Computers & Operations Research 39 (2012) 1837–1849

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

# A strategy-oriented operation module for recommender systems in E-commerce

Hsiao-Fan Wang [*], Cheng-Ting Wu

*Department of Industrial Engineering and Engineering Management, National Tsing Hua University, No. 101, Section 2 Kuang Fu Road, Hsinchu, Taiwan 30013, ROC*

ARTICLE INFO

Available online 1 April 2010

*Keywords:*
Electronic commerce
Recommender system
Marketing strategy
Clique-effects collaborative filtering

ABSTRACT

Electronic commerce (EC) has become an important support for business and is regarded as an efficient system that connects suppliers with online users. Among the applications of EC, a recommender system (RS) is undoubtedly a popular issue to make the best recommendation to the users. Even if many approaches have been proposed to perfect the recommendation, a comprehensive module comprising of essential sub-modules of input profiles, a recommendation scheme, and an output interface of recommendations in the RS is still lacking. Besides, the fundamental issue of profit consideration for an EC company is not stressed in general terms. Therefore, this study aims to construct an RS with a strategy-oriented operation module regarding the above aspects; and with this module, an approach named clique-effects collaborative filtering (*CECF*) for predicting the consumer's purchase behavior was proposed. Finally, we applied our proposed module to a 3C retailer in Taiwan, and promising results were obtained.

*Scope and Purpose:* This study aims to construct a comprehensive module for the recommender systems. The proposed strategy-oriented operation module comprises the essential parts of a recommender system. By utilizing the proposed module with marketing strategies and an effective on-line interface scheme, the recommender system could emphasize not only the customer's satisfaction as conventional recommender system suggested, but also the supplier's profit which shall be an important issue to an E-commerce company. Thus, a better recommendation environment could be displayed.

## 1. Introduction

Electronic commerce (EC) has been widely used by online users to perform different daily activities through the Internet. Online shopping is one of the popular applications among these activities. Instead of conventional shopping, EC provides alternative ways for users to get information on products such as price, availability, suppliers, substitutes, and even manufacturing process [39,54]. For competitiveness, EC companies need to develop higher business interoperability on their electronic market places by improving the electronic market functions [52,53]. The enhancement of electronic market functions could lead to an overall reduction of interaction cost for business interoperation on all types of electronic market places [15]. However, among the numerous EC functions which provide so much available information, it is difficult for online users to make quick and effective decisions [48]. Facing fierce market competition and impatient users, a personalized decision support system is urgent and essential for an EC company. By providing more helpful information to users, faster and more satisfactory decisions can be made; and thus, opportunities of retaining customers and gaining profits are higher.

Many EC suppliers use the recommender systems (RSs) to find out the preferences of target users so that the right products can be suggested [45]. A well-established RS can add value to an EC company in several ways—(1) users can retrieve product information easily, (2) cross-selling for users can be enhanced, and (3) users' loyalty can be sustained by good service. There are numerous studies in the fields of social networks [34] and information filtering techniques [42]. In social networks, people with similar characteristics tend to associate with each other. The use of social network structure generally allows the EC to identify the products of likely interest to the target users based on some information provided by the members of the network [19,28]. On the other hand, information filtering techniques that analyze users' preferences and help EC Web sites achieve accurate product selection. By filtering the information provided by the users, the techniques aim to track the purchase behavior of users and recommend proper products. Among information filtering techniques, collaborative filtering (*CF*) [25,45,46] is one of the

[*] Corresponding author. Tel.: +886 3 5742654x42654; fax: +886 3 5722685.
*E-mail address:* hfwang@ie.nthu.edu.tw (H.-F. Wang).

0305-0548/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2010.03.011

most commonly adopted method. The concept of the *CF* is much related to the social network. The *CF* technique uses collaborative information from ''neighbors,'' which are defined as users with similar behavior to the target user. *CF* is also regarded as the most effective method for the RS. However, *CF*'s drawback is that no recommendation could be made if a user's related data are sparse [26]. On the other hand, excessive emphasis on recommendation performance could lead to the neglect of the profit, which is also an essential concern for an EC company. Aside from this, although there are different approaches to retrieve the needed information for recommendation, a systematic and comprehensive decision module is still lacking. Therefore, the time spent on data retrieval can be long, and the recommended products may not match the users' desires. In particular, without a structural module, documenting the recommending procedure becomes difficult, and achieving the goal of ''the right goods for the right person'' becomes impossible.

With these concerns, we aim to propose a strategy-oriented operation module that could be comprehensively applied to EC Web sites as a decision support mechanism so that the choice of various marketing strategies that consider profit for both suppliers and users can be developed. In addition, under the framework of the proposed recommender module, we also propose a clique-effects collaborative filtering (*CECF*) technique to predict users' purchase behavior.

In particular, this paper presents the modeling perspective to the e-service system i.e. the recommender system. The proposed RS module aims to fulfill the profits of the customers and suppliers; the final stage of product selection is described as a linear bi-objective model, of which all required arguments are derived from the offline database and the *CECF*.

The paper is organized as follows. Section 2 discusses the literature related to the framework, issues, and the further development of an RS. The strategy-oriented operation module applied to an RS will be developed along with the proposed *CECF* in Section 3. Then we apply our proposed RS to a 3C retailer as a case study in Section 4. Finally, concluding remarks are given in Section 5, with suggestions on further research.

## 2. Literature review of the infrastructure of recommender systems

Schafer et al. [45,46] and Montaner et al. [37] have investigated the infrastructure of an RS in the framework of three sub-modules: (1) *input sources* of the users' profiles, (2) *output of recommendations*, and (3) *recommendation methods* as the interface between the two. In this section, we shall briefly review the current developments with respect to these three sub-modules.

