listed coffee maker item being the best choice and rank them accordingly. Then it shows the query results ranked high to low, so that the user can delightfully see the most wanted choices at the first sight. In other words, the query results will be a personalized ranking based on personalized preferences rather than on simple relevance or prices that the current websites are using. The personalized preferences refer to user preferences predicted by the system, which are learned from users' historical purchasing behavior. Then, given query results, the user will make comparison based on their interested attributes, such as products' prices, ratings, etc. The final purchasing action from this query will be recorded and join the previous training data, which is used for the improvement of next purchasing prediction upon the next query. The historical user purchasing behavior is defined as the item the user purchased among the given query list after comparison, which is based on a given set of attributes (e.g., price, reputation). Therefore, the given information that could be learned only includes three parts, the set of candidate items, the set of item

[1] If the user did not purchase coffer makers before, then the proposed personalized ranking algorithm will learn from a *default user*.

attributes, and the purchased item.

The difference between our scenario and the ones in most previous researches is that we use implicit behavior as the input. The online user purchasing behavior is considered as implicit behavior since the user does not explicitly transform their feelings which tell the system the satisfaction level of each item (e.g., the rating of a watched movie from Netflix or the rating of a sold item from Amazon or eBay) they have compared. Instead, the users' actions are directly recorded by the system. For example, the system in our scenario records whether an item in the queried list has been purchased or not. Other implicit user actions include browsing histories, wishing lists, shopping carts, etc.

The most obvious advantage of using personalized ranking is that users can see the most suitable items/products/services within the top recommended choices. In this way, there is no need to click the second page or spend more time on choices out of the top ones. After several experiences, users will begin to trust this system and make decisions more efficiently with less efforts, which solves the targeted information overload problem. Further advantage of this system is to make use the predictions to serve businesses for better marketing strategies, such as to properly adjust prices of products or design services/promotions that attract consumers and make more profit.

Note that the proposed scenario above does not consider any side information[2]. Instead, it focuses on the abstract model that analyzes historical user purchasing behavior and models personalized preference that cannot be acquired directly. Although online stores commonly have additional tools to interactively get more user information to improve RS performance, these are not the focus of this thesis. Similarly, the hybrid solution such as combining previous RS solutions to improve the proposed model is also out of the consideration of this thesis.

[2]The side information is part of the user profile that can not be acquired from users' interaction with items. Instead, it can be acquired from the users interactions with the system, such as providing personalized information directly to the system, selecting various filters on websites to narrow down the selections, answering a small questionnaire during registration, etc. For example, many travel agent try to feed consumers with suitable hotels and restaurants by considering side information, such as users' locations, consumption levels, tastes, preferences, etc.

1.3 Motivation

While most prior recommendation solutions focused on studying explicit user behavior, such as ratings, reviews, votes, etc., the proposed implicit online user purchasing scenario only considers implicit user behavior, which is under-investigated from the state-of-art. Compared to explicit user behavior modeling, implicit user behavior modeling has the following three advantages that should be emphasized.

First, implicit user behavior provides much more information. Explicit user behavior requires users to explicitly transform and publish their feelings into absolute scores or opinions, but most users are reluctant to do so for time or privacy reason. The RS is also restricted by the available rating/reviewing aspects, which actually limits the available information. RS using implicit user behavior could directly tracks all kinds of user actions, such as browsing history, wish list, shopping cart, bidding history, purchasing records, etc., without disrupting users. The enlarged information source could relieve the sparsity problem which hampered the traditional RS [7]–[9].

Second, recommendation based on implicit user behavior is more feasible. Except for big online vendors, many online platforms do not have rating systems, which means they cannot implement traditional recommender systems. However, implicit user actions could instead provide enough information for all of them to make predictions and to understand their customers better.

Third, the training data obtained from implicit behavior is more reliable. There are two main reasons. Fist and the most obvious reason is due to the users' reluctancy of rating or reviewing things. This means that the collected ratings and reviews are very likely to be extreme evaluations, since extreme experiences more easily trigger users to populate their feelings. For example, it is easy to see the reviews of walk-in clinic on Google Maps always have a very low rating, such as 1.9, 2.1, etc., as most of the reviewers were the ones with bad experiences. Second, prior researches have shown that users are more consistent when making real comparisons (e.g., compare two products and pick up the better one) than giving absolute scores [10], [11]. In contrast, transformation from feelings to absolute scores introduces information inconsistency and distortion. These are due to 1) users' limited ability to recall feelings precisely and 2) the lack of an agreement on the meaning of score systems. For example, most studies assume that the five-star system uses same scales between adjacent scores, which may not be able to express users' preferences well in practice. For example, the interpretation of 3 stars may be a moderate rating for some users that is only a slightly difference from 4, but for other users, it may be a low quality rating closer to $1 - 2$ stars [12].

In summary, the study of implicit online user behavior has promising advantages than the study of explicit user behavior and is the focus of this thesis.

1.4 Outline

The rest of the thesis is organized as follows. Chapter 2 goes through the previous literature, demonstrates limitations of previous studies, and gives a summary of the proposed study and the contributions. Chapter 3 presents the proposed framework and introduces relative economics concepts used in this thesis. Chapter 4 describes in detail the proposed algorithms for user preference estimation and personalized ranking. Chapter 5 gives out simulation results and discussions. Chapter 6 shows real user test, the results and discussions. Chapter 7 concludes the proposed work and discusses the future work.

Chapter 2

Literature Review

2.1 Recommender Systems

While many efforts in the literature have been made to tackle the problem of overloading information, recommender system (RS) is always the most active research area that uses users' past behaviors to learn their preferences and make predictions. The goal of RS is to assist users with their decision making on multiple alternatives (items). It realizes the assistance by making prediction/recommendation on items that have not been observed, e.g., viewed, used, purchased, etc. The quality of the predictions/recommendations is measured by how relative it is with users' actual need/interests [4], [13], [14]. Therefore, the definition of RS can be expressed as follows [4]. Denote two sets, $Users$ and $Items$, which contain all the users using the system $u \in Users$ and all the available items $i \in Items$, respectively. A utility function is defined as $R : Users \times Items \rightarrow R_0$. In the recommender systems, the utility function expressing user's personalized preferences is assumed to exist but is not known by the system nor by users themselves. The system is supposed to retrieve personalized preferences by estimating the utility function $R(u, i)$ and to recommend top-N items with the highest $R(u, i)$.

2.1.1 Non-personalized Recommenders and Personalized Recommenders

RS is often classified into non-personalized recommenders and personalized recommenders. Non-personalized recommenders often aggregate ratings/votes among all users, e.g., Zagat Guide for New York city restaurants averages ratings of criteria such as food, decor, service and cost; the Conde Nast Traveller provides the percentage of people who rate a particular hotel, cruise, etc., as "very good" or "excellent"; YouTube presents the total number of upvotes and downvotes, etc. However, non-personalized recommenders suggest popular items that may not be useful. On one hand, popular items may not match personalized needs. On the other hand, they may only provide trivial information. For example, recommending popular products such as banana for supermarket customers is very accurate, but it is trivial since banana will be bought by most costumers without any recommendation assistance. Personalized recommenders use a variety of techniques to learn personalized user preferences in different scenarios and make proper recommendations accordingly. This is the focus of our work. Personalized recommenders can be classified into content-based (CB), collaborative filtering (CF), and hybrid approaches [15].

2.1.2 Content-Based (CB) Recommenders

The CB approach recommends similar items to the liked one experienced in the past. It assumes relative stable user preferences. Thus it is suitable in areas such as news feeding (e.g., the user prefers stories on technology, music and political issues), clothing (e.g., the user prefers cotton, warm colors, and casual), hotels (e.g., the user prefers breakfast, wifi, and swim pool), movies (e.g. the user prefers Tom Hanks, Comedy and Romantics), etc. In these areas, user preference will not change a lot in a short time. To find the utility function $R(u, i)$, the CB approach builds user and item profiles in vectors. Similarities between user and item vectors are then calculated and the most similar items are recommended to the user. One of the famous algorithms is the term frequency and inverse document frequency (TFIDF), which is a numerical statistic indicating the importance of terms/attributes in a document. The term frequency is the number of occurrences of a term in the document, and the inverse document frequency is the inverse of the occurrences of documents containing the testing term among all the documents. By multiplying the two frequencies, the TFIDF represents the importance of certain terms, automatically demotes stopwords and common terms, and promotes core terms over incidental ones.

The advantage of the CB approach is the easy computation, the understandable profiles and the flexibility to integrate with query-based and case-based systems. However, compared to other existing approaches, it has difficulties in handling interdependencies, e.g., the user interests in computer games and sport television do not indicate the preference in sport games. Also, it is lack of diversity and serendipity. When considering online purchasing scenario in this thesis, CB can only recommend similar items the user has experienced but cannot deal with trade-offs of conflicting attributes. In particular, when the user profile contains conflicting criteria that need to be trade-off, recommendation based on similarity may end up proposing the inferior item that is more similar to past liked items rather than the more competitive one. Consider the case that a user u purchased product A at $\$100$ with 4.2 rating stars in the past. When u makes a new query, the available products are product B, which is $\$105$ with 4 stars, and product C, which is $\$75$ with 4.9 stars. The CB approach will not be suitable here since it will recommend B rather than C, since B is more similar to A.

2.1.3 Collaborative Filtering (CF) Recommenders

The CF approach can be mostly divided into two categories, memory-based CF and model-based CF. Both approaches assume that users'(items') past agreement could predict their future agreement. Memory-based CF usually does not have training phases [16]. Instead, it measures user behaviors similarities directly and finds neighbors for the active user online. The well used example is Amazon's "Customers Who Bought This Item Also Bought" approach. Concretely, the predicted rating of item i for an active user u_a can be the normalized weighted sum of neighbors' past ratings:

$$p_{u_{ai}} = \frac{\sum_{u=1}^{n} r_{ui} \times w_{u_a u}}{\sum_{u=1}^{n} w_{u_a u}}, \tag{2.1}$$

where u is a neighbor of u_a, and $w_{u_a u}$ measures the similarity between u_a and u. To consider rating deviations among different users, the prediction could be

$$p_{u_{ai}} = \bar{r}_{u_a} + \frac{\sum_{u=1}^{n}(r_{ui} - \bar{r}_u) \times w_{u_a u}}{\sum_{u=1}^{n} w_{u_a u}}, \tag{2.2}$$

where $\bar{r}_u$ is the average rating of user u.

Similarity measurements for calculating the weight $w_{u_a u}$ can be cosine similarity, Pearson correlation, and Spearman's rank correlation [17]. Vector similarity is the cosine similarity of two user vectors. Pearson correlation is

$$w_{u_a u} = \frac{\sum_{i=1}^{m}(r_{u_a i} - \bar{r}_{u_a})(r_{ui} - \bar{r}_u)}{\sigma_{u_0}\sigma_u}, \tag{2.3}$$

where $\bar{r}_{u_a}$ and $\bar{r}_u$ are average ratings of active user u_a and user u, respectively, and σ_{u_0} and σ_u are standard deviation of ratings of active user u_a and user u, respectively. Pearson correlation indicates the extent that two users vary together from their averaged ratings. Spearman rank correlation is to apply Pearson correlation with ranks rather than ratings.

In contrast to memory-based CF which executes most computation online, model-based CF builds model first through some training phases, and executes most of the computation offline. It makes the prediction for the active user using the builded/estimated models [16]. There are mainly three classes of the model-based CF. The first is item-item CF. Rather than calculating the similarity between users, item-item CF calculates similarities between items through the user-item matrix of past ratings. It is different from CB in that it gathers information from users' past experiences, e.g., ratings, votes, etc., rather than item profiles. This approach is superior to user-user CF when dealing with sparse data in the per user vector. In other words, item vector contains lots of user information with the large number of users, while users vector often contains few item information since most users only experienced/observed few items. In addition, since item profiles are always more stable than user preferences, it can be pre-computed offline before query to reduce the expense of online computation on large user-item matrices. For example, the top-N recommendation using item-item CF would be an instant feedback since the system already has the similarities in the database. The disadvantage of this approach is the lower diversity and serendipity compared to user-user CF. The second model-based CF approach is the probabilistic modeling approach. It calculates the probability that a rating is a particular value. By training the data, it clusters items and/or users into classes and predicts ratings for the active user by using the ratings in classes that fit in best with the active user and/or items to be rated [16]. One example is the flexible mixture model (FMM) [16], which uses two latent variables Z_u and Z_i to determine a single rating r of user u on item i. The third model-based CF approach is to reduce the dimension of user-item matrix using singular value decomposition (SVD). It decomposes the user-item matrix into three matrices as $X = USV^T$. A reduced diagonal matrix $\bar{S}$ with rank k and corresponding $\bar{U}$ and $\bar{V}$ are then used to make the best rank-k approximation $\bar{X} = \bar{U}\bar{S}\bar{V}^T$. This approach largely reduces the computational complexity. It also improves the accuracy by denoising and by avoiding overfitting.

The advantage of the CF approach is that it can recommend alternatives with higher serendipity, diversity, coverage, novelty, etc[18]. However, basic CF cannot deal with multiple attributes and there is no explicit model that depicts the relationship between users' choices and items attributes [18]. This results in delay and bias of recommendation when dealing with changing user/item profiles, e.g., an item may enter or exit the market or change its prices, reputations, etc.; a user can change his/her status such as changing the residing location or changing from working environment to entertaining environment; a user can change his/her preferences – they may prefer iPhone 7 once it issued instead of any old version even if all of his/her preference neighbors bought the old versions in the past. Without knowing the reason of user behavior, the CF approach can only refer to neighbor's previous behaviors and the recommendation will be biased until enough history occurs after the user profile has been changed. In contrast, a model that explicitly takes multi-attributes into account could make changes immediately. For example, if a user cares more about distance of restaurants, the recommendation after the user location change could be changed accordingly rather than being biased by his/her old neighbors' experiences.

2.1.4 Hybrid Recommenders

The hybrid approach of recommendation combines CB, CF, and other approaches in different ways [19], [20] to solve particular issues in different scenarios. Possible approaches include combining algorithm scores using weighted sum aggregation, switching algorithms in various conditions, using one algorithm as an input to another algorithm, etc. One typical example is the winning Netflix Prize algorithm which is a linear combination of more than 100 algorithms. For experienced users with no user-entered user profile, researchers have used CF to identify attributes and objectives that the customer may care about as the first step, and then used the obtained information in other approaches[18]. In addition, for inexperienced user with user-entered profiles, one can use these explicit user profiles as the constraints before CF. Other examples of hybrid approaches are as follows. The proliferation of social networks provides new methods that could be integrated in traditional recommender systems to improve its reliability and accuracy, and accelerate the development of social recommender systems[21]–[23]. To consider trust information, researchers incorporate user trustworthiness with relatedness computation, such as to weight users by trust information from social networks before computing their similarities [24]. The trust-aware recommender systems (TARS)[25] exploit information from explicitly trusted neighbors in social networks, It provides independent information and alleviates the cold-start problem [25]–[28]. Recently, researchers use online product reviews to enhance recommendations and put efforts on sentiment analysis[29]–[31]. Similar approaches are well investigated in various forms of trust metrics.

Despite the various methods stated above, there are two issues that inherently exist in RS which is not suitable for implicit user behavior modeling. First, the traditional RS only fills in missing data in the user-user/item-item matrix, but keeps records unchanged. This will raise issues in implicit purchasing behavior scenario, where same items may appear in multiple dynamically changing markets. Concretely, in the traditional recommendation scenario, once the rating is recorded, it will be treated as known data and will not be estimated anymore. For example, the personalized movie ratings in the database will not change and will be utilized as training/validation data for predicting ratings of movies that the user has not seen. In contrast, in the implicit user purchasing scenario, an item that is not purchased in the current query list does not mean that it will never be purchased in the future and vice versa. All the items in the current queried list could be in the next queried list and should all be re-estimated.

Second, it is not proper to follow RS and use numerical values for the purchasing records. Intuitively, assigning 1 for the purchased item and 0 for unpurchased ones seems correct. However, this will lose information from the purchasing action. When a purchasing action occurs, in the given query list with N items, one item is preferred than all other $N-1$ items. This means the active user implicitly assigned $N-1$ relative preference relationships by comparison. However, the unpurchased items should not be the same as dislike or 0. Some candidates should be better than others and may only have a slight difference than the purchased one. In addition, once the market changes, the item purchased before may not be the best choice any more. It does not simply mean that the user changes his/her preference from like (1) to dislike (0). On the contrary, it may be because the previous best choice is masked by new items with lower prices and higher reputations. In summary, transforming purchasing records to numerical values poses more bias to the recorded user behavior.

2.2 Multi-Criteria Recommender Systems (MCRS)

As discussed above, the online purchasing scenario needs explicit multi-attribute model to elicit the relationships between multiple conflicting item attributes and user behavior outcomes. The solution to it is hybrid recommender systems with multi-criteria decision making theory (MCDM), and the new system is called multi-criteria recommender systems (MCRS).

MCRS has gained increasing interests in recent years and has been regarded as one of the important issues for the next generation RS [4]. Example applications of MCRS include eBay, who allows users to provide detailed ratings of transactions on several aspects; the Zagat Guide for restaurant recommendations, who has four criteria, food, decor, service and cost; the Rakuten, who mainly sells electronics and provides multi-criteria ratings such as overall satisfaction, value, ease to use and performance.

MCDM deals with theory and methodological issues in multiple criteria (objectives, goals, attributes) decision making environments. It generally has four steps to solve a problem [32]: define the object of decision (the alternatives to be recommended in MCRS); define a consistent family of criteria (multi-attribute item profiles, e.g., the detailed ratings of transactions in eBay); develop a global preference model; choose appropriate method from each of the previous step. In MCRS, the first two steps are often defined or constrained by particular applications. The most important part is to develop a proper global preference model that will be used for recommendation.

Algorithms of MCRS can be classified into heuristic-based approach and model-based approach in the literature [33]. Heuristic-based techniques compute utility function in real time based on observed user behaviors on a certain heuristic assumption. The heuristic-based collaborative filtering approach is the most popular in traditional RS, which assumes two users that had similar preferences before are also expected to have similar preferences in their later purchasing of items that have not been observed yet. Model-based techniques learn a predictive model, which is often an explicit preference model that represents user preference based on observed user behaviors, using statistical or machine-learning methods. The learned model could then be used to estimate users' future behavior on future items. By directly modeling the user preference with multiple criteria/attributes, the model-based approach overcomes the problems in CB and CF stated before. It deals with conflicting attributes which CB cannot, and its explicit model for relationship between multi-attributes and user behavior will reflect changes immediately, without delayed or biased outcomes (which CF suffers)

The main model-based approach is the aggregation function. It synthesizes marginal preferences ($v_1, ..., v_k$) of each criterion into an aggregation function [33], [34].

$$v_0 = f(v_1, ..., v_k). \tag{2.4}$$

First, to get each marginal preference, the problem is decomposed into k single-rating recommendation problems [33] under k criteria. Methods in traditional RS can be used in this step. Then, the aggregation function is determined using domain expertise, statistical techniques, or machine learning techniques. Concretely, the aggregation function can be the simplest additive multi-attribute model commonly used in web-based applications $v(x_1, x_2, ...x_k) = \sum_{i=1}^{k} v_i(x_i)$, where v_i is the marginal preference and is defined as a single-attribute value function over criterion/attribute x_i [18]. A variation of this function is $v(x_1, x_2, ...x_k) = \sum_{i=1}^{k} w_i v_i(x_i)$, which tunes the value functions based on the weight that indicates relative importance of each criterion, and $\sum_{i=1}^{k} w_i = 1$. The value functions are often simplified to a scale in $[0, 1]$.

However, there are conditions that are required to ensure the proper use of this additive value function. First, the trade-offs among any two criteria are not affected by common outcomes on the remaining n-2 criteria [35]. Second, if we use the weighted sum variation, an even stronger condition is required as the difference independence [36]. Specifically, the common performance levels on attributes should not affect the individual's preferences for changing in the performance levels of any other attribute. In other words, the change of one v_i should not affect other v_j. This may be violated in practical online user purchasing scenario. For example, in online purchasing, when the price of a product is already very low, the satisfaction level v_i of a 3 star reputation may be relatively fair. But if the product has a higher price, the same reputation may gain lower satisfaction level, e.g., user may raise their expectation of reputation and a 3 star item is no longer acceptable. Therefore, the value of v_i should be lower than before. The violence of this condition calls issues of multi-criteria recommender system for further investigation. For example, they may cause convex hull problem stated in [37], in which some of the items never have a chance to be the best choice of users. The case study can be found in [37] using the most popular weighted sum function in recommender systems. The work in [37] also shows in simulation that the weighted sum regression function fails dramatically with slightly more sophisticated user behavior. More generally, the reason for this issue is due to the undetermined user behavior. The predetermined cardinal utility functions used in MCRS could not capture all kinds of user preference. Even if a v_i could be defined as non-linear function, it should be learned precisely to make sure the aggregation function is linear while keeping it independent from others. With the various user behaviors and the limited information provided by users, to adjust parameters in the learning function for each user precisely is difficult and costly.

There are several other methods that extend traditional model-based CF to MCRS, such as the probabilistic modeling approach and the SVD approach[38]. However, they both base on CF that cluster similar past behaviors for future prediction. There is no explicit user behavior model that depicts the relationship between item attributes and user behaviors.

In summary, the advantage of MCRS is that it could potentially improve the prediction performance by utilizing detailed information from items–the multiple attributes[33]. In this way, the new model could distinguish two users with similar behaviors, e.g., two users rate the same movie similarly but have different opinions in story, action, and visual effects. This could not be done with traditional CF methods. The drawbacks of MCRS are as follows. The problem of dependent attributes indicates that a more flexible model is needed to depict the tied attributes. For example, the model should be able to consider the case that a user may require a high reputation if he/she purchases expensive items. In addition, a more comprehensive model should also be able to capture different user behaviors among different users, and consider the changes of user preferences with time.

2.3 Multi-Attribute Probabilistic Selection framework (MAPS)

To address the above issues, the work in [37] proposes a personalized Multi-Attribute Probabilistic Selection framework (MAPS). In [37], each attribute is considered as one dimension in a multi-dimensional space, and every item to be compared is mapped to a point in the space. They use visual angle, the angle of the line connecting an item and the origin, to represent items and exploit information from implicit user purchasing behaviors. MAPS records each of the purchased item using Gaussian distribution in visual angle and combines multiple records to form the learned density function, which is used to model users' personal preference.

The visual angle approach in MAPS is the first research effort in the literature that investigates both the inter-attribute tradeoff and the inter-item competition. Inter-item competition happens when one item that was promising in previous markets is masked by other more competitive items newly entering the market. In this case, the old item's probability of being chosen as the best choice will become much lower. MAPS also avoids using cardinal utility functions to get rid of convex hull problem [37]. However, the approach has the following drawbacks. 1)The attributes' dimension is reduced by one; 2) the approach may lead to loss of some useful information; 3) the approach may not have a complete or accurate characterization of users' preference; and 4) the approach violates the monotonicity assumption which prevents the approach from further study.

2.4 Other Approaches

With the fast growing e-commercial market, there are more research efforts recently that take purchasing records as the main consideration in their proposed RS. However, without much investigation in modeling implicit user behavior in the past RS, most of the efforts combine side information and translate past behaviors to explicit ratings. After that, existing approaches in RS are utilized to build the models and to make prediction. For example, 4-level explicit ratings 1 to 4 are used to represent click, collect, add to cart and payment, respectively in [39], indicating different preference levels. A preference degree is defined in [40] where higher preference degree comes from more clicks or longer reading durations. Although implicit user behaviors are recorded and learned in these approaches, they are not directly used to model user preferences. On the contrast, the translation of implicit user behaviors to explicit ratings is done by self-defined algorithms, which will surely introduce biases due to the bounded ability of the algorithms and the limited range of ratings. In addition, some researchers attempt to bring price sensitivity to prediction [41]. This approach uses large-scale information such as the item, user, time, category, product, and quantity to make predictions [41]. It has similar intention as the proposed work in this thesis, that is, filling the gap between RS which is good at prediction based on past behaviors and Microeconomics that pays more attention to the influence of price and demand [41]. However, the approach in [41], along with the two approaches in [39] and [40], do not consider multi-attribute trade-off in decision making. They do not exploit the relationship between the purchased item and unpurchased items either. From such relation, some valuable information can be extracted, as shown in our work.

2.5 Proposed Research

The prominence research efforts in recommender systems provide lots of possible solutions, such as CB, CF, hybrid approaches, etc. However, most of them are based on similarity measurements and are biased or delayed due to the lack of explicit user behavior model that depicts personalized balances of multiple conflicting attributes. In addition, the numerical input used in traditional RS also generates bias when users transform intuitive feelings to abstract numbers, and when user action is translated to scales with limited ranges. Furthermore, the traditional learning process only fills the missing data, while in online user purchasing scenario, all queried items should be reevaluated every time. MCRS tried to solve the problem of multi-attribute balancing, but has the dependency restriction, which causes problems such as convex hull problem. MAPS attempted to establish an explicit user behavior model that is multi-attribute friendly and used a novel ordinal way to record user behavior to avoid information bias from numerical records. It was also capable to reevaluate items every time. However, it violated the monotonicity assumption due to dimension reduction and caused information loss, and only considered inter-item competition among visual-angle neighbors instead of global items.

In our work, we propose a novel probabilistic ranking framework using the concept of *indifference curve* from microeconomics. Different from traditional RS that used cardinal utility, the newly designed framework uses ordinal utility that enables multi-attribute balancing with simple records directly from users' purchasing actions. While cardinal utility is numerically measurable, ordinal utility, on the contrast, could not express user's satisfaction in numerical terms. It could only indicate whether the item provides more or less satisfaction to the user compared to the other item. As a result, the model could well depict personalized user behavior without any bias or loss of information from user or system. It avoids the attributes' independence restriction in MCRS and offers a flexible way to model any kind of user behavior. As MAPS, it can also take each item into account whether or not they are recorded before. In contrast to MAPS, there is no information loss in the proposed framework and it considers inter-item competition globally instead of just among neighbors.

In addition, different from all prior works that assume users compare all the items simultaneously, the proposed framework addresses the fact that a user has bounded rationality and can only compare a few items at the same time [42]–[44]. It models the user's decision making as a two-step selection process, where a user first selects a few candidates and then makes detailed comparison. Furthermore, the proposed framework outputs the probability that an item is the user's best choice, which provides important guidelines on appropriate pricing schemes, estimations of the market demand, and marketing strategies.

2.6 Potential Significance and Impact of the Research

The thriving online user activities make it increasingly important to investigate the study of user behavior. In the user oriented ubiquitous network, understanding online user behavior will accelerate businesses and trigger more purchasing behaviors. Concretely, understanding the relationship between user preferences and user behaviors will impact the design of online business models as well as provide sellers with guidelines to build optimized pricing strategy. The potential contributions of this thesis include:

1. novel ordinal user purchasing behavior modeling that records implicit user behavior directly. This avoids information bias due to the transformation of user feelings to abstract numbers, and prevents information loss when translating user actions to limited range of scaled ratings.

2. the use of indifference curve (IC) and the marginal rate of substitution (MRS) in microeconomics to establish the ordinal model, which flexibly covers all kinds of user behaviors on multi-attribute balancing. It also avoids the attribute independence restriction in MCDM and the convex hull problem.

3. the new learning process that takes every items in the market into account, whether or not they are recorded before, while at the same time it uses historical records to improve prediction.

4. the consideration of inter-item competition that takes every item in the market into account;

5. new mechanism of two-step user selection processes that can better depict user behaviors by dividing the decision process into two selection steps, where users select a subset of items first and then compare them in details;

6. new algorithm that achieves high ranking accuracy with a short learning curve, which is preferable for practical online businesses in the sparse data environment;

7. probabilistic results that provide physical meanings related to sellers competitiveness in the online market, which could benefit individual sellers with their online businesses.

Chapter 3

The Proposed Framework

3.1 Problem Formulation

In this thesis, as stated before, we only use purchasing history of the user as our information resource. Although interactive assistants or social information may solve the cold start problem, we will only treat it as an assistant to help with user behavior modeling, and focus on the ranking algorithm only. In addition, we only consider the fixed-price buy-it-now market but not those requiring auctions. Also, we consider a dynamic market, where items can enter or exit the online market at any time, and where price and reputation change with time.

3.1.1 Basic Definition of Variables

We consider ranking items in an online shopping platform with two conflicting attributes, price and reputation, and our work can be extended to ranking items with more than two conflicting attributes. Consider a user query, which returns a list of matching items. For a matching item with price $P \in [P_{MIN}, P_{MAX}]$ and reputation $R \in [R_{MIN}, R_{MAX}]$, we normalize both reputation and price into the range $[0, 1]$ using simple linear mapping functions $p = (P_{MAX} - P)/(P_{MAX} - P_{MIN})$ and $r = (R - R_{MIN})/(R_{MAX} - R_{MIN})$. In this work, after checking with products on eBay, we choose $P_{MAX} = 10^3$, $P_{MIN} = 10$, $R_{MAX} = 10^6$ and $R_{MIN} = 0$ to cover most of the products, and observe similar trends for other values and other normalization functions. After normalization, for both attributes p and r, a larger value indicates a higher preference of the user. In contrast to some of the existing models (i.e., weighted sum) who use more complex normalization functions, our model uses the above two simple normalization functions and leaves all the complexity to the proposed algorithm and keeps the raw data undistorted. Since the matching items are commonly sold by different sellers, we denote an item/product in the query list as s and index it using i. We use the term an *item's utility* $U(s_i)$ to quantify the user's personal level of satisfaction with item s_i, and a larger utility means higher preference. We describe an item using the two attributes p and r, which can uniquely identify an item in this work, so that we have $s_i = \{p_i, r_i\}$. We can also denote the item's utility as $U(p_i, r_i)$.

3.1.2 Rational User Behavior Assumption

In this work, we consider rational and consistent user behavior with the following three assumptions [1]. The first is the *monotonicity assumption*, where we assume that an item's utility is higher when an attribute value is higher with the other(s) fixed. If $p_i \leq p_j$ and $r_i = r_j$, then $U(s_i) \leq U(s_j)$. For example, users always prefer the higher reputation when prices are the same, and prefer the lower price when reputations are the same. Second, we have the *diminishing value assumption*, which says users are assumed to have diminishing additional level of satisfaction with the increase of a certain attribute's value. $U(p_i, r_i + \Delta r) - U(p_i, r_i) \leq U(p_i, r_i) - U(p_i, r_i - \Delta r)$. That is, with the other attribute values fixed, as one attribute value increases, the additional level of satisfaction that the user obtains decreases [1]. To understand this concept, consider the example of two listed items with reputations of 50 and 5000, respectively, and both of them gain additional 1000 reputation. When the reputation of the first item is raised from 50 to 1050, the extra satisfaction (utility) that a user gains is larger than that when the reputation of the second item is raised from 5000 to 6000. Last, we assume that users have *bounded rationality* [42], [43] and can only compare a few (usually 3 to 5) multi-attribute items at a time [44]. Therefore, the comparison process within a transaction is assumed to be divided into two steps. A user will target some interested items first, then he/she will make the detailed comparison within these items and determine the best choice.