### 2.1. Input sources

Usually, input sources include users' individual profiles which could be used to gather preferences for specific items, item attributes, ratings, and keywords or even purchase history [46]. Schafer et al. have classified input sources into two types [46]: (1) single users' profiles—the preferences of the target user for whom we are recommending, and (2) communities' opinions as an input regarding the general community of other users, that is, the target user is represented by the community. The two types of inputs allow the RS to make suggestions for different reasons.

For a target user, the individual profiles are inputted to the recommender agent to provide personalized information, whereas the input profiles of the community are fed into the RS to reflect opinions from multiple individuals as a whole. Therefore, these two types can be applied at different levels of personalization. In particular, the community's opinions as input are helpful in reinforcing or complementing information that can be retrieved from single user's profiles. This could be specified by the well-known issue of the ''new user'' problem, which is one of the cases in the ''ramp-up'' problem [27]. Recommendation for new users faces the challenge that the neighbors are hard to identify in a start-up company since the new users' profiles are lacking. When this phenomenon is translated into a user–item relation matrix, the matrix will be sparse. In particular, if a highly dimensional database is developed for an RS, the problem of identifying neighborhood becomes severe from the sparse user–item relation matrix. In order to solve the problems of sparse data or missing values, many approaches based on *CF* have been proposed. The issues of sparse matrix or missing values are often tackled with dimensionality reduction techniques [7,14,24,43]. Several dimensionality reduction techniques have been developed and applied to Jester, Movielens and EachMovie datasets. And in Eigentaste, Goldberg et al. [14] divided the recommendation process into two stages: online and offline operations. In the offline stage, the authors exploited the principal component analysis (PCA) to facilitate dimensionality reduction so that user's profiles which are formed through rating the gauge set are projected into an eigen plane. Consequently, in the online stage, the target user is asked to rate the gauge set to receive recommendations.

An alternative approach to estimate the missing values and to reduce the dimensionality of user–item relation matrix is the method of singular value decomposition (SVD), which has been exploited by Sarwar et al. [43]. SVD appears to be a common method for matrix factorization that results in the best lower rank approximations of the user–item relation matrix; however, Sarwar et al. suggested that the SVD-based method would yield better results in dense datasets of which a start-up company does not possess. Kim and Yum [24] further suggested an evolved PCA-iterative method, in which SVD is performed iteratively to improve the accuracy of imputed values based on prior results.

Nevertheless, to accommodate the dimensionality reduction to the recommendation process, the new user usually requires to rate on the specifically designated item set, for example, the gauge set, which could contain items that the new user never knows; besides, the size of designated item set should also be carefully controlled in case of driving the impatient customers out of the system. As indicated by Herlocker et al. [18] and Linden et al. [31], using PCA- or SVD-based techniques for dimensionality reduction would cause a lower recommendation quality since recommendations for items are more restricted to specific subjects; examining a small user sample such as the gauge set, the chosen neighborhoods are less similar with the target user. Moreover, Bell et al. [5] argued that the methods using imputed ratings, which significantly outnumber the original ratings, rely on imputation risk; and such risk would distort the data due to inaccurate imputation.

To realize a user's purchase behavior, the information revealed by a user's profiles is often investigated. Generally, there are two kinds of user's profiles that are commonly searched and collected. These are the user's ratings [45] and market basket data [35]. User's ratings refer to the scores given to item attributes by a user, and the user's ratings are often analyzed to define preference. On the other hand, market basket data contain a user's purchase history and probably demographic features. Specifically, each item presented in a user's basket data could either be "0" or "1" to denote whether an item is purchased , "1", or not, "0". There are always a number of transactional data in the market baskets; hence, management of these input profiles should be easier to maintain and retrieve.

The usual techniques used to maintain user's profiles are the history-based model [37] and the vector space model [11,40]. A history-based model lists purchase records, navigation history, or the contents of e-mail boxes to define users' profiles. In the vector space model, items are represented with a vector of features or attributes, usually words or concepts (such as a binary column to denote the purchased state or a column to denote the attributive value of an item), with an associated value. The vector space model is more efficient for computation, so it is often used for large amounts of data. For this reason, it is also the model adopted in this paper to maintain the database.

### 2.2. *Output of recommendations*

In general, the output is a suggestion of product(s) containing information on item type, quantity, and appearance [46]. The simplest form of a suggestion is the recommendation of a single item. A single item increases the chance that a user will seriously consider it desirable. More commonly, an RS provide an ordered or unordered recommendation list for a user [38]. Some advertising strategies can also be embedded in the recommendation, displaying bundled items, which could help enhance cross-selling and up-selling. By comparing bundled items with a recommendation list, bundled items may include products that are not related to the users since they are generated for promotions. In contrast, a recommendation list shows a set of products that satisfies users' preferences to a certain degree.

### 2.3. *Recommendation methods*

Recommendation methods are concerned with the accuracy and efficiency of prediction and presentation of the recommended items according to users' input sources. For an RS, it is critical to know users' preferences systematically. An essential concept is to use a relational database which is constructed offline. Then by mapping a new user to the database, a product that has been purchased by the same type of historical users can easily be picked up for the target user [29].

Clustering analysis is the technique that groups users/items with similar characteristics/properties into one group. By clustering, the search dimensionality can be reduced which speeds up the mapping process. A wide range of applications have been implemented by clustering techniques, and one of these is used to predict unknown users based on the group they belong to [49]. By analyzing the properties of the groups, we can learn about the characteristics of new users by identifying the group they belong to and thus provide them with the items that the same group has mostly bought. Besides, clustering analysis is also a very useful tool for looking for the ''neighbors'' in the information filtering technology. That is, the users called the *neighbors* are chosen by certain methods, such as clustering techniques, to support the prediction [6].