3.1.3 Skyline Items

Given a set of items, an item is a *skyline item* if and only if its attributes are not all worse (smaller) than those of any other items [45]. From the monotonicity assumption, a non-skyline item whose attributes are all smaller than those of a skyline item has a lower utility, and thus will never be picked by the consistent user. Therefore, in our work, we consider skyline items only.

Denote the set of N skyline items as $S = \{s_i = \{p_i, r_i\}\}$, and without loss of generality, we sort them in the ascending order of normalized reputation with $r_1 < \cdots < r_N$, then we should have $p_1 > \cdots > p_N$. We map all items into points in a two-dimensional space with the X and Y axes being the normalized price and reputation, respectively. In the following, we will use the two terms "point" and "item" interchangeably to represent the same concept.

3.1.4 Two-Step Ranking

Given N skyline items in S, a user considers the trade off between price and reputation based on his/her personal preference, and chooses his/her personal best choice s_b. Due to the bounded rationality, we assume that a user first pre-selects n interested items/candidates from the set S. He/she then makes detailed comparison within the n items and finds the best choice s_b. For a given user, the goal of the proposed multi-attribute ranking algorithm is to understand the user's personal preference between the conflicting attributes and to rank the items accordingly, such that the user's true best choice s_b is ranked as high as possible.

3.2 Indifference Curve (IC) and Marginal Rate of Substitution (MRS)

3.2.1 Indifference Curve (IC)

In this work, we use the concept of *indifference curve* (IC) from microeconomics to model users' personal preference. An indifference curve is a graph showing different combinations of factors among which a user is indifferent, and points

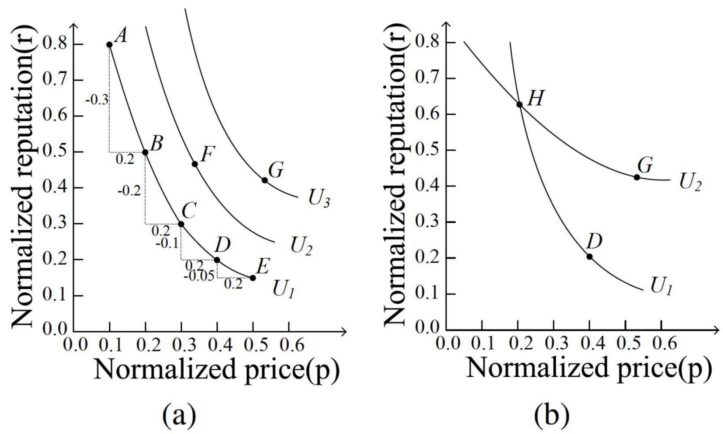


Fig. 3.1. (a) An indifference map with $U(G) > U(F) > U(A) = \cdots = U(E)$[1], and (b) indifference curves cannot intersect.

on the same indifference curve have the same utility value [1]. For example, in Fig.3.1(a), point A, B, C, D, E are on the same indifference curve with the same utility, $U(A) = U(B) = U(c) = U(D) = U(E) = U_1$. An indifference map is a collection of indifference curves with different utility values for a user, as shown in the example in Fig. 3.1(a).

Indifference curves have three properties [1] that can help with the ranking algorithm. *First*, an indifference curve is always a non-increasing function. Otherwise, an item with higher values on both attributes will have the same utility value as the item under comparison, which contradicts the monotonicity user behavior assumption. *Second*, different indifference curves do not intersect. Otherwise, all the points on the two intersected curves will have the same utility. To prove this, consider the example in Fig. 3.1(b) where two indifference curves with utility U_1 and U_2 intersect at point H. From the definition of indifference curve, all points on the two curves have the same utility as point H, including point G and D with $r_G > r_D$ and $p_G > p_D$, as shown in Fig. 3.1(b). However, it contradicts the monotonicity assumption which says that point G is preferred to D. Thus, it shows that two indifference curves cannot intersect. *Third*, when comparing two indifference curves, the top one has a higher utility and is preferred by the user. In the example of Fig. 3.1(a), we have $U_3 > U_2 > U_1$. Similarly, a point above/below an indifference curve has a higher/lower utility.

With the above properties, the problem of user purchasing behavior modeling can be changed to the estimation of the indifference curves, from which we can easily rank items according to which indifference curve the items resides on. However, in real applications, the only information of indifference curve that could be obtained from the purchasing record is that the indifference curve passing through s_b lies above all the indifference curves passing through other items in S. Since we have limited number of the user transaction (purchasing) records, it is insufficient to obtain the complete indifference map.

3.2.2 Marginal Rate of Substitution (MRS)

To address this issue, we use another concept from microeconomics, the *marginal rate of substitution* (MRS),which is the maximum amount of the attribute on the Y axis (normalized reputation r) that a user is willing to give up to obtain one additional unit of the attribute on the X axis (normalized price p)[46]. For example, in Fig. 3.1(a), MRS between points A and B is -1.5, since the marginal rate of substitution is $\Delta y/\Delta x$ where $\Delta x = 0.2$ and $\Delta y = 0.3$, which means a user is willing to give up at most $\Delta y = 0.3$ units of reputation to obtain 0.2 additional unit of price. Likewise, MRS between B and C is 1. To simplify the analysis, we assume in this work that the indifference curves are continuously differentiable. Then, MRS at a given point is the slope of the indifference curve evaluated at that point [46]. In this work, we use MRS as partial knowledge of the indifference curve to help model users' personalized preference.

MRS has an important *diminishing property* that can help us extract information of the indifference curves and users' preference from a finite number of purchasing records. From the diminishing value assumption of user purchasing behavior, as the p/r increases, the additional satisfaction that user gains with one more unit increment of p/r decreases. Consequently, with user's satisfaction fixed, as the p/r increases, the user is willing to give up less on r/p to gain an additional unit of price (reputation). Therefore, The magnitude of MRS decreases as the price increases along the curve, and the indifference curves are convex.

Indeed, with only a few purchasing records, we cannot extract perfect information of MRSs of the complete price-reputation plane. However,we can estimate the MRS ranges at most points in the purchasing records. Still, as will be demonstrated later, these estimated slope ranges can help capture users' preference with minimum information loss and offer good ranking quality.

3.3 Proposed Framework

The proposed framework is shown in Fig. 3.2, where the two shaded blocks, the preference learner and the Probability of best choice predictor, are the two key processes. The Preference Learner is responsible for the learning process that generates the MRS range profiles. As explained before, based on the two attributes p and r, the historical record corresponds to one complete transaction that contains the set of matching items from query and the best choice. For one transaction, we denote the historical record as $h = \{s_b, S\}$. The records are used to generate the learned user preference profile represented by the MRS range profiles. As shown Fig. 3.3, the preference learner contains three steps, MRS ranges estimation, MRS ranges refinement and the consistency check, to narrow down the MRS range to be closer to the real value. The Probability of Best Choice Predictor is responsible for predicting the probability of each item to be the best choice, using the estimated MRS range profiles. After learning, the MRS range profiles are used to estimate MRS ranges for the items in the new market – the matching items from the new query. Then, we calculate the probability of each item to be the best choice based on the two-step selection model. We first take a summation of all the probability that the candidate sets containing the current item is being pre-selected, and second calculate the probability that this item is the best choice using the MRS ranges we have within each candidate set. The above two key blocks are going to be explained in detail in next section using the concepts and properties described in this section.

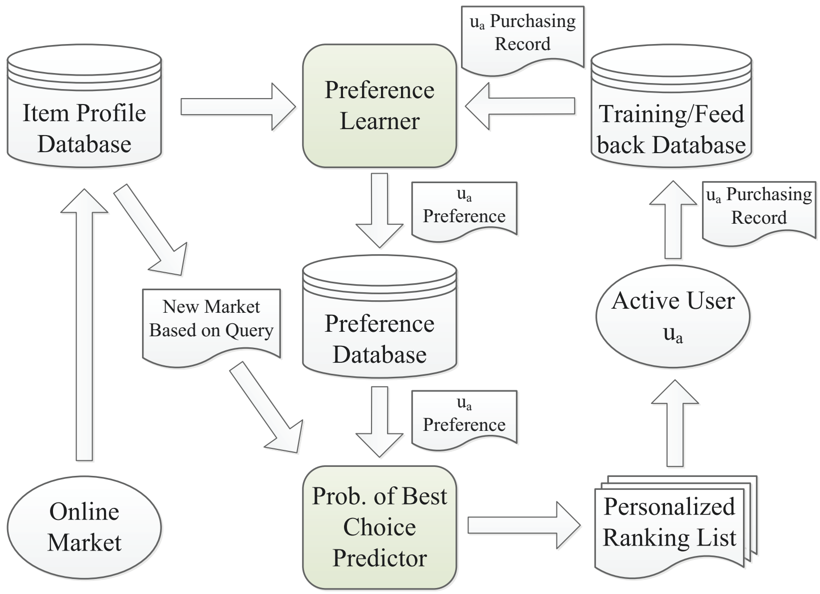


Fig. 3.2. Framework of proposed work

3.4 Performance Evaluation

Given the top-down ranking list of the N skyline items in the set S, let v_b be the ranking position of the (known) user's best choice s_b. $v_b = 1$ when s_b is ranked first by the ranking algorithm and considered to have the highest probability of being purchased by the user, and $v_b = N$ when it is considered to be the least favorable item for the user. We use two metrics to measure the performance of the proposed personalized ranking algorithm. First, we use *ranking quality*

$$rq = \frac{N - v_b}{N - 1}, \tag{3.1}$$

the percentage of items ranked worse than s_b with lower ranking positions than v_b [37]. A larger value of ranking quality indicates better accuracy, where $rq = 1$ when the best choice is accurately ranked the first, and $rq = 0$ when s_b is ranked the last. We also use *success rate* with parameter m,

$$sr = \frac{\text{Number of times } s_b \text{ is ranked among the top } m \text{ items in the list}}{\text{Total number of times of ranking}}, \tag{3.2}$$

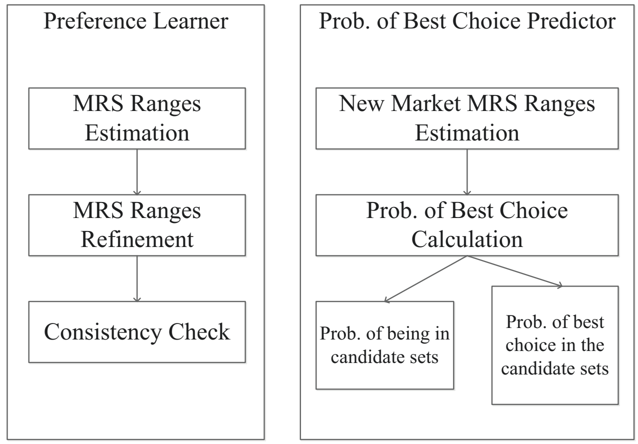


Fig. 3.3. Steps of learning and prediction module

the percentage of successful times to evaluate the performance of the proposed scheme. A larger value of success rate indicates that the algorithm has a higher chance to put s_b in the front of the list, which means a higher accuracy and better user experience.

Chapter 4

Probabilistic Ranking with MRS Range Estimation

In this chapter, we provide details of the learning and predicting module in the proposed framework in Fig. 3.3.

4.1 MRS Range Estimation – The Preference Learner

In this section, we use the concept and properties of MRS to estimate user preferences. Concretely, in the 2D price-reputation plane, let k_{s_i} be the slop of indifference curve at point s_i. It is also defined as the MRS of item s_i from the previous chapter. Based on the properties of indifference curves, we have a basic MRS range $k_i \in (-\infty, 0]$. Given a set of the user's historical purchasing records, we study in the following the method to acquire a set of narrowed down estimated MRS ranges (EMR) at multiple points. The set of learned EMRs are used as user preference profiles for the prediction module.

4.1.1 MRS Range Estimation from Single Transaction

We first consider one single transaction record $h = \{s_b, S\}$, where among a set of skyline items S with items being sorted as $r_1 < \cdots < r_N$ and $p_1 > \cdots > p_N$, the user purchases s_i, $s \in [1, N]$ as his/her best choice. Note that all the price and reputation we use in the following derivation are normalized.

Given the best choice $s_b = s_i$, we first divide the skyline items set $S - \{s_i\}$ into two subsets: the first subset includes all points above the best choice s_b. Denote the

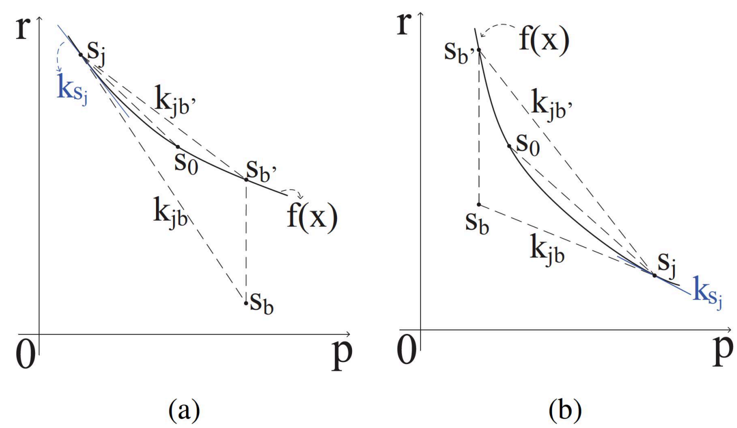


Fig. 4.1. (a) Slope range of $s_j \in S_+$, and (b) slope range of $s_j \in S_-$.

set as $S_+ = \{s_{i+1}, \cdots, s_N\}$. The "above" indicates the higher position in the 2-D domain of items in set S_+. The second subset includes all points below s_b. Denote the set as $S_- = \{s_1, \cdots, s_{i-1}\}$. We study the two sets separately and have Theorem 1 to calculate the EMR.

Theorem 1. Given S as the set of skyline items, for an item $s_j \neq s_b$, let k_{jb} be the slope of the line connecting s_j and the best choice s_b in the 2-D plane. For all $s_j \in S_+$, we have $k_{s_j} \leq k_{jb}$; and for all $s_j \in S_-$, we have $k_{jb} \leq k_{s_j} \leq 0$.

***Proof*:** In Fig. 4.1(a) where $s_j \in S_+$, assume the slope of the indifference curve at point s_j is $k_j > k_{jb}$. We will prove this theorem by contradiction. Since s_b is the best choice, we have $U(s_b) > U(s_j)$. Define $f(x)$ as the function of the indifference curve crossing s_j, and $k_{jb'}$ as the slope of line $s_j s_{b'}$, where $s_{b'} = (p_b, f(p_b))$. Since f is a convex function due to the diminishing MRS property, for any p_1 and p_2 and any $t \in [0, 1]$, we have $f(tp_1 + (1-t)p_2) \leq tf(p_1) + (1-t)f(p_2)$. Based on this property, for any point $s_0 = (p_0, f(p_0))$ on the indifference curve $f(x)$ with $p_j < p_0 < p_b$, we have

$$p_0 = \frac{p_b - p_0}{p_b - p_j} p_j + \frac{p_0 - p_j}{p_b - p_j} p_b,$$

$$f(p_0) \leq \frac{p_b - p_0}{p_b - p_j} f(p_j) + \frac{p_0 - p_j}{p_b - p_j} f(p_b),$$

$$\frac{f(p_0) - f(p_j)}{p_0 - p_j} \leq \frac{f(p_b) - f(p_j)}{p_b - p_j}.$$

Since $k_j = \lim\limits_{(p_0 - p_j) \to 0^+} \frac{f(p_0)-f(p_j)}{p_0-p_j}$ and $k_{jb'} = \frac{f(p_b)-f(p_j)}{p_b-p_j}$, we have $k_{s_j} \leq k_{jb'}$. Therefore, $k_{jb'} \geq k_{s_j} > k_{jb}$, and for point $s_{b'}$ and s_b we have $f(p_b) = y_j + k_{jb'}(p_b - p_j) > y_i + k_{jb}(p_b - p_j) = y_b$. Since s_b and s_b' have the same price p_b, we have $U(s_{b'}) > U(s_b)$ according to the monotonicity assumption of user behavior. Since s_j and $s_{b'}$ are on the same indifference curve, we have $U(s_j) = U(s_{b'})$, so $U(s_j) > U(s_b)$. This result contradicts to the fact that s_b is the best choice. Therefore, we should have $k_{s_j} \leq k_{jb}$. Using the same method, for $s_j \in S_-$ in Fig. 4.1(b), we could proof that $k_{jb} \leq k_j \leq 0$. □

Note that no MRS information for k_{s_b} could be extracted from the purchasing record, since k_{s_b} could be any slope to satisfy that s_b is preferred to s_j.

Therefore in each transaction, every point except s_b acquires the EMR $k_{s_j} \leq k_{jb}$ or $k_{jb} \leq k_{s_j} \leq 0$. Denote $\bar{k}_{s_j}$ as the upper bound of MRS range of s_j when $s_j \in S_+$, and $\underline{k}_{s_j}$ as the lower bound of MRS range of s_j when $s_j \in S_-$, so we have $\bar{k}_{s_j} = k_{jb}$ and $\underline{k}_{s_j} = k_{jb}$ in each case. Denote $\mathbb{K}_{s_j+}$ as the EMR for the first case and $\mathbb{K}_{s_j-}$ as the EMR for the second case, so we have $\mathbb{K}_{s_j+} = (-\infty, \bar{k}_{s_j}]$ and $\mathbb{K}_{s_j-} = [\underline{k}_{s_j}, 0]$.

In summary, From Theorem1, for every user online purchasing transaction record h with N skyline items, we could obtain $N-1$ narrowed down EMR. From multiple transaction records, we could obtain a combined EMR set used to describe user purchasing behavior or user preference profile.

It is worth to notice that EMR preserves all the preference relationships between $N-1$ items pairs by distinguishing the $\{s_b, s_j\}$, $s_j \in S - \{s_b\}$ pairs utilizing the properties of indifference curve and the different positions of item in the 2-D plane. It majorly distinguishes different user feelings when he/she compared s_b with different non-best-choice (NBC) items s_j that has not been chosen in this transaction. When compared to the boolean records that only use 0 or 1 to distinguish the purchasing behavior in the previous researches, the EMR preserves mush more information using the same source of information.

4.1.2 EMR Refinement within single transaction

Given the above initial EMRs, we can further refine the estimation by narrowing down some EMRs using the diminishing MRS property. Consider the example in Fig. 4.2(a), there are two items s_{i+1} and s_{i+2} above the best choice $s_b = s_i$ and $k_{(i+1)b} < k_{(i+2)b}$. From Theorem 1, initially, we have $\bar{k}_{s_{i+2}} = k_{(i+2)b}$ and $\bar{k}_{s_{i+1}} = k_{(i+1)b}$. Note that from the diminishing property of MRS, we should always have $k_{s_{i+1}} \geq k_{s_{i+2}}$, so we should have $\bar{k}_{s_{i+1}} \geq \bar{k}_{s_{i+2}}$. Therefore, we can update the upper bound $\bar{k}_{s_{i+2}}$ to $\bar{k}_{s_{i+2}} \leq \bar{k}_{s_{i+1}} \leq k_{(i+1)b}$. In other words, the inherited diminishing MRS property helps improve the accuracy of EMR. On the other hand, in the case of Fig.4.2(b), since we already have $k_{(i+2)b} < k_{(i+1)b}$, the transaction records provide a more accurate information than the inherit property. There is no improvement needed.

For a transaction record that has more than two items in S_+, the refinement is as follows. Let $s_j = s_{i+2}$ if $i + 2 \leq n$. Compare k_{jb} with $k_{(j-1)b}$, update the slope range as above and increase j by 1 at each step until $j = n$. As an example in Fig. 4.2(c), we first update $\bar{k}_{s_{i+2}}$ to $\bar{k}_{s_{i+2}} = k_{(i+1)b}$ since $k_{(i+2)b} > k_{(i+1)b}$. Secondly, since $k_{(i+3)b} < k_{(i+1)b}$, no improvement is needed for $\bar{k}_{s_{i+3}}$. Last, $\bar{k}_{s_{i+4}}$ should be updated to $\bar{k}_{s_{i+4}} = k_{(i+3)b}$ since $k_{(i+4)b} > k_{(i+3)b}$. Therefore, the results after the refinement would be $\bar{k}_{s_{i+1}} = k_{(i+1)b}$, $\bar{k}_{s_{i+2}} = k_{(i+1)b}$, $\bar{k}_{s_{i+3}} = k_{(i+3)b}$, $\bar{k}_{s_{i+4}} = k_{(i+3)b}$. Similarly, we can update lower bound of the EMRs for all items in S_- accordingly from s_{i-2} to s_1.

4.1.3 EMR Refinement for Multiple Transactions

From Theorem 1, with one purchasing record, we can refine the upper bounds of EMRs for all items above the best choice and the lower bounds of EMRs for all items below the best choice. Now, we combine EMRs from multiple transaction records and consider refinement for the summarized user preference profile.

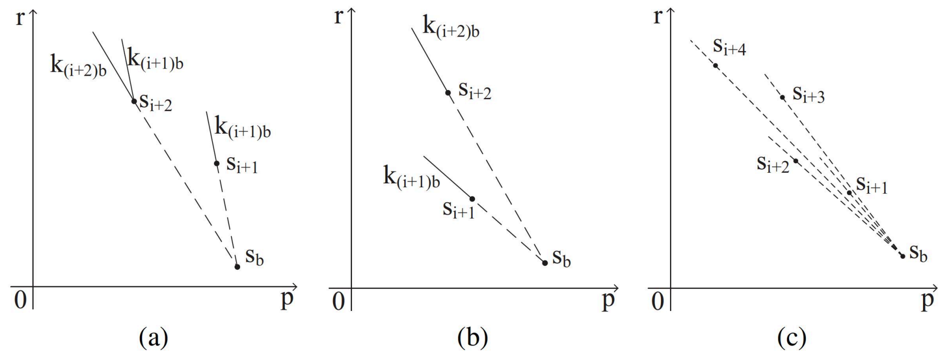


Fig. 4.2. Refinement of estimated MRS ranges in single transaction: (a)condition requiring update, (b)condition not requiring update, and (c) example of updating all the MRS ranges.

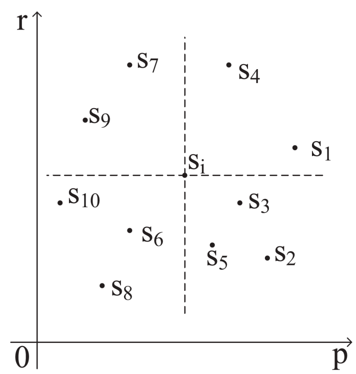


Fig. 4.3. Refinement of estimated MRS range with multiple transactions.

4.1.3.1 Refinement of one item from multiple EMRs

First, since an item s_i can belong to S_+ or S_- in different transactions, it may have both ranges in Theorem 1. Assume two transactions are combined together, there are four cases described in the following. EMRs from more than two transactions can be combined in the same way. Denote $\mathbb{K}^1_{s_i}$ and $\mathbb{K}^2_{s_i}$ as EMR of s_i acquired from the first and the second transaction record; denote $\mathbb{K}_{s_i}$ as the combined EMR.

- If $\mathbb{K}^1_{s_i} = (-\infty, \bar{k}^1_{s_i}]$, $\mathbb{K}^2_{s_i} = (-\infty, \bar{k}^2_{s_i}]$, and $\bar{k}^1_{s_i} \leq \bar{k}^2_{s_i}$, then the combined EMR $\mathbb{K}_{s_i} = (-\infty, \bar{k}^1_{s_i}]$.

- If $\mathbb{K}^1_{s_i} = [\underline{k}^1_{s_i}, 0]$, $\mathbb{K}^2_{s_i} = [\underline{k}^2_{s_i}, 0]$ and $\underline{k}^1_{s_i} \leq \underline{k}^2_{s_i}$, then the combined EMR

$\mathbb{K}_{s_i} = [\underline{k}^2_{s_i}, 0]$.

- If $\mathbb{K}^1_{s_i} = (-\infty, \bar{k}^1_{s_i}]$, $\mathbb{K}^2_{s_i} = [\underline{k}^2_{s_i}, 0]$, and $\underline{k}^2_{s_i} \leq \bar{k}^1_{s_i}$, then the combined EMR $\mathbb{K}_{s_i} = [\underline{k}^2_{s_i}, \bar{k}^1_{s_i}]$.
- If $\mathbb{K}^1_{s_i} = (-\infty, \bar{k}^1_{s_i}]$, $\mathbb{K}^2_{s_i} = [\underline{k}^2_{s_i}, 0]$, and $\underline{k}^2_{s_i} \geq \bar{k}^1_{s_i}$. This case is filtered out by consistency check, and the detailed reason is stated later.

In general, for a consistent user in the first three cases, the combined EMR $\mathbb{K}_{s_i}$ is the intersection of EMR for this item in all transactions:

$$\mathbb{K}_{s_i} = \bigcap_{j=1,\ldots,M} \mathbb{K}^j_{s_i}. \tag{4.1}$$

Define the default upper bound as $\bar{k}_{s_i} = 0$ and the default lower bound as $\underline{k}_{s_j} = -\infty$, then the combined EMR is

$$\mathbb{K}_{s_i} = [\max_{j=1,\ldots,M} \underline{k}^j_{s_i}, \min_{j=1,\ldots,M} \bar{k}^j_{s_i}] = [\underline{k}_{s_i}, \bar{k}_{s_i}]. \tag{4.2}$$

The reason of executing the intersection is that the true MRS belongs to all the EMRs if the user is consistent. It narrows down the EMRs to the greatest extent and accurately depicts the personal preference given the limited number of items in the markets.

4.1.3.2 Refinement of one item from other EMRs

For a given item s_i in the 2-D plane containing all the acquired EMRs from previous transactions, we divide the remaining items into four subsets: $S_i^I = \{s_j : p_j > p_i, r_j > r_i\}$, $S_i^{II} = \{s_j : p_j \leq p_i, r_j > r_i\}$, $S_i^{III} = \{s_j : p_j \leq p_i, r_j \leq r_i\}$ and $S_i^{IV} = \{s_j : p_j > p_i, r_j \leq r_i\}$, where each s_j has its own EMR. In the example in Fig. 4.3, $S_i^I = \{s_1, s_4\}$, $S_i^{II} = \{s_7, s_9\}$, $S_i^{III} = \{s_6, s_8, s_{10}\}$ and $S_i^{IV} = \{s_2, s_3, s_5\}$.

To further refine the estimation results from multiple records, we again use the diminishing property of MRS, and use items in S_i^{II} to update the lower bound of s_i's EMR $\underline{k}_{s_i}$, and use items in S_i^{IV} to update the upper bound of s_i's EMR $\bar{k}_{s_i}$. The diminishing property of MRS says $\bar{k}_{s_i} \leq \bar{k}_{s_j}$ for all $s_j \in S_i^{IV}$ and $\bar{k}_{s_i} \geq \bar{k}_{s_j}$ for all $s_j \in S_i^{II}$. Therefore, we refine

$$\bar{k}_{s_i} = \min_{s_j \in S_i^{IV} \cup s_i} \{\bar{k}_{s_j}\}, \tag{4.3}$$

$$\underline{k}_{s_i} = \max_{s_j \in S_i^{II} \cup s_i} \{\underline{k}_{s_j}\}. \tag{4.4}$$

.

4.1.4 Consistency Check

The last step is to check the consistency of the EMRs at each point. For point s_i, if its EMS is $[\underline{k}_{s_i}, \bar{k}_{s_i}]$, it should satisfy $\underline{k}_{s_i} \leq \bar{k}_{s_i}$. If $\underline{k}_{s_i} > \bar{k}_{s_i}$, it means that the user shows inconsistent behavior in the historical records by violating the diminishing MRS, and the corresponding personalized record h should be discarded to ensure accurate information collection.

In summary, the preference learner described above enables us to convert the historical purchasing records into *user preference profile* UProf in the form of narrowed down EMRs of every NBC items in the 2-D price-reputation plane. We denote the set of user preference profile as $\mathbb{H} = \{\mathbb{K}_{s_i}, i = 1, ..., N\}$.

4.2 Probability of Best Choice Predictor

By receiving a new query in the dynamically changing online market, the user obtains a new matching list of items. This section discusses details of predicting the probability of best choice for each item in the new matching list using the acquired user preference profile $\mathbb{H}$. The results are then used for personalized ranking.

4.2.1 MRS Range Estimation in the New Market

In the new matching list, items may have appeared in previous transactions and have their EMRs stored in $\mathbb{H}$. For these items, the EMRs will be used directly. However, there may be other items that do not have EMRs. There are two cases. First, items may have been chosen as the best choice in previous transactions and have no EMRs recorded. Second, new items entered the dynamic market recently and have not been in any transactions before. For these items, we estimate the EMRs using EMRs in $\mathbb{H}$.

For a new item s_i, according to equation (4.2), we need to estimate the upper bound $\bar{k}_{s_i}$ and lower bound $\underline{k}_{s_i}$ of its EMR. To estimate the upper bound, we first search all items in the historical data and find a set of its closest neighbors S' whose upper bounds are non-zero and $\bar{k}_{s_j} < 0$ for all $s_j \in S'$. We then estimate the upper bound of $\bar{k}_{s_i}$ using weighted sum $\bar{k}_{s_i} = \sum_{s_j \in S'} w_j \bar{k}_{s_j}$, where the weight $w_j = (d_{ij})^{-1} / \sum_{s_j \in S'}[(d_{ij})^{-1}]$ is inversely proportional to the distance d_{ij} between s_i and s_j. In this way, items closer to the estimated item will have larger influence on the EMRs. Similarly, to estimate the lower bound of $\bar{k}_{s_i}$, we find a set of its closest neighbors S' among all items in the historical data whose lower bounds are finite, and estimate $\underline{k}_{s_i}$ using $\underline{k}_{s_i} = \sum_{s_j \in S'} w_j \underline{k}_{s_j}$. We use 3 as the size of neighbor size in this work and observe similar results for other size of sets. Note that EMR refinement within single transaction stated in Section 4.1.2 is used again to ensure that items in the current market satisfy the diminishing MRS. There is no need to use the refinement for multiple transactions, since all the items within the new market (matching list of queried item) are skyline items.

4.2.2 Two-Step Ranking

After obtaining all EMRs for the new matching list of items, we have two steps to make the prediction. With the assumption in Section 3.1 that users can only compare a few items at the same time, we consider the scenario where a user first preselects n candidates that he/she might be interested in, and then makes detailed comparison within the n items. Define the preselected candidate set of item s_i as the Interested Set (IS) S_{IS_i}. To find the probability that an item s_i is the best choice, we first need to find all possible preselected candidate sets $S^j_{IS_i}$ that contains s_i, and then for each such set $S^j_{IS_i}$, find the probability that s_i has the largest utility compared to all other items in $S^j_{IS_i}$.