Information filtering technology has the ability to define user preferences with little effort. It is divided into two main categories [26]—collaborative filtering (*CF*) and content-based filtering (*CBF*). *CF* is the most popular approach to predict the probability that a user will purchase a specific item based on other users' preferences [21]. A *CF* method functions by matching people with similar interests and then making recommendations. However, in the initial state of an RS, the main problem would be insufficient users' profiles sustain the prediction basis while using *CF*. Consequently, the drawback of *CF* is the requirement of some relevant rating data given by the target user. Usually, by clustering users into groups before predicting, group influences could be utilized by recommendation methods on the target user to prevent poor prediction due to rarely relevant information [44]. Furthermore, because the conventional *CF* approach utilizes preferences of neighbors to make a prediction for a target user, it leaves additional influences of non-neighbors out of consideration. As a result, research tends to discriminate the impacts of neighbors from non-neighbors [23]; by integrating the effects caused by the two sources, better performance could be also expected.

*CBF* is the technology of analysis based on terms in the content such as texts or documents on the Web site. It considers term frequency in the content and its relation to the user's preference. However, with other media such as music or movies, its performance is not as good as text content because these objects are not easily indexed. In addition, the maintenance of numerous heterogeneous electronic product catalogues on the Internet is still a tough task [16]. Nevertheless, *CF* is still most commonly used since it is flexible and easily adaptable to an EC's RS [7]. Therefore, in this paper, we would incorporate the concept of *CF* into our system as the basic recommendation mechanism.

In addition to *CF* and *CBF*, another technique requires the private information of a user. *Demographic filtering* (*DF*) explains users by their personal demographics [17]. A *DF* approach uses descriptions of people to learn the probability that an item is most preferred by what type of persons. Therefore, this method would lead to the same recommendation if the users have similar personal data. However, the *DF* approach requires more information regarding a user's privacy; therefore, *DF* is confronted with the problem that it is not easy to collect users' demographic descriptions. Consequently, the *DF* method requires collaborating with other methods such as *CF* or *CBF* [37].

Besides the aforementioned filtering techniques, rules derived from the market basket analysis between items in large databases also account for an RS. Market basket analysis has been a popular system in finding the correlation among baskets [2,41]. One of the techniques is the famous *association rules* method, which was first introduced by Agrawal et al. [3]. Association rules have been used to find the pattern of the probability of buying a specific product when another product is purchased. In such a recommending environment, many rules have been developed on how the different purchase behaviors of users can be treated [20]. Therefore, Sarwar et al. also proposed a method of a*ssociation-rule based recommendation* (*ABR*) in 2000 [42]. However, for the huge amount of transaction data, there may be many biased rules that would affect the precision of the recommendation. Therefore, the market basket analysis shall be conducted with the aid of filtering techniques such as *CF*, and the common concept of the *CF* method adapted to the binary market basket data as proposed by Mild and Reutterer [36].

### 2.4. *Roles with their goals in a recommender system*

In the current RS, there are three common roles involved: the supplier, the system developer, and the user. In Table 1, we list possible considerations for constructing an RS. In the fields of EC trading, Li and Wang proposed a multi-agent-based model with a win-win negotiation approach of which the agents seek to strike a fair deal that also maximizes the payoff for everyone involved [30]. However, such kind of win-win negotiation mechanism has not been discussed in the RSs with more comprehensive scale. For the existing research, the ''performance of recommendation'' is an attribute that benefits users. Therefore, when ''more is better'' is stressed, only the number of sold products is maximized but not necessarily the profit. In other words, an RS is usually constructed from a user's standpoint. Only a few RSs could be regarded as built from a supplier's perspective. For instance, Liu and Shih developed

**Table 1**
Roles and resolution in recommender systems.

| | User | Supplier | System developer | |
|---|---|---|---|---|
| Objective | $O(u)$ | $O(s)$ | ***Win–win strategy***<br>$O(u)$ & $O(s)$ | ***Maximal profit strategy***<br>$O(s)$ |
| Constraints | $C(u)$ | $C(s)$ | $C(u)$ & $C(s)$ | $C(u)$ & $C(s)$ |
| Problem types | Maximization problem | Maximization problem | Multi-objective problem | Maximization problem |

*Note*: $(u)$: user; $(s)$: supplier.$O(u)$ Objective of the user: fulfill the demands of oneself, $O(s)$ objective of the supplier: maximize profit or products sold, $C(u)$ constraint of the user: budgets in hand and $C(s)$ constraint of the supplier: fulfill demands of users.

a weighted RFM-based method for an RS [32,33], where RFM means recency, frequency, and monetary; it considers the user's lifetime value which is helpful in extending market share in the long run. However, for an RS constructed from the viewpoint of system developer, issues should be considered that not only to fulfill the user's needs (preferences, budgets) but also to raise the supplier's profit. Changchien et al. discussed sales promotion based on businesses' marketing strategies, pricing strategies, and users' purchasing behavior, which could potentially be a win-win situation [9]. However, the study prioritized the probability of an inequitable supplier so that it may be difficult to keep a user's loyalty. Therefore, it is also necessary to construct an RS that allows both parties to justify their priorities.