Mathematically, let $\mathcal{P}[S^j_{IS_i}]$ be the probability that $S^j_{IS_i}$ is the preselected

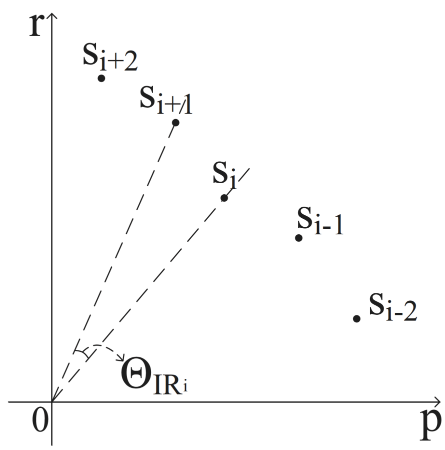


Fig. 4.4. Interested region and interested set

candidate set, and let $\mathcal{P}[s_i = \text{best}|S^j_{IS_i}]$ denote the probability that s_i is the preferred item among all in $S^j_{IS_i}$. Then, the probability that s_i is the user's best choice in the whole new matching list is

$$\mathcal{P}_{s_i} = \sum_{S^j_{IS_i}: s_i \in S^j_{IS_i}} \mathcal{P}[s_i = s_b | S^j_{IS_i}] \mathcal{P}[S^j_{IS_i}]. \tag{4.5}$$

4.2.2.1 Interested Set (IS)

To model the candidate pre-selection process, we adopt the *visual angle* model in [37]. In particular, for skyline item s_i, define its visual angle as $\psi_i = \arctan(r_i/p_i)$. The visual angle model is capable of roughly describing users preferences, e.g., a user who cares about reputation more than price will more likely to prefer items with large visual angles, and vice versa. Therefore, to find S_{IS_j}, we first let the interested visual angle range be $\Theta_{IR_j} = [\psi_j, \psi_{j+1})$ as shown in Fig. 4.4. A user's preferences are described by the probability that the user is interested in Θ_{IR_j} over the 2-D plane in the range of $[0, \pi/2]$. We divide the interested visual angle range by items so that the whole plane can be covered without overlapping, and $\sum_j \mathcal{P}[S^j_{IS_i}] = 1$. For the interested range below item s_1, denote $\Theta_{IR_0} = [0, \psi_1)$. For the interested range above item s_n, denote $\Theta_{IR_n} = [\psi_n, \pi/2]$. Given Θ_{IR_j}, we assume that a user is aware of two closest neighbor items. Therefore, in addition to the items in the interested set, the user also looks for two closest neighbors for comparison. The assumption is reasonable based on the *bounded rationality* stated in section 3.1.2 and usual user behavior.

Under this case, the user will preselect $n = 4$ candidate items for the corresponding $S^j_{IS_i}$ before making detailed comparison. For example, for Θ_{IR_i} in Fig. 4.4, we have $S^i_{IS_i} = \{s_{i-1}, s_i, s_{i+1}, s_{i+2}\}$. Exceptions happen for corner cases that $S_{IS_0} = \{s_1, s_2\}$, $S_{IS_1} = \{s_1, s_2, s_3\}$, $S_{IS_{n-1}} = \{s_{n-2}, s_{n-1}, s_n\}$, $S_{IS_n} = \{s_{n-1}, s_n\}$. Therefore, item s_i in Fig. 4.4 is included in four candidate sets $S^{i-2}_{IS_i}$, $S^{i-1}_{IS_i}$, $S^i_{IS_i}$ and $S^{i+1}_{IS_i}$, so $j \in \{i-2, i-1, i, i+1\}$, and the corresponding interested regions are $\Theta_{IR_{i-2}} = [\psi_{i-2}, \psi_{i-1})$, $\Theta_{IR_{i-1}} = [\psi_{i-1}, \psi_i)$, $\Theta_{IR_i} = [\psi_i, \psi_{i+1})$, and $\Theta_{IR_{i+1}} = [\psi_{i+1}, \psi_{i+2})$.

To calculate $\mathcal{P}[S^j_{IS_i}]$, a preference density function $f(\psi)$ is used to model the probability that a user is interested in items at angle ψ in the price-reputation plane, and we use the same method as in [37] to estimate $f(\psi)$. Therefore, the probability that $S^j_{IS_i}$ is the pre-selected candidate set is

$$\mathcal{P}[S^j_{IS_i}] = \int_{\psi_j}^{\psi_{j+1}} f(\psi)\mathrm{d}\psi. \tag{4.6}$$

4.2.2.2 Probability of Best Choice in S_{IS_j}

The next step is to compute $\mathcal{P}[s_i = s_b | S_{IS_j}]$, the probability that s_i is the best choice in S_{IS_j}.

We first consider a simple scenario of comparing two items s_A and s_B where s_A is below s_B, as shown in Fig. 4.5(a). To determine the probability that s_A is preferred to s_B, we first consider the indifference curve IC_{s_A} that passes through s_A. Since all points on IC_{s_A} have the same utility as s_A, we can compare s_B with any point on IC_{s_A}. In this work, we choose the point $s_{A'}$ whose distance to s_A is the same as that between s_A and s_B. Let k_{AB} denote the slope of the line connecting s_A and s_B, and $k_{AA'}$ be the slope of the line connecting $s_{A'}$ and s_A. We define function $\theta(k) = \pi + \arctan(k)$ to convert slope k to angle θ, and we have $\theta_{AB} = \theta(k_{AB})$ and

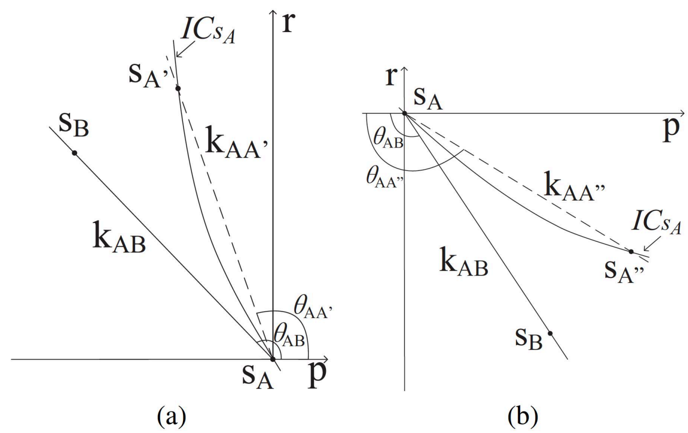


Fig. 4.5. Two-item comparison: (a) s_A is below s_B, and (b) s_A is above s_B.

$\theta_{AA'} = \theta(k_{AA'})$, as shown in Fig. 4.5(a). Since $s_{A'}$ and s_B have the same distance to s_A, comparing their positions is equivalent to comparing the two angles θ_{AB} and $\theta_{AA'}$. From Fig. 4.5(a), it is easy to see that when $\theta_{AB} > \theta_{AA'}$, $s_{A'}$ is above s_B and $U(s_A) > U(s_B)$, and vice versa.

To compare θ_{AB} and $\theta_{AA'}$, note that the indifference curve IC_{s_A} is convex, and we have $-\infty < k_{AA'} \leq k_A \leq \bar{k}_A$ where k_A is the true slope of IC_{s_A} at point s_A, and $\bar{k}_A$ is the estimated upper bound of k_A. Define $\bar{\theta}_A = \theta(\bar{k}_A)$ and $\theta_A = \theta(k_A)$. Without any prior knowledge of $\theta_{AA'}$ or the position of $s_{A'}$, we assume that $\theta_{AA'}$ is uniformly distributed in the range $[\pi/2, \bar{\theta}_A]$. Note that if $\theta_{AB} > \bar{\theta}_A$, we have $\theta_{AB} > \theta_{AA'}$, and thus s_A is always preferred to s_B. Therefore, the probability that s_A is preferred to s_B is

$$\mathcal{P}[U(s_A) \geq U(s_B)] = \begin{cases} \frac{\theta_{AB}-\pi/2}{\bar{\theta}_A-\pi/2} & \text{if } \pi/2 \leq \theta_{AB} \leq \bar{\theta}_A, \\ 1 & \text{if } \bar{\theta}_A < \theta_{AB}. \end{cases} \tag{4.7}$$

By the same token, when s_A is above s_B as shown in Fig. 4.5(b), we have

$$\mathcal{P}[U(s_A) \geq U(s_B)] = \begin{cases} \frac{\pi-\theta_{AB}}{\pi-\underline{\theta}_A} & \text{if } \underline{\theta}_A \leq \theta_{AB} \leq \pi, \\ 1 & \text{if } \theta_{AB} < \underline{\theta}_A. \end{cases} \tag{4.8}$$

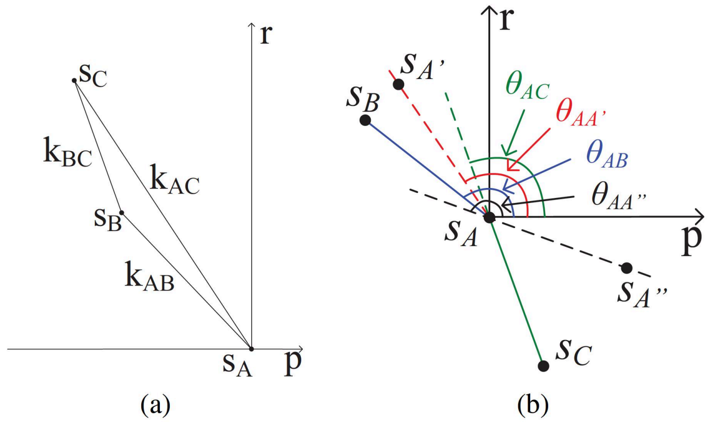
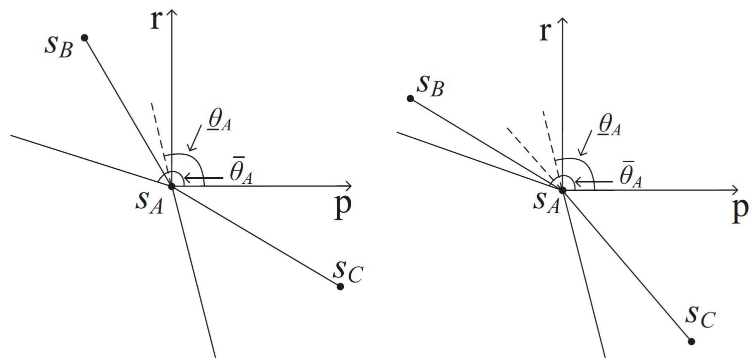


Fig. 4.6. Three-item comparison: (a) s_A is below s_B and s_C , and (b) s_A is below s_B but above s_C.

Now we consider the scenario where we compare three or more items at the same time. We first consider the three-item comparison and calculate $\mathcal{P}[s_i = \text{best}|S_{IS_j}] = \mathcal{P}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)]$. We divide the comparison into three cases. In the first case, s_A is below s_B and s_C as shown in Fig. 4.6(a). In this case, we choose the most competitive item from s_B and s_C, and use it to compare with s_A based on (4.7), so that the three-item comparison can be reduced to two-item comparison. By competitive, we first find out the item with the smallest slope of the line connecting itself and s_A. In the example of Fig. 4.6(a), denote $\theta_{AB} = \theta(k_{AB})$ and $\theta_{AC} = \theta(k_{AC})$, where k_{AB} and k_{AC} are the slope of the lines connecting s_A and s_B, and s_A and s_C, respectively. According to (4.7), since $\theta_{AB} > \theta_{AC}$, we have $\mathcal{P}[U(s_A) \geq U(s_B)] \geq \mathcal{P}[U(s_A) \geq U(s_C)]$, which means in the respective of s_A, it has less chance to beat s_C than to beat s_B, thus s_C is considered more competitive than s_B. One exception is that if $\bar{k}_B < k_{BC}$, according to (4.7), we have $P[U(s_B) \geq U(s_C)] = 1$, then s_B is always preferred to s_C and is the most competitive item. Otherwise, we assume s_C as the most competitive one. Using the same method, in the second case of three-item comparison, when s_A is above both s_B and s_C, we could also find the most competitive item from s_B and s_C , and reduce the three-item comparison to a two-item comparison in (4.8). In the third case where s_A is below s_B but above s_C as shown in Fig. 4.6(b),



(a) (b)

Fig. 4.7. Three-item comparison: s_A is below s_B but above s_C: (a) $\theta_{AB} \leq \theta_{AC}$, and (b) $\theta_{AB} > \theta_{AC}$.

contrary to the first two cases, $\mathcal{P}[U(s_A) \geq U(s_B)]$ and $\mathcal{P}[U(s_A) \geq U(s_C)]$ are incomparable since they use different formulas (4.7) and (4.8). The comparison is as follows. Same as Fig. 4.5(a), we consider the indifference curve IC_{s_A} that passes through s_A and find the point $s_{A'}$ on IC_{s_A} that has the same distance to s_A as the distance to s_B. The definitions of $\theta_{AA'}$, θ_{AB} and $\bar{\theta}_A$ are the same as above. Similarly, we find another point $s_{A''}$ on IC_{s_A} that has the same distance to s_A as the distance to s_C. Let k_{AC} be the slope of the line connecting s_A and s_C, and define $\theta_{AC} = \theta(k_{AC})$. The definitions of $\theta_{AA''}$ and $\underline{\theta}_A$ are the same as above. Note that the indifference curve IC_{s_A} is convex. Thus, we have $k_{AA'} < k_{AA''}$ and $\theta_{AA'} < \theta_{AA''}$. Therefore, the probability that s_A has the largest utility among the three is equivalent to the probability that $\theta_{AA'} \leq \theta_{AB}$ and $\theta_{AA''} \geq \theta_{AC}$ under the constraint that $\theta_{AA'} < \theta_{AA''}$. Without any prior knowledge of the positions of $s_{A'}$ and $s_{A''}$, we assume that $\theta_{AA'}$ and $\theta_{AA''}$ are independent. Following the same analysis as (4.7) and (4.8), the calculation can be divided into following cases.

- If $\mathcal{P}[U(s_A) \geq U(s_B)] = 1$ or $\mathcal{P}[U(s_A) \geq U(s_C)] = 1$, we have

$$\begin{aligned}
&\mathcal{P}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)] \\
&\qquad = \mathcal{P}[U(s_A) \geq U(s_B)] \cdot \mathcal{P}[U(s_A) \geq U(s_C)]. \quad (4.9)
\end{aligned}$$

Since s_A is always preferred to at least one of s_B and s_C, the $\mathcal{P}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)]$ is reduced to $\mathcal{P}[U(s_A) \geq U(s_B)]$, $\mathcal{P}[U(s_A) \geq U(s_C)]$, or 1.

- For other cases as shown in Fig. 4.7, we first integrate all the possible combinations of $\theta_{AA'}$ and $\theta_{AA''}$ under the constraint that $\theta_{AA'} < \theta_{AA''}$, and the integrated area is

$$\mathcal{I} = \int_{\pi/2}^{\underline{\theta}_A} (\pi - \underline{\theta}_A) d\alpha + \int_{\underline{\theta}_A}^{\bar{\theta}_A} (\pi - \alpha) d\alpha = -\frac{1}{2}(\bar{\theta}_A - \pi)^2 - \frac{1}{2}(\pi/2 - \underline{\theta}_A)^2 + \frac{\pi^2}{8}. \tag{4.10}$$

  - If $\theta_{AB} \leq \theta_{AC}$ (Fig. 4.7(a)), the condition that $U(s_A) \geq U(s_B)$ and $U(s_A) \geq U(s_C)$ is $\pi/2 \leq \theta_{AA'} \leq \theta_{AB}$ and $\theta_{AC} \leq \theta_{AA''} \leq \pi$, thus we have

$$\mathcal{I}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)] = (\theta_{AB} - \pi/2)(\pi - \theta_{AC}). \tag{4.11}$$

    Combining (4.10) and (4.11), we have

$$\begin{aligned} &\mathcal{P}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)] \\ =&\frac{\mathcal{I}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)]}{\mathcal{I}} \\ =&\frac{-2 * (\theta_{AB} - \pi/2)(\pi - \theta_{AC})}{(\bar{\theta}_A - \pi)^2 + (\pi/2 - \underline{\theta}_A)^2 - \pi^2/4}. \end{aligned} \tag{4.12}$$

  - If $\theta_{AB} > \theta_{AC}$ (Fig. 4.7(b)), the condition that $U(s_A) \geq U(s_B)$ and $U(s_A) \geq U(s_C)$ is $\pi/2 \leq \theta_{AA'} \leq \theta_{AB}$ and $\theta_{AC} \leq \theta_{AA''} \leq \pi$ under the constraint that $\theta_{AA'} < \theta_{AA''}$, thus we have

$$\begin{aligned} &\mathcal{I}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)] \\ =&(\theta_{AC} - \pi/2)(\pi - \theta_{AC}) + \int_{\theta_{AC}}^{\theta_{AB}} (\pi - \alpha) d\alpha \\ =&-\frac{\pi^2}{2} + \frac{\pi}{2}\theta_{AC} + \pi\theta_{AB} - \frac{1}{2}{\theta_{AB}}^2 - \frac{1}{2}{\theta_{AC}}^2. \end{aligned} \tag{4.13}$$

Combining (4.10) and (4.13), we have

$$\begin{aligned}&\mathcal{P}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)]\\=&\frac{\mathcal{I}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)]}{\mathcal{I}}\\=&\frac{\pi^2 - \theta_{AC}\pi - 2\theta_{AB}\pi + {\theta_{AB}}^2 + {\theta_{AC}}^2}{(\bar{\theta}_A - \pi)^2 + (\pi/2 - \underline{\theta}_A)^2 - \pi^2/4}.\end{aligned} \tag{4.14}$$

Therefore, for three-item comparison where s_A is not always better than s_B and s_C, we have

$$\begin{aligned}&\mathcal{P}[U(s_A) \geq U(s_B), U(s_A) \geq U(s_C)]\\=&\begin{cases}\frac{-2*(\theta_{AB}-\pi/2)(\pi-\theta_{AC})}{(\bar{\theta}_A-\pi)^2+(\pi/2-\underline{\theta}_A)^2-\pi^2/4} & \text{if } k_{AB} \leq k_{AC},\\ \frac{\pi^2-\theta_{AC}\pi-2\theta_{AB}\pi+{\theta_{AB}}^2+{\theta_{AC}}^2}{(\bar{\theta}_A-\pi)^2+(\pi/2-\underline{\theta}_A)^2-\pi^2/4} & \text{if } k_{AB} > k_{AC}.\end{cases}\end{aligned} \tag{4.15}$$

Using the same method, given s_A, s_B and s_C, we can also calculate the probabilities that s_B/s_C are preferred among the three, respectively. The comparison among four or more items is similar.

In summary, for each skyline item in a new market, we use (4.5) - (4.15) to compute the probability that it is the user's best choice. We then rank all items in the descending order of $\mathcal{P}_{s_i}$.

Chapter 5

Simulation Results

To validate the performance of the proposed indifference curve based (IC-based) method, in this chapter, we use synthetic markets to simulate practical online shopping queries and use synthetic user behaviors to determine best choice items for each transaction. Then, we evaluate the performance of this work using ranking accuracy and success rate defined in Chapter 3, and compare them with prior works.

5.1 Market Simulation

For each query, we generate a synthetic market with N skyline items, where N is a randomly chosen number in the range of $[20, 100]$. Then, we generate N price values in the range $[10, 10^3]$ and N reputation values in the range $[0, 10^6]$, both following the power law distribution. Without loss of generality, we order the reputation and price values from low to high with $P_1 < \cdots < P_N$ and $R_1 < \cdots < R_N$. Then, N skyline items are generated with (P_i, R_i) being the ith item. We follow Section 3.1 to normalize all prices and reputations and use $\{p_i, r_i\}$ for analyses.

5.2 User Behavior Simulation

To simulate users' purchasing behavior, we first use the widely used Cobb-Douglas model in economics [1] to model the user's level of satisfaction for an item. For an item with normalized price p and normalized reputation r, the item's utility function is $U(p, r) = p^{\alpha} \cdot r^{\beta}$ with $\alpha \geq 0$ and $\beta \geq 0$ being the parameters quantifying the importance of price and reputation to the user, respectively. We simulate the user purchasing behavior using this synthetic user behavior model, and let predicting algorithms learn the user behavior and make predictions. We consider five different categories of users, summarized as follows:

Type 1: Users in this category consider price and reputation to be equally important, and we use $\alpha = \beta = 1$ as an example.

Type 2: Users in this category consider price to be more important than reputation, and we choose $\alpha = 2$ and $\beta = 1$ as an example. For other values of α and β with $\alpha > \beta$, we observe the same trend.

Type 3: These users consider that reputation is more important than price and we use $\alpha = 1$ and $\beta = 2$ as an example.

Type 4: Users in this category consider price only and always choose the cheapest item. We use $\alpha = 1$ and $\beta = 0$ as an example.

Type 5: Users in this category consider reputation only and always choose the item with the highest reputation in the market. We use $\alpha = 0$ and $\beta = 1$ as an example.

For a user, his/her utility function is used as the ground truth to choose his/her best choice. For each type of users, we repeat the simulations 30,000 times and average the results.

5.3 Simulation Results and Discussions

Table 5.1 and Fig. 5.1 compare the IC-based method with MAPS [37] and the weighted sum approach [18], [47], [48] using ranking quality and success rate. The IC-based method and MAPS learn users' personal preference based on their previous $L = 5$ purchasing records. Since the performance of the weighted sum approach is very sensitive to the selected weight parameter γ, we walk through all the weights and take its best, worst, and average performance into comparison. Table 5.1 shows the ranking quality results and Fig. 5.1 gives the success rate results.

We can see that for both ranking quality and success rate, IC-based method outperforms both the weighted sum approach and MAPS in most cases. For extreme users Type-4 and Type-5, all three algorithms perform very well due to the fixed user

TABLE 5.1
SIMULATION RESULTS OF THE RANKING QUALITY OF DIFFERENT MULTI-ATTRIBUTE RANKING ALGORITHMS.

		Type 1	Type 2	Type 3
IC-based		96.55%	94.63%	97.01%
MAPS		90.71%	86.06%	93.56%
Weighted sum	Max	86.56%	76.25%	92.87%
	Min	15.04%	25.16%	9.97%
	Average	68.20%	61.61%	72.47%

		Type 4	Type 5
IC-based		99.87%	99.41%
MAPS		98.79%	99.14%
Weighted sum	Max	99.99%	100%
	Min	0%	1.57%
	Average	48.0%	75.83%

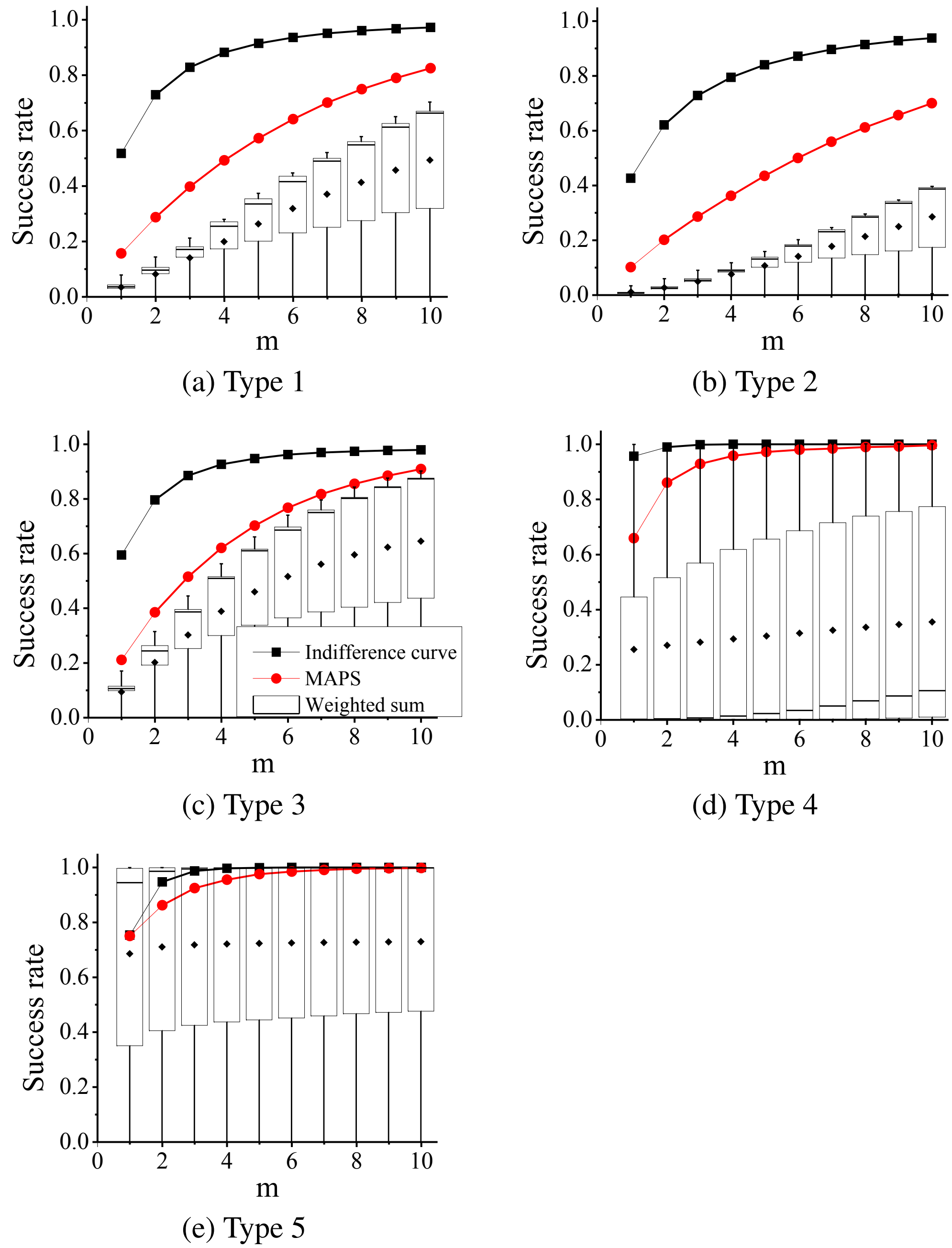
behavior. However, the results shows that the weighted sum method requires careful selection of the weight, otherwise, the results will be much worse. In contrast, the proposed method is relatively insensitive to such parameter selection. In addition, IC-based method shows more advantage in success rate. In Fig. 5.1, when $m \leq 3$, the improvements of proposed work from MAPS are 43.04%, 44.24%, 37.0% of user Type 1, 2, and 3. Even for extreme user Type 4 and 5 where everyone performs well, the largest improvements are 31.24% for user Type 4 with $m = 1$, and 8.48% for user Type 5 with $m = 2$. Even more improvement can be seen when compared to weighted sum approach. The significant improvement in success rate highlights that the proposed work performs better in ranking the ground truth higher, especially in the very top rankings. Noted that we have $[20, 100]$ items in a market. This means that users will have a better experience in that they can have higher chances to meet their best choice within the first few listed items/the first page by using the proposed algorithm.

The previous results evaluated ranking performances for different user behaviors. The following will analyze the algorithm from other aspects and consider facts that may affect the performances. We used user Type 1 as an example to show the trends and set all the parameters default as stated before, except the one under analysis. The four curves in each figure in Fig. 5.2 represent the ranking qualities and success rates from the IC-based method and MAPS. Fig. 5.2(a) evaluates the impact of number of historical purchasing records used in learning algorithms. The IC-based method and MAPS have similar rate of improvement with the increase of history, and both become stable at around 5 records for ranking quality. In addition, both systems do not lose accuracy dramatically even with only 1 or 2 records, indicating short learning curve without the need of large number of user past records. The short learning curve is more practical in the real systems that always face with sparse data, and it is better to keep only the most recent records, since user preferences change with time.

Fig. 5.2(b) shows the assessment of the impact of number of candidate neighbors that a user considers each time. Since MAPS does not use the two-step ranking, the accuracy of MAPS does not change. The accuracy for IC-based method is higher with more items. However, there is no dramatic improvement. Small number of neighbors, such as 4, could already produce satisfied results. This feature shows that the IC-based method has the capacity to capture personalized preference with bounded rationality, and at the same time, it benefits the system by reducing computational complexity.

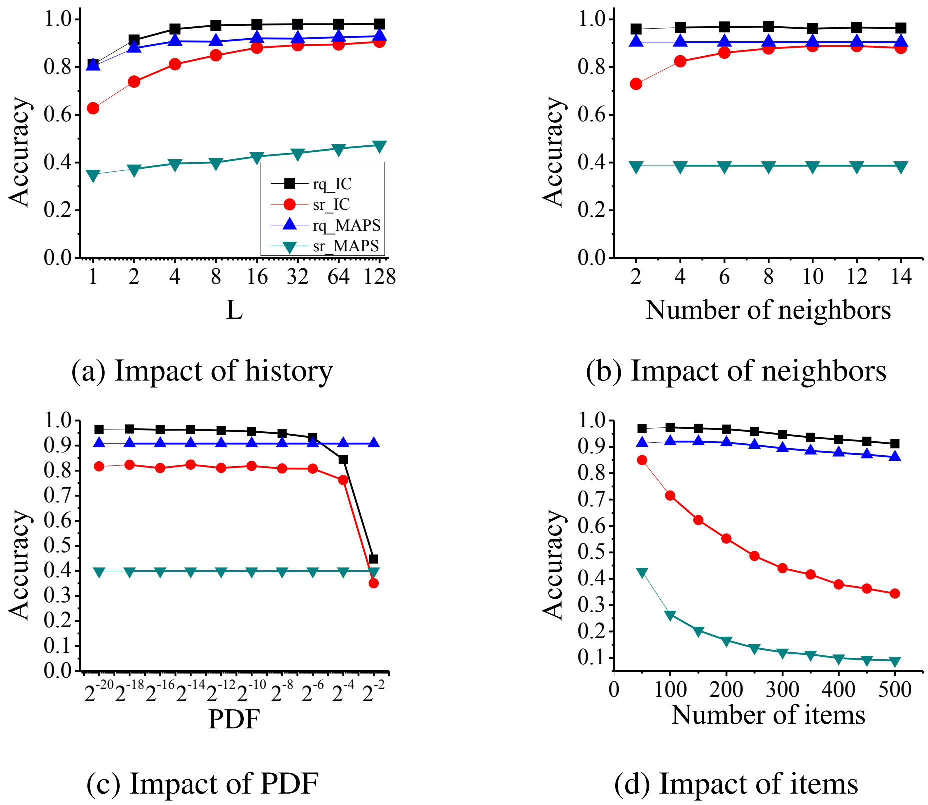
Fig. 5.2(c) is to evaluate the impact if we first exclude the condition when $\mathcal{P}[S_{IS_j}]$ is too small. In other words, if the user is highly unlikely to be interested in certain visual angle region that $\mathcal{P}[S_{IS_j}] < \theta$ (where θ is the threshold), we set $\mathcal{P}[S_{IS_j}] = 0$ and skip the computation of $\mathcal{P}[s_i = s_b | S_{IS_j}]$. The θ is set to be 0.5^i, $i = 2, 4, \ldots, 20$ in Fig. 5.2(c). The result shows that when $threshold < 0.5^6$, there is not much difference in accuracy, which is equivalent of reducing 60% of the computation complexity in general from the simulation. The percentage of the reduction depends on the shape of the personalized density function. The more consistent the user is with the interested region, the more concentrated the density function is, and the more reduction we can achieve. This feature is important for websites to preserve good feeding performance when queries are large. With the fact that most users have a consistent interested region, the proposed work performs much better than MAPS with not too much complexity added.