### 2.5. Summary and discussion

From the brief review of the recent RSs, some aspects could be emphasized to improve an RS. First, it should be noted that so far, there is no complete manipulated module that supports all sub-modules of input module, output module, and recommendation interface in an RS. The researchers also realized that through quantitative measurement, the performance of the system can be better controlled and evaluated. This triggers our main goal in this study to develop an operation module for systematic analysis and general applications in an RS. From the viewpoint of managing an EC site and its RS, it is more robust and convenient if an analytical model comprising the three sub-modules can be imported to facilitate the product selection process. With this regard, developing a comprehensive module that can achieve the transparent requirements of the decision-support process and provide a good solution for recommendation purposes is necessary and would be presented in this study.

Second, we found merits and deficiencies in each of the existing recommendation approaches. Since RSs have different types of input sources such as user's ratings or market basket data, the corresponding recommendation method will be a key sub-module that determines the success of an RS. As the applications in *CF*, personal profiles of target users are first used to match their neighbors'; the purchase behaviors of the neighbors are then exploited to predict target users' choices. However, for an EC Web site that is a start-up or is selling products with high prices, it would be confronted with the problem that not enough basket data support the market basket analysis (dataset is sparse or with missing values); therefore, that recommendation performance would be very poor. Since the new user with few personalized information is difficult to categorize, the community's opinions could be adopted to complement the insufficient information. For a user whose personal profiles are already known, the community's opinions reinforce the user's identity [23].

With the above concerns, in this study, we propose a strategy-oriented operation module for the RS comprising (1) an offline database, (2) *CECF*, and (3) the analytical model. An offline database that could be mathematically supported for the RS is developed. The database consists of three parts—user-group data, item-group data, and the relations in between. The offline database is designed with the two characteristics: (1) the users and the items are classified into groups according to their respective features/attributes (see Sections 3.1.1 and 3.1.2). As suggested in the literatures, PCA- or SVD-based approaches may lose prediction accuracy due to excessively restricted dataset from which the neighborhood is formed. Thus we adopt the classification technique for dimensionality reduction. We regard any individual in a group as an information provider, which is especially important to a start-up RS with rare data, (2) the group effects are much easier to be retrieved. By bringing out additional effects from the groups of users and items, we aim to dilute the imprecise prediction caused by rare data, and to prevent inconsistent imputed data like average scores.

However, to avoid the imputed group effects predominating over prediction, the priority of group effects shall be well-arranged. Therefore, under the proposed offline database, we base on *CF* to propose a clique-effects approach, namely, *CECF*. With the scheme of adjustable weights between individual's and group's effects, *CECF* is likely helpful in solving the situation of sparse data and the so-called "ramp up" problem. In addition, we also introduce an analytical model proposed by Wang and Wu [51]. The analytical model could allow the system developer to actively adjust the priority between the supplier's profit and the user's satisfaction level. Therefore, in the next section, we shall propose the strategy-oriented operation module whose cores consist of *CECF* and the analytical model; the module aims to describe the recommendation process and provide better recommendation performance for the RS.

## 3. The proposed recommendation module

Based on the issues specified in Section 2.4 that an RS shall provide three roles to be switched and the summary in Section 2.5, we propose an RS (Fig. 1) with the recommendation module composed of three sub-modules—input, the recommendation method, and output. The input sub-module deals with the input profiles of a target user; the types of profiles considered in the system would be the demographic information, the binary basket data, and the target user's requests of the desired satisfaction level and budget limit. The output sub-module would provide the recommended items from the result of the recommendation method. Both input and output sub-modules are categorized into *online* operations. The recommendation method, which is the core

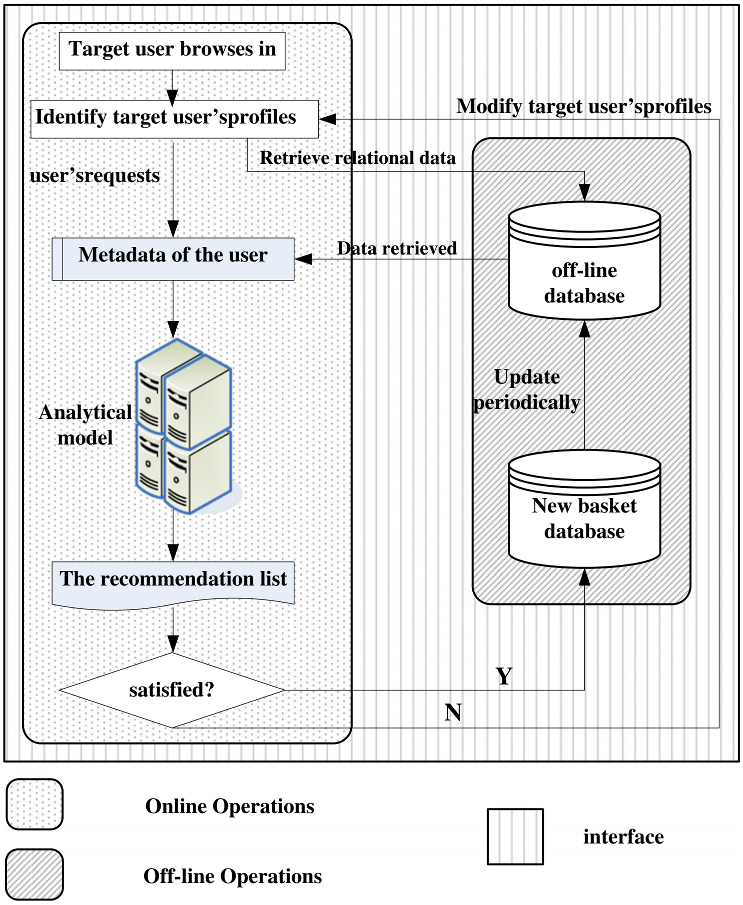

**Fig. 1.** The proposed recommendation module.

of the recommending module, functions with an *online* analytical model under the *offline* database constructed from three parts—user-group end, item-group end, and the relations in between. Exploiting the proposed *CECF* approach, the offline database provides required information retrieval of the target user's purchase probability measure on each item. The analytical model is then run by metadata composed of the target user's request and what has been retrieved from the offline database. In particular, the analytical model uses a bi-objective function that would allow choice between the *win–win strategy* and the *maximal profit strategy*, which were proposed by Wang and Wu [51]. The *win–win strategy* not only matches the user's taste but also enhances the supplier's profit, whereas the *maximal profit strategy* recommends products based on maximization of profit.

This section is organized as follows. First, we would specify the construction of the offline database including the user-group and item-group data. Then the proposed clique-effects approach based on *CF* (*CECF*) would be presented in Section 3.2. Finally, we would clarify online and offline operations as well as present the analytical model in Section 3.3.

**Table 2**
Classification rules when $K=3$.

| Class | Attribute labels |
|---|---|
| 1 | Non |
| 2 | $\{\alpha_1\}\backslash\{\alpha_2,\alpha_3\}$ |
| 3 | $\{\alpha_2\}\backslash\{\alpha_1,\alpha_3\}$ |
| 4 | $\{\alpha_3\}\backslash\{\alpha_1,\alpha_2\}$ |
| 5 | $\{\alpha_1,\alpha_2\}\backslash\{\alpha_3\}$ |
| 6 | $\{\alpha_1,\alpha_3\}\backslash\{\alpha_2\}$ |
| 7 | $\{\alpha_2,\alpha_3\}\backslash\{\alpha_1\}$ |
| 8 | |

## 3.1. *Offline operations*

In this section, we would specify the construction of the offline database including the user-groups data and item-groups data.

### 3.1.1. *Item-groups with their properties*

Let $D$ be the items in the market basket, with each item denoted as $p_d$, where $d=1,\ldots,D$. Define $\Psi_{p_d}=[\alpha_1,\alpha_2,\ldots,\alpha_k,\ldots,\alpha_K]_{p_d}$ to be an attribute vector of $p_d$, then the set of items in the database is $P=\{p_d(\alpha_k)|d=1,2,\ldots,D\}$. All items in the database are further classified into mutually exclusive item-groups as $P^i=\{p_{d^i}(\alpha_k)|d^i=1^i,2^i,\ldots,D^i,i=1,2,\ldots,I\}$, each with $|P^i|=D^i$, and thus $\bigcup_{i=1}^{I} P^i=P$ and $\sum_{i=1}^{I} D^i=D$. In particular, we classify the items with respect to the item attributes. A threshold of each attribute value is given; each item with specific attribute values above those thresholds will be assigned to the corresponding group. The number of attributes ($K$) would be referred with its power set and then $2^K$ item-groups are generated. For instance, in Table 2, the number of item-groups generated is 8 when $K$ is 3; an item would be distributed into Class 5 only if its attribute values in $\alpha_1$, $\alpha_2$ are higher than the thresholds of $\alpha_1$, $\alpha_2$ as well as its $\alpha_3$ value lower than the thresholds of $\alpha_3$ as denoted by $\{\alpha_1,\alpha_2\}\backslash\{\alpha_3\}$. Thus, the classification rules will provide exclusive groups so that one item belongs to only one group. Properties of each item-group could be also easily and clearly identified by observing attribute labels. The *selling prices* of items in the market basket are defined as $\mathbf{s}=[s_{d^i}|d^i=1^i,2^i,\ldots,D^i,i=1,2,\ldots,I]$; possible *profits* are defined as $\mathbf{c}=[c_{d^i}|d^i=1^i,2^i,\ldots,D^i,i=1,2,\ldots,I]$ where $s_{d^i}$ and $c_{d^i}$ represent the corresponding price and profit of $p_{d^i}$. Therefore, for the items database, it will be stored by each item-group with its items and specified properties.

### 3.1.2. User-groups with their profiles

Denote a user as $u_f$ with $f\in N$. Let $U=\{u_f(\omega_g)|f\in N\}$ be a set of users labeled by the demographic features $\omega_g\in\{\omega_1,\omega_2,\ldots,\omega_g,\ldots,\omega_G\}$. To facilitate analysis—providing solutions for the "new user" problem and exploiting clique effects, the users are classified into mutually exclusive user-groups and assumed to behave similarly as the *DF* method suggests. The user-groups are formed by the following rules: assume each demographic feature $\omega_g$ could be divided/categorized into $v_g$ intervals/categories denoted by $\omega_g^{v_g}$, and then we define $U^j: U\to\omega_1^{v_1}\times\omega_2^{v_2}\cdots\times\omega_G^{v_G}$, we have $U^j=\{u_f(\omega_g)|\omega_g\in\omega_g^{v_g},\ g=1,2,\ldots,G,j=1,2,\ldots,J\}$. Then each user-group could be represented as $U^j=\{u_{f^j}(\omega_g)|f^j=1^j,2^j,\ldots,\ F^j,j=1,2,\ldots,J\}$, $|U^j|=F^j$ and thus $\bigcup_{i=1}^{J}U^j=U$. For instance, we define the demographic features to be *gender* ($\omega_1$) and *age* ($\omega_2$); $\omega_1$ is categorized into $v_1=2$ categories as male and female; $\omega_2$ is divided into $v_2=4$ intervals as $(0,\ 20]$, $[20,30]$, $[30,40]$, $[40,\ \infty)$. Then we define the user-groups as $U^j: U\to\omega_1^2\times\omega_2^4$ and eight user-groups yield as $U^j,j=1,2,\ldots,8$.