Finally, Fig. 5.2(d) represents the impact of the size of item sets $N (N = 50, 100, \ldots, 500)$. The performances of ranking quality is not affected much by different N. The success rates are approximately inversely proportional to N, which means the performances do not get worse.



(a) Type 1 (b) Type 2 (c) Type 3 (d) Type 4 (e) Type 5

Fig. 5.1. Simulation results of the success rates of different multi-attribute ranking algorithms.



(a) Impact of history

(b) Impact of neighbors

(c) Impact of PDF

(d) Impact of items

Fig. 5.2. Impact of parameters

Chapter 6

Real User Test Results

To further validate the performance of the IC-based method in the real world, we established the real user test and compared it with prior researches. All the input data used in the real user test are from real data instead of synthetic ones. Concretely, we collected online real data from eBay, processed it to create a survey suitable for real user test, and obtained survey results from 21 subjects.

6.1 Data Collection and Processing

6.1.1 Online Data Processing

To study user behavior in different price ranges, we collected online data from three products, coffee maker as the low-price range product, Itouch as the middle-price range product and SLR camera as the high-price range product. The online data collection steps are as follows. Take the Cuisinart Coffee Maker (price around $\$100$) as an example, first, enter key words for inquiry, e.g., 'Cuisinart coffee maker dcc 1200'; second, go to advanced search and select the specific area, e.g., 'Small Kitchen Appliances', exclude key words that indicate it to be accessories/not the exact product, e.g., 'filter', 'cup', etc, choose 'buy it now' to get the selling price, choose 'new' to exclude second-hand items, and choose 'price + shipping: lowest first' to make finding skyline sellers afterwards easier. After search, we collected 28 sellers with item No., price and reputation, and their total prices are sorted from low to high. Likewise, we collected 50 sellers from 'Itouch 5th generation' (price around $\$200$) and 50 sellers from 'canon 5d mark ii' (price around $\$2000$).

6.1.2 Offline Data Processing

After obtaining the raw data, we processed the data offline to create a survey suitable for real user test. For each product, we created 15 groups of listed items from the raw data. Subjects(i.e., users) will choose one item from each group as the one they want to buy based on given information. Table 6.1 shows a example.

TABLE 6.1
SAMPLE: A GROUP OF CUISINART COFFEE MAKER DCC 1200

Item No.	Price	Overall Rating	Your Choice
601981	75.99	3	
262971	97.4	3.978	
654079	109.2	4.024	
689214	113.72	4.181	
748151	121.13	4.991	

After several round of small scale pre-tests with feedbacks, we decided to use the normalized 5-star rating system instead of the raw accumulated ratings in eBay to reduce unnecessary wrong decisions. The reasons are as follows.

- Some products, e.g., daily necessities, sell in large amount and their accumulated ratings maybe much higher than products that are only needed by one per household and are not bought frequently, such as microwaves, strollers, tables, etc. With normalized 5-star rating system, subjects could use the same personalized preference scale to make decision rather than frequently adjust their preference on different products in this short test. This will reduce the inconsistent decision makings.

- 5-star system is more visualized and people have more experiences on how the seller or item are expected. On the other hand, the accumulated ratings are large obscure numbers, which could easily distract subjects from simply making decisions to figuring out large numbers.

- The lack of upper bound in accumulated ratings could also confuse users with different ratings. For example, two sellers may have ratings 5000, and 50000.

It is not necessarily means the first seller is much worse than the second. 5000 may be already good enough and equivalent to 4.9 rating in the 5-star rating system. Subjects facing accumulated ratings may make wrong decisions due to not looking at the numbers carefully. This case is also supposed to be reduced.

- In this controlled survey, the subjects are required to only use the provided two dimensional information as shown in Table 6.1. They should not consider the percentage of positive feedback, the positive/negative feedback in the past 1/6/12 months, the location of the product, the quality of communication with sellers, shipping time and shipping and handling charges, etc. However, the accumulated rating score is suspicious to subjects from their past experience of online shopping. For example, they know that sellers who are not good enough could also accumulate a high rating with lower percentage of positive feedback in eBay. This concern will distract subjects' attention from making decisions to wondering issues irrelevant to the test. In contrast, the 5-star rating system always tells user the overall expectation the user should have for this purchasing behavior.

We used the following steps to generate the groups of listed sellers.

1. We normalized the reputations to the 5-star system and set the reputation ranges as $[3, 5]$ to let all the ratings in the survey acceptable to most subjects.

$$r_i = \frac{R_i}{\max\{R_i, i = 1, ..., N\}} \times 2 + 3, \tag{6.1}$$

where r_i is the normalized reputation, R_i is the real reputation as defined before, $\max\{R_i, i = 1, ..., N\}$ is the maximum reputation of the listed sellers, and N is the number of listed sellers in the group.

The reason to set 3 as the lowest reputation is because the average rating is 3. According to the 'A Statistical Analysis of 1.2 Million Amazon Reviews'[49], more than half of the reviews give a 5-star rating. Aside from perfect reviews, most reviewers give 4-star or 1-star ratings, with very few giving 2 stars or 3 stars relatively. This indicates that most reviewers are so polite when giving ratings and it makes 3 stars a quite unacceptable rating for most products. Therefore, setting the lowest reputation to be 3 can let every item in the survey competitive. In contrast, non-competitive items are not necessary to appear in the survey.

2. Under the sorted sellers with non-decreasing reputation, we excluded certain number of wholesale sellers so that the normalized reputations could spread well in the 5-star system instead of getting clustered. An example of clustered ratings is shown in Table 6.2, where normalized ratings of item 1, 2 and 3 are clustered. Their ratings are very similar although the original accumulated ratings are quite different. The reason is that the highest accumulated rating $\max\{R_i, i = 1, ..., N\}$ is much higher than most of the other ratings, and most of the normalized reputations will be compressed to be very close to 3.

   Clustered ratings cannot reflect the original ratings well and most user behaviors will become extreme, which will make the real user test trivial. Concretely, in this extreme case, some subjects will choose item 1 since with similar reputations in item 1, 2, and 3, item 1 has the lowest price. Others will choose item 4 to have a reasonable reputation. In contrast, in the real case, item 2 and 3 would have a good chance to be the best choice. With the existence of wholesale sellers, most testing groups will have clustered ratings, and the user behavior will be compressed to the extreme behavior as stated above, which makes the experiment meaningless.

TABLE 6.2
SAMPLE GROUP OF CLUSTERED RATINGS

Item No.	Price	Accumulated Rating	Re-scaled Rating	Your Choice
1	75.99	7	3.00	
2	78.99	7773	3.03	
3	85.99	14525	3.06	
4	89.95	292942	4.26	

TABLE 6.3
RELATIVE ENTROPY OF REPUTATIONS AFTER DELETING 0-9 WHOLESALE SELLERS

	0	1	2	3	4
Coffee Maker	3.259	2.979	2.137	1.320	1.496
Itouch	3.316	3.301	3.031	2.716	1.601
SLR Camera	3.830	3.092	2.624	2.468	1.559
	5	6	7	8	9
Coffee Maker	1.403	1.206	1.333	1.570	1.513
Itouch	1.364	1.351	1.440	1.347	1.494
SLR Camera	1.587	1.336	1.344	1.256	1.395

To solve the clustered problem and have ratings well spread, relative entropy[50] is used as a criterion. Relative entropy measures the distance between reputations' distribution and uniform distribution. In equation (6.2), $p(x)$ is density function of uniform distribution and $q(x)$ is the density function of reputations. Discrete density function with scale 0.1 is used. The results in Table 6.3 show the relative entropy of remaining sellers' reputations if i ($i = 0, 1, ...9$) sellers with the highest reputation are excluded. According to the results, we chose the first entry whose relative entropy is less than half of the original one in the 0-th column in Table 6.3. Thus, we exclude 3 wholesale sellers for coffee maker, and 4 for Itouch and camera.

$$D(p||q) = \sum_{x \in \mathscr{X}} p(x) \log_2 \frac{p(x)}{q(x)}. \tag{6.2}$$

3. For each product, we re-normalize the reputations to the 5 star system with the remaining sellers after excluding the wholesale sellers as in equation (6.3).

$$r_i = \frac{R_i}{\max\{R_i, i = 1, ..., N'\}} \times 2 + 3. \tag{6.3}$$

Algorithm 6.1 specifies the procedure for offline data processing. Note that N_groups is the number of generated groups, and $N\_sellers$ is the number of sellers in each group.

By applying the above steps, we have 3 products in the survey and each product

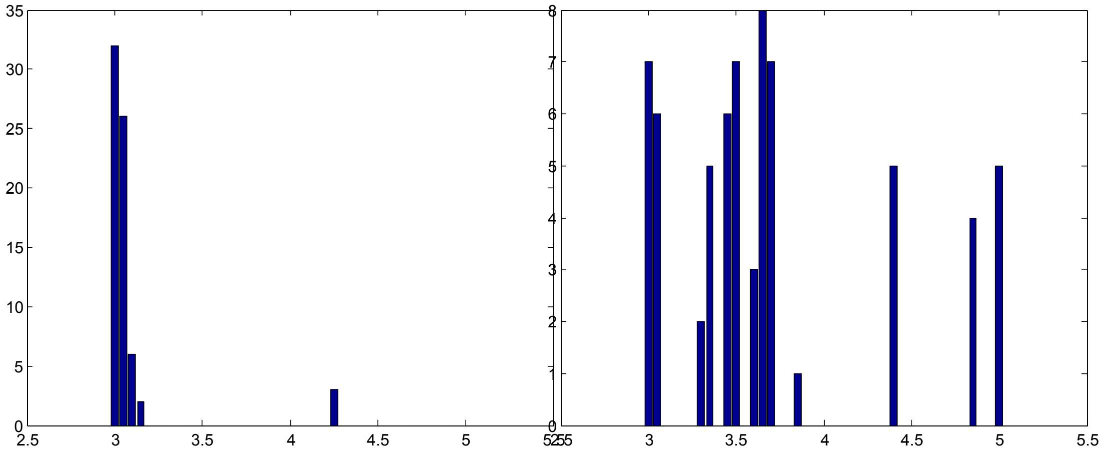
Algorithm 6.1 Offline data processing

Require: Available sellers' price and reputation from online query
Ensure: 15 different groups of skyline sellers (groups size > 3)
  Sort sellers lowest ratings first
  Convert accumulated ratings to 5-star ratings using equation (6.1)
  Calculate relative entropy
  while $D(p||q) > 0.5\dot{D}_0(p||q)$ **do**
    Delete the seller with the highest rating
    Convert accumulated ratings 5-star ratings using equation (6.3)
    Calculate relative entropy
  while $N\_groups < 15$ **do**
    Randomly select 10 sellers
    Filter out non-skyline sellers
    if $N\_sellers > 3$ **then**
      if The generated group is different from existing groups **then**
        Save the generated group
        $N\_groups + 1$
  return 15 generated groups

contains 15 groups of listed items for real user test. Fig. 6.1 shows the reputation histogram before and after removing the wholesale coffee maker sellers. The items/entries after whole sale sellers removal are the ones used in the real survey. We can see that the items are distributed well and are not clustered.

6.2 Real User Test Results

We conducted the real user test using the survey generated by Algorithm 6.1. 21 subjects have gone through the survey. To cover most of items in eBay, we used fixed price and reputation ranges $P \in [0, 10^5]$ and $R \in [0, 10^6]$ for normalization. To verify that IC-based method is insensitive to this normalization method, we calculated the averaged variance of the accuracy performances based on different price and reputation ranges. For each survey, we proceeded the test results 100 times by using any two price range and reputation range combination from $[0, p_{MAX}], \ldots, [0, 10p_{MAX}]$ and $[0, r_{MAX}], \ldots, [0, 10r_{MAX}]$, where p_{MAX} (r_{MAX}) is the maximum item price (reputation) in the survey. The results are in Table 6.4, which shows very small changes of accuracy. This validates the use of the above fixed normalization method.



(a) All Sellers (b) 3 Wholesale Sellers Removed

Fig. 6.1. Reputation Histogram of 15 Groups of Cuisine Coffee Maker Sellers

TABLE 6.4
VARIANCE OF IC-BASED ACCURACY

	Coffee Maker	Itouch 5th generation	SLR Camera
Ranking Quality	2.87E-4	7.26E-5	1.15E-4
Success Rate	1.17E-3	1.10E-4	9.04E-4

Table 6.5 and Table 6.6 show the basic performance results. The accuracy was calculated per subject and per product, and then averaged through all subjects for each product. The average accuracy of ranking quality in Table 6.5 shows that the proposed work always performs better than MAPS, and large improvements can be seen from coffee maker and SLR camera. Although the improvement is smaller on Itouch, the smaller variance from the proposed work indicates more stability of the prediction accuracy. Table 6.6 shows the success rate of predicting the best choice ranked first. With only $4 - 6$ items in a group, ranking the best choice as the first one is much more important than other ranking positions. From the results we can see that IC-based method can always rank the best choice first in more than half of the times. This is a critical improvement since this will greatly improve online user experiences. Furthermore, there are obvious improvements in both averaged results and variances, indicating both better prediction accuracy and

TABLE 6.5
AVERAGED RANKING QUALITY OF DIFFERENT PRODUCTS

	Coffee Maker		Itouch	
	Ave	Var	Ave	Var
IC-based	78.47%	1.02E-2	75.13%	1.92E-2
MAPS	66.54%	0.82E-2	71.34%	2.60E-2
Difference	11.93%	0.20E-2	3.79%	-0.68E-2

	SLR Camera	
	Ave	Var
IC-based	77.35%	0.55E-2
MAPS	64.79%	1.03E-2
Difference	12.56%	-0.48E-2

TABLE 6.6
AVERAGED SUCCESS RATE OF DIFFERENT PRODUCTS

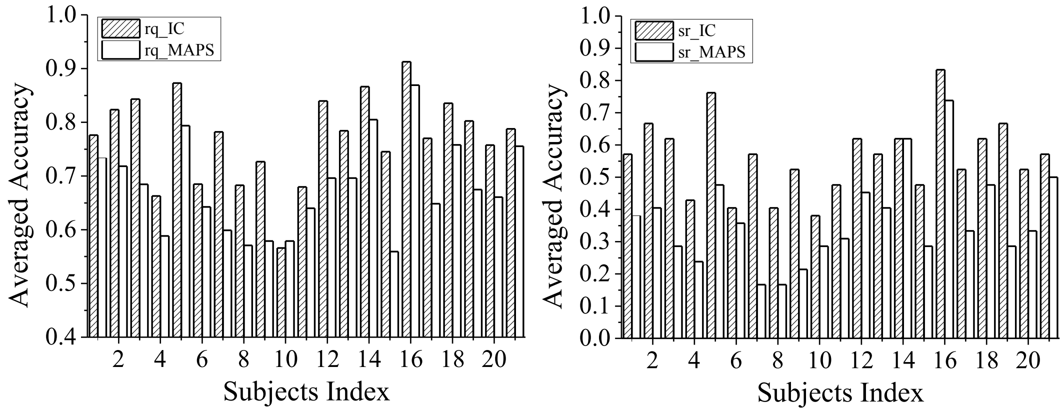
	Coffee Maker		Itouch	
	Ave	Var	Ave	Var
IC-based	56.46%	2.55E-2	57.14%	3.21E-2
MAPS	29.93%	3.04E-2	45.92%	7.22E-2
Difference	26.53%	-0.49E-2	11.22%	-4.01E-2

	SLR Camera	
	Ave	Var
IC-based	55.36%	1.42E-2
MAPS	33.93%	2.17E-2
Difference	21.43%	-0.75E-2

better prediction stability.

Fig. 6.2 shows per-subject accuracy averaged through the three products. Each bar is the averaged accuracy for each subject through the whole survey. It is obvious that IC-based method can always perform better than MAPS, except one subject who obtained similar prediction performance from both methods in both ranking quality and success rate.