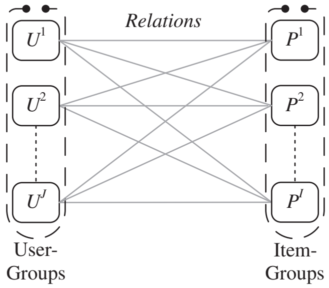

**Fig. 2.** Framework of relations among user-groups and item-groups.

### 3.2. Derivation of relations among users and items—CECF

In the proposed offline database, the framework of bipartite grouping connects users and items (Fig. 2). The relations embedded in the framework are regarded as clique effects of the purchase probability measured for a target user. The clique effects result mainly from the grouping of users. Users in the same clique with the target user (the so-called *neighbors* in *CF*) could provide collaborative information to measure purchase probability. However, users in different cliques may also provide collaborative information to the target user to a certain degree. In this respect, we propose the following concept to measure the purchase probability of the target user with respect to a predicted item as

$$P_{r(\text{user, item})}=\theta\cdot P^{\text{in-clique}}_{r(\text{user, item})}+(1-\theta)\cdot P^{\text{out-of-clique}}_{r(\text{user, item})},\tag{1}$$

where the probability $P_{r(\text{user, item})}$ is a convex combination of two distinct probabilities: one is the purchase probability predicted by collaboration of users in the same clique (the *neighbors*) with the target user, and the other is predicted by collaboration of users in the different cliques. The composition of the proposed probability measure is illustrated in Fig. 3.

Let us refer to Fig. 3. First, note that arrows 3 and 4 jointly represent the "in-clique" purchase probability measure used by conventional *CF*. The common concept of the *CF* method with adaptation to the binary market basket data [6,35] is presented as

$$P^{\text{in-clique}}_{r(u_{f^j},p_{d^i})}=\kappa_1\sum_{u_{f^\tau}\in U^j}sim(u_{f^j},u_{f^\tau})\times C_{u_{f^\tau},p_{d^i}},\tag{2}$$

where $P^{\text{in-clique}}_{r(u_{f^j},p_{d^i})}$ is the probability that target user $u_{f^j}$ purchases item $p_{d^i}$ by using a collaboration of neighbors' preferences; $\kappa_1$ is a normalizing factor to ensure the absolute values of probability sum to unity; $sim(u_{f^j},u_{f^\tau})$, which refers to arrow 4, is the similarity between the target user $u_{f^j}$ and the neighbors $u_{f^\tau}$; and $C_{u_{f^\tau},p_{d^i}}$, which refers to arrow 3, is the binary choice whether a user $u_{f^\tau}$ purchases $p_{d^i}$ or not. It is noteworthy that for the similarity measure between the target user $u_{f^j}$ and the neighbors $u_{f^\tau}$, as specified in Eq. (2), the neighbors are chosen from the user-group to which the target user belongs; this is in compliance with the structure of our proposed RS, which assumes that users in the same demographic group would tend to behave similarly.

Second, for the probability measure of "out-of-clique" based on the concept of *CF*, two factors should be considered: (1) the similarity between the target user-group and other user-groups as well as (2) other user-groups' purchase priorities on the predicted

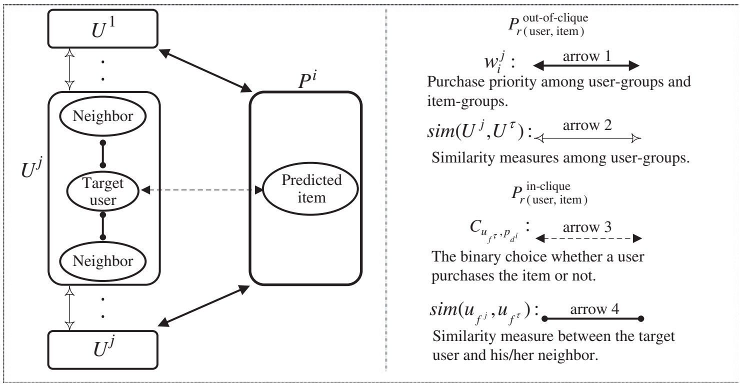
$U^1$
$P^i$
$U^j$
Neighbor
Target user
Neighbor
Predicted item
$U^j$
$P^{\text{out-of-clique}}_{r(\text{user, item})}$
$w_i^j$ :
arrow 1
Purchase priority among user-groups and item-groups.
$sim(U^j,U^\tau)$ :
arrow 2
Similarity measures among user-groups.
$P^{\text{in-clique}}_{r(\text{user, item})}$
$C_{u_{f^\tau},p_{d^i}}$ :
arrow 3
The binary choice whether a user purchases the item or not.
$sim(u_{f^j},u_{f^\tau})$ :
arrow 4
Similarity measure between the target user and his/her neighbor.

**Fig. 3.** Various probability measurements of the target user on the predicted item.

item-group. For the former, the similarity measures would refer to arrow 2 in Fig. 3. For the latter that refers to arrow 1 in Fig. 3, the relative purchase frequency in the binary basket analysis has been adopted as the prediction of purchase priority [10]:

$$w_i^j = \frac{C(U^j,P^i)}{S(U^j)}, \tag{3}$$

where $C(P^i,U^j)$ is the relative frequency that users in $U^j$ purchase items in $P^i$; $S(U^j)$ is the total number of market baskets for $U^j$. Therefore, the probability measure of "out-of-clique" purchase can be presented as

$$P^{\text{out-of-clique}}_{r(u_{fj},p_{d^i})} = \kappa_2 \sum_{\tau \neq j} sim(U^j,U^\tau) \times w_i^\tau, \tag{4}$$

where $sim(U^j,U^\tau)$, which refers to arrow 2, is the similarity measure between the target user-group $U^j$ and other user-group $U^\tau$; $\kappa_2$ is a normalizing factor to ensure the absolute values of probability sum to unity. Therefore, the probability measure of a target user $u_{fj}$ purchasing item $p_{d^i}$ would be represented as

$$\begin{aligned} P_{r(u_{fj},p_{d^i})} = \partial^{u_{fj}}_{p_{d^i}} = \theta \cdot \overbrace{\left(\kappa_1 \sum_{u_{f\tau} \in U^j} sim(u_{fj},u_{f\tau}) \times C_{u_{f\tau},p_{d^i}}\right)}^{\text{in-clique}} \\ +(1-\theta) \cdot \overbrace{\left(\kappa_2 \sum_{\tau \neq j} sim(U^j,U^\tau) \times w_i^\tau\right)}^{\text{out-of-clique}}, \end{aligned} \tag{5}$$

where the probability measure $P_{r(u_{fj},p_{d^i})}$ is replaced by $\partial^{u_{fj}}_{p_{d^i}}$ for simplicity; and $\theta$ is an adjustable weight on the in-clique probability measure. The way of the probability measure in Eq. (5) would lead us into the consideration on how to select similarity functions. Note that the *CF* performance depends on the choice of similarity measures. Conventionally, the similarity function for market basket data is based on the Jaccard coefficient [10,22,36] as

$$sim(u_{fj},u_{f^\tau}) = \frac{|S(u_{fj}) \cap S(u_{f^\tau})|}{|S(u_{fj}) \cup S(u_{f^\tau})|} = \frac{|S(u_{fj}) \cap S(u_{f^\tau})|}{|S(u_{fj})| + |S(u_{f^\tau})| - |S(u_{fj}) \cap S(u_{f^\tau})|}, \tag{6}$$

where $S(u_{fj})$ is the item set purchased by user $u_{fj}$; $S(u_{fj}) \cap S(u_{f^\tau})$ is the common item set purchased by user $u_{fj}$ and $u_{f^\tau}$; $S(u_{fj}) \cup S(u_{f^\tau})$ is the item set purchased by user $u_{fj}$ or $u_{f^\tau}$. However, as indicated in [36], the Jaccard coefficient missed the information that two users do not choose the same items simultaneously. The non-common item set would affect the similarity measure between two objects; as a result, the similarity function shall take the influence of non-common item set into consideration. Therefore, on the grounds of effects caused by non-common item set between users' purchase histories, we propose the similarity measure between two users based on the similarity function considering non-common item set as

$$sim(u_{fj},u_{f^\tau}) = \frac{|\overline{S}(u_{fj}) \cap \overline{S}(u_{f^\tau})|}{|\overline{S}(u_{fj}) \cup \overline{S}(u_{f^\tau})|}, \tag{7}$$

where $\overline{S}(\cdot)$ represents the non-purchased item set and the complement set of $S(\cdot)$. Consequently, Eq. (7) preserves the information of items that are not commonly purchased by two compared users.

However, the similarity measure of the non-common item set is not very appropriate in a large-scale database. The reason is that the value of this indicator would be probably close to one when comparing two users (see Appendix A). As a consequence, we suggest that they are compared on the grounds of group purchase behavior, which is given as

$$sim(U^j,U^\tau) = \frac{|\bigcup_{j=1}^{J} S(U^j) - (S(U^j) \cup S(U^\tau))|}{|\bigcup_{j=1}^{J} S(U^j) - (S(U^j) \cap S(U^\tau))|}, \tag{8}$$

where $sim(U^j,U^\tau)$ is the similarity measure between the target user-group $U^j$ and the other user-group $U^\tau$. Therefore, the similarity measures indicated in the Appendix A could be computed as shown in Appendix B, in which the similarity measure is more appropriate.

### 3.2.1. Summary of the proposed CECF

In this section, we have proposed the *CECF* containing users' purchase probability measure as Eq. (5), which is a convex combination of two distinct probability measures from in-clique effects of Eq. (2) and out-of-clique effects of Eq. (4). The classification of the target user into in-clique users as well as out-of-clique users, the proposed probability measure function provides different insight from that of conventional *CF* method.

As for the probability measure of in-clique users, we adopt the traditional *CF* method, whereas for the measure of out-of-clique users, we propose an alternative similarity function by incorporating the items not purchased simultaneously by each pair of compared users to find the similarity among user-groups. Then the proposed probability measure is predicted by the purchase and non-purchase behaviors of the users, which could be expected to provide more information in expounding the users. Therefore, to facilitate flexible applications, under the proposed *CECF*, we have two schemes in the recommendation method, namely, *CECF-C* and *CECF-NC*. *C* and *NC* represent the choice of similarity functions applied in computing the similarities among user-groups. *C* is based on the Common item set, whereas *NC* is based on *N*on-*C*ommon item set. It is worthy to discuss the hybrid of *C* and *NC* in measuring similarities among user-groups. We would not focus on a hybrid approach currently since the adjustment of weights would make the module more complex for analysis. Note that measuring similarities between in-clique users still apply the concept of common item set since their basket sizes are much smaller. In Table 3, we list all recommendation schemes that would be compared in Section 4.

## 3.3. The analytical model and recommendation procedures

In this section, we would discuss the analytical model proposed by Wang and Wu [51] as well as the operation procedures of the proposed module.