Fig. 6.3 shows the impact of history length. It is obvious that both IC-based method and MAPS improve fast in accuracy at the beginning, and could gain stable accuracy with more than 3 historical records. This means both algorithms do not need many historical records to reach stable performance. This feature is



(a) Ranking quality (b)Success rate

Fig. 6.2. Averaged accuracy per subject

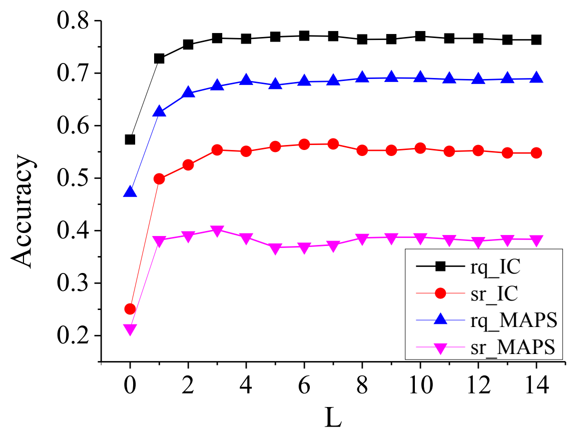


Fig. 6.3. Impact of history length

essential when dealing with a very sparse data environment in practice, e.g., users' purchasing records are far less than the number of products in online markets. The improvement of ranking quality from MAPS in Fig. 6.3 keeps stable with the growth of history length. On the other hand, two success rates are almost the same at the beginning and IC-based method improves much more with the first historical record. In other words, IC-based method uses the user purchasing information more thoroughly and more efficiently, so that it could gain much more user preferences from the history even from only one historical record and make better predictions.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

With the thriving user activities online, a dedicated investigation of user behavior becomes more important in the user oriented ubiquitous network, since user preference will lead the trend of businesses that want to trigger more purchasing behavior. Specifically, understanding the relationship between user preference and user behaviors can impact the design of online business model as well as provide sellers with guidelines to build optimized pricing strategy. Implicit user behavior analysis for online shopping platform is not widely studied due to the lack of quantized information that can be provided. However, compared to explicit user behavior which more previous researches relied on, the implicit user behavior could provide greater amount of information and it is more feasible and more reliable.

As discussed in the thesis, content-based (CB) approaches have difficulties in dealing with multi-attribute balancing, and collaborative filtering (CF) approaches do not have explicit multi-attribute model, which causes bias recommendation under dynamically changing market and users. Besides, the traditional recommender systems only predict for missing data, which is not suitable for online purchasing scenario that takes every candidate into account every time. The cardinal model used in previous research works also introduces information distortion during the time of translating user actions to limited range of scaled ratings. MCRS provides multi-attribute balancing through MCDM, but the cardinal functions used have independence restriction on attributes that causes convex hull problem. MAPS used the ordinal visual angle approach to avoid above issues, and took inter-item competition into account. However, the dimension reduction causes information loss which leads to the violation of monotonicity assumption and prevents the approach from further study.

In this work, we proposed a novel multi-attribute probabilistic ranking framework, which uses indifference curves to address multi-attribute balancing and to model users' personal preference. The novel ordinal model offers a flexible way to explicitly model multi-attribute balancing without the independence restriction on attributes. Thus, it solves the issues mentioned above and eliminates information bias/loss and avoids convex hull problem. The model also considers inter-item competition globally. In addition, the proposed framework also considers the limitation of human beings and models the user's decision process as a two-step selection process. Furthermore, according to simulation and real user test results, the proposed algorithm significantly outperforms existing multi-criteria ranking algorithms by achieving higher ranking accuracy with short learning curve.

7.2 Future Work

The model proposed in this thesis could then be utilized by individual sellers. One of the consequences of the thriving e-commercial environment is the high competition among online sellers. Online sellers are always overwhelmed with enormous amount of information and easily lose their awareness of factors that affect their business success. To solve this problem, services emerged to provide automated summarized reports to help sellers with their business management [51], [52]. Examples include providing demographic information, summary of purchasing time, frequency, past sales, etc. Advanced services include automatic price monitoring and price adjustment under sellers' settings [53]. However, no technique provides integrated information regarding the relationship between sales performance and sellers' profiles, which could directly help individual sellers strategically improve their competitiveness in online market.

In the research area, increasing number of researchers have been studying sales prediction of certain product/service based on past knowledge[54]. Their objectives are to forecast in advance and provide valuable intelligence for vendors' business, such as helping lower inventory level as well as satisfying demand just-in-time[54], [55]. Existing research efforts cover a wide range of offline and online businesses with/without considering user behavior. The work in [56] focused on forecasting fresh food sales in convenience stores by comparing basic machine learning methods and time series analysis. The work in [55] examined the impact of quality of reviews on the sales prediction based on IMDB movie database. The work in [57] built a trigger model to combine several data mining techniques to achieve higher accuracy for online books sales prediction. The work in [58] executed books sales prediction and observed high correlations between sales and searching/browsing user behaviors on some book categories. The work in [59] observed that user behaviors, such as calling number and the relevant search engine query data, are positive factors on air ticket sales.

However, previous works only examine factors (season, fashion, weather, holidays, cultural, religious, etc.) that affect the total demand of certain product rather than sales performance of individual sellers. To provide personalized service for individual sellers, the relationships between sale performances and sellers' profiles (e.g., prices, service reputations, locations, shipping time and cost, etc.) should be explicitly modeled. In this way, sellers will know their customers better in terms of multi-attribute sellers' profile, and react fast and strategically to remain competitive and profitable. Although some researches tried to predict with online user behaviors, they only considered behaviors that are directly linked with sales performances, such as searching/browsing or sentiment in the reviews. There is no explicit model which quantifies the reason for sales success.

Based on the completed work in this thesis, the future work will utilize the ordinal approach and the learned EMRs on each recorded item. Different from the completed work that predicts for individual users, the future work will predict for individual sellers who interact with lots of users. An integrative model is needed to combine EMRs from multiple users. Since the accuracy of EMRs on items is different and is dependent on item locations on the multi-attribute plane, to identify EMRs' accuracy could be an important issue in the future work.

References

[1] R. S. Pindyck and D. L. Rubinfeld, *Microeconomics*. Prentice Hall, 2012.

[2] S. Brin and L. Page, “Reprint of: The anatomy of a large-scale hypertextual web search engine,” *Computer networks*, vol. 56, no. 18, pp. 3825–3833, 2012.

[3] W. Kießling, M. Endres, and F. Wenzel, “The preference sql system-an overview.” *IEEE Data Eng. Bull.*, vol. 34, no. 2, pp. 11–18, 2011.

[4] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.

[5] F. Ricci, L. Rokach, and B. Shapira, *Introduction to recommender systems handbook*. Springer, 2011.

[6] P. Melville and V. Sindhwani, “Recommender systems,” in *Encyclopedia of machine learning*. Springer, 2011, pp. 829–838.

[7] J. O’Donovan and B. Smyth, “Trust in recommender systems,” in *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM, 2005, pp. 167–174.

[8] R. Reshma, G. Ambikesh, and P. S. Thilagam, “Alleviating data sparsity and cold start in recommender systems using social behaviour,” in *International Conference on Recent Trends in Information Technology (ICRTIT)*, 2016, pp. 1–8.

[9] P. Yu, L. Lin, and J. Wang, “A novel framework to alleviate the sparsity problem in context-aware recommender systems,” *New Review of Hypermedia and Multimedia*, pp. 1–18, 2016.

[10] S. Agarwal, J. Wills, L. Cayton, G. Lanckriet, D. J. Kriegman, and S. Belongie, “Generalized non-metric multidimensional scaling,” in *International Conference on Artificial Intelligence and Statistics*, 2007, pp. 11–18.

[11] M. Schultz and T. Joachims, “Learning a distance metric from relative comparisons,” *Advances in neural information processing systems (NIPS)*, p. 41, 2004.

[12] Y. Koren and J. Sill, “Ordrec: an ordinal model for predicting personalized item rating distributions,” in *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 2011, pp. 117–124.

[13] J. A. Konstan, “Introduction to recommender systems: Algorithms and evaluation,” *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 1–4, 2004.

[14] P. Resnick and H. R. Varian, “Recommender systems,” *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, 1997.

[15] M. Balabanović and Y. Shoham, “Fab: content-based, collaborative recommendation,” *Communications of the ACM*, vol. 40, no. 3, pp. 66–72, 1997.

[16] L. Si and R. Jin, “Flexible mixture model for collaborative filtering,” in *ICML*, vol. 3, 2003, pp. 704–711.

[17] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, “An algorithmic framework for performing collaborative filtering,” in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999, pp. 230–237.

[18] J. C. Butler, J. S. Dyer, J. Jia, and K. Tomak, "Enabling e-transactions with multi-attribute preference models," *European Journal of Operational Research*, vol. 186, no. 2, pp. 748–765, 2008.

[19] R. Burke, "Hybrid recommender systems: Survey and experiments," *User modeling and user-adapted interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[20] Y. Shi, M. Larson, and A. Hanjalic, "Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges," *ACM Computing Surveys (CSUR)*, vol. 47, no. 1, p. 3, 2014.

[21] I. Guy and D. Carmel, "Social recommender systems," *Int. Conf. Companion World Wide Web*, pp. 283–284, 2011.

[22] I. King, M. R. Lyu, and H. Ma, "Introduction to social recommendation," *Int. Conf. World Wide Web*, pp. 1355–1356, 2010.

[23] R. Zafarani, M. A. Abbasi, and H. Liu, *Social media mining: an introduction*. Cambridge University Press, 2014.

[24] P. Massa and P. Avesani, "Trust-aware collaborative filtering for recommender systems," in *OTM Confederated International Conferences on On the Move to Meaningful Internet Systems*. Springer, 2004, pp. 492–508.

[25] ——, "Trust-aware recommender systems," in *Proceedings of the 2007 ACM conference on Recommender systems*. ACM, 2007, pp. 17–24.

[26] C. Birtolo and D. Ronca, "Advances in clustering collaborative filtering by means of fuzzy c-means and trust," *Expert Systems with Applications*, vol. 40, no. 17, pp. 6997–7009, 2013.

[27] Y. A. Kim and R. Phalak, "A trust prediction framework in rating-based experience sharing social networks without a web of trust," *Information Sciences*, vol. 191, pp. 128–145, 2012.

[28] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, “Recommender systems with social regularization,” in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 287–296.

[29] R. West, H. S. Paskov, J. Leskovec, and C. Potts, “Exploiting social network structure for person-to-person sentiment analysis,” *arXiv preprint arXiv:1409.2450*, 2014.

[30] S. Rosenthal, P. Nakov, S. Kiritchenko, S. M. Mohammad, A. Ritter, and V. Stoyanov, “Semeval-2015 task 10: Sentiment analysis in twitter,” in *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval*, 2015.

[31] D. H. Alahmadi and X.-J. Zeng, “Ists: Implicit social trust and sentiment based approach to recommender systems,” *Expert Systems with Applications*, vol. 42, no. 22, pp. 8840–8849, 2015.

[32] B. Roy, *Multicriteria methodology for decision aiding*. Springer Science & Business Media, 2013, vol. 12.

[33] G. Adomavicius, N. Manouselis, and Y. O. Kwon, “Multi-criteria recommender systems,” *Recommender Systems Handbook*, pp. 769–803, 2011.

[34] G. Adomavicius and Y. O. Kwon, “New recommendation techniques for multicriteria rating systems,” *IEEE Intelligent Systems*, pp. 48–55, 2007.

[35] R. L. Keeney and H. Raiffa, “Decision with multiple objectives,” 1976.

[36] J. S. Dyer and R. K. Sarin, “Measurable multiattribute value functions,” *Operations research*, vol. 27, no. 4, pp. 810–822, 1979.

[37] Q. Feng, L. Liu, Y. Sun, T. Yu, and Y. Dai, “Enhancing personalized ranking quality through multidimensional modeling of inter-item competition,” *IEEE Int. Conf. Collaborative Computing: Networking, Applications and Worksharing*, pp. 1–10, 2010.

[38] N. Sahoo, R. Krishnan, G. Duncan, and J. P. Callan, "Collaborative filtering with multi-component rating for recommender systems," in *Proceedings of the sixteenth workshop on information technologies and systems*, 2006.

[39] Z. Fang, L. Zhang, and K. Chen, "A behavior mining based hybrid recommender system," in *Proceedings of IEEE International Conference on Big Data Analysis (ICBDA)*, 2016, pp. 1–5.

[40] L. Duo and J. Su, "A recommender system based on contextual information of click and purchase data to items for e-commerce," in *Third International Conference on Cyberspace Technology (CCT 2015)*, 2015, pp. 1–6.

[41] M. Wan, D. Wang, M. Goldman, M. Taddy, J. Rao, J. Liu, D. Lymberopoulos, and J. McAuley, "Modeling consumer preferences and price sensitivities from large-scale grocery shopping transaction logs," in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, pp. 1103–1112.

[42] H. A. Simon, *Models of Man: Social and Rational*. Wiley, 1957.

[43] G. Gigerenzer and R. Selten, *Bounded rationality: The adaptative toolbox*. MIT Press, 2002.

[44] G. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological review*, vol. 63, no. 2, p. 81, 1956.

[45] S. Borzsony, D. D. Kossmann, and K. Stocker, "The skyline operator," *IEEE Int. Conf. Data Engr.*, pp. 421–430, 2001.

[46] H. R. Varian, *Intermediate Microeconomics: A Modern Approach (Eighth Edition)*. W.W. Norton Co., 2009.

[47] U. Junker, "Preference-based search and multi-criteria optimization," *National Conf. Artificial Intelligentce*, pp. 34–40, 2002.

[48] P. Viappiani, B. Faltings, and P. Pu, "The lookahead principle for preference elicitation: Experimental results," *Flexible Query Answering Systems*, pp. 378–389, 2006.

[49] M. Woolf, "A statistical analysis of 1.2 million amazon reviews," 2014, [Online]. Available: http://minimaxir.com/ 2014/06/reviewing-reviews/. [Accessed: 02- Nov- 2015]. [Online]. Available: http://minimaxir.com/2014/06/reviewing-reviews/

[50] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[51] A. G. H. Ltd., "Shu ju mo fang," http://mofang.taobao.com/, 2015.

[52] C. Dawson and D. Wilson, "ebay & ecommerce tools & services guide 2012," 2012.

[53] C. Corp., "Repricer," http://www.channeladvisor.com/, 2015.

[54] W.-I. Lee, B.-Y. Shih, and C.-Y. Chen, "A hybrid artificial intelligence sales-forecasting system in the convenience store industry," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 22, no. 3, pp. 188–196, 2012.

[55] X. Yu, Y. Liu, X. Huang, and A. An, "A quality-aware model for sales prediction using reviews," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 1217–1218.

[56] C.-Y. Chen, W.-I. Lee, H.-M. Kuo, C.-W. Chen, and K.-H. Chen, "The study of a forecasting sales model for fresh food," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7696–7702, 2010.

[57] W. Huang, Q. Zhang, W. Xu, H. Fu, M. Wang, and X. Liang, "A novel trigger model for sales prediction with data mining techniques," *Data Science Journal*, vol. 14, 2015.

[58] H. Yuan, W. Xu, and M. Wang, “Can online user behavior improve the performance of sales prediction in e-commerce?” in *proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2014, pp. 2347–2352.

[59] H. Yuan, W. Xu, and C. Yang, “A user behavior-based ticket sales prediction using data mining tools: An empirical study in an ota company,” in *Proceedings of 11th IEEE International Conference on Service Systems and Service Management (ICSSSM)*, 2014, pp. 1–6.