### 3.3.1. The analytical model with two marketing strategies: maximal profit strategy and win–win strategy

After the offline operations, three databases were constructed, namely item-group database defined by $P^i = \{p_{d^i}(\alpha_k) | d^i = 1^i, 2^i, \ldots, D^i, i = 1,2,\ldots,I\}$; user-group database defined by $U^j = \{u_{fj}(\omega_g) | f^j = 1^j, 2^j, \ldots, F^j, j = 1,2,\ldots,J\}$; their relations constructed by *CECF* of Eqs. (4)–(6), and (8). When a user is online, we could identify a user's preferences through the corresponding information retrieved from the databases. The retrieved data as well as the user's requests (desired satisfaction level and budget limit) are

**Table 3**
Recommendation schemes.

| Schemes | Function of user similarity<br>In-clique effects | Function of user-group similarity<br>Out-of-clique effects |
|---|---|---|
| *CF* | Common item set | – |
| *CECF-C* | Common item set | Common item set |
| *CECF-NC* | Common item set | Non-common item set |

regarded as the user's metadata input into the analytical model, which has been proposed by Wang and Wu as shown in Eqs. (9.1)–(9.5):

$$\text{Maximize} \quad \sum_{j=1}^{J} \sum_{f^j=1}^{F^j} \mathbf{c}\mathbf{x}^{f^j}, \tag{9.1}$$

Subject to

$$\mathbf{a}^{f^j}\mathbf{x}^{f^j} \geq b^{f^j}, \quad f^j = 1^j, 2^j, \ldots, F^j, \quad j = 1,2,\ldots J, \tag{9.2}$$

$$\mathbf{s}\mathbf{x}^{f^j} \leq B^{f^j}, \quad f^j = 1^j, 2^j, \ldots, F^j, \quad j = 1,2,\ldots J, \tag{9.3}$$

$$\sum_{i=1}^{I} \sum_{d^i=1}^{D^i} x_{id^i}^{f^j} \geq 1, \quad f^j = 1^j, 2^j, \ldots, F^j, \quad j = 1,2,\ldots J, \tag{9.4}$$

$$x_{id^i}^{f^j} \in \{0,1\}, \tag{9.5}$$

where $\mathbf{x}^{f^j} = [x_{d^i}^{f^j}]_{\sum_i D^i \times 1}$, $i = 1,2,\ldots,I$, $d^i \in \{1,2,\ldots,D^i\}$, $j = 1,2,\ldots J$, $f^j = 1^j, 2^j, \ldots, F^j$, $x_{d^i}^{f^j} = 1$ if item $p_{d^i}$ is recommended to $u^{f^j}$; otherwise, $x_{d^i}^{f^j} = 0$. **c** and **s** are the corresponding profit and price of $p_{d^i}$. $b^{f^j}$ is the satisfactory level requested by $u^{f^j}$; $B^{f^j}$ is the budget limit given by $u^{f^j}$. $\mathbf{a}^j = [\partial_{p_{d^i}}^{u_{f^j}}]_{1 \times \sum_i D^i}$, $\partial_{p_{d^i}}^{u_{f^j}}$ to be the purchase probability measure of user $u_{f^j}$ on $p_{d^i}$. This model maximizes the profits of an EC company (9.1) when the items recommended to users satisfy their satisfactory level as shown in constraint (9.2); the total prices spent on the items should not exceed the budget of the user as shown in constraint (9.3). Constraint (9.4) provides a tool for strategic uses by recommending different number of items of which at least one item should be recommended to a user at each time.

Under the basic model, two strategies could be provided for different marketing strategies—the *maximal profit strategy* and *win–win strategy*. When the recommending processes use only the supplier viewpoint, the goal will be to *maximize the profits* of the goods under a set of items that satisfy the users' preferences and budgets. When this is intended, denote the reduced decision-variable vector and the corresponding coefficients by "′" to mean that all items left for consideration are at least above the requested satisfactory level, namely $b^{f^j}$. Model (10) will immediately reflect such strategy.

$$\begin{aligned} &\text{Maximize} \quad \sum_{j=1}^{J} \sum_{f^j=1}^{F^j} \mathbf{c}'\mathbf{x}^{f^{j'}} \\ &\text{Subject to} \quad \mathbf{s}'\mathbf{x}^{f^{j'}} \leq B^{f^j}, \quad f^j = 1^j, 2^j, \ldots, F^j, \quad j = 1,2,\ldots J \\ &\sum_{i=1}^{I} \sum_{d^{i'}=1}^{D^i} x_{id^{i'}}^{f^j} \geq 1, \quad f^j = 1^j, 2^j, \ldots, F^j, \quad j = 1,2,\ldots J, \quad x_{id^{i'}}^{f^j} \in \{0,1\} \end{aligned} \tag{10}$$

Although *maximal profit strategy* will bring about the highest income to the suppliers, from the management viewpoint, it only passively satisfies users' desires to the minimal levels and thus is not a strategy to provide good services. Alternatively, the *win–win strategy* which actively takes both suppliers' profit and users' preferences into account is proposed. Model (11) realizes such strategy in which the first objective function maximizes the supplier's profit as previously done; meanwhile, the second objective function represents the maximization of the user's satisfaction. Model (11) is a bi-objective programming model. Since there are a lot of prominent literatures discussing and solving this kind of bi-criterion problems [1,4,8,13,50] we do not focus on how to solve the proposed models. In the manner of convex combination of the two objectives: introducing a weighting parameter $\beta$, $\beta \in [0,1]$, Model (11) can be transformed into a single objective programming model as Model (12). While Model (11) with $\beta = 1$ yields Model (10) for implementing *maximal profit strategy*; that with $\beta = 0$ will emphasize the users' benefit as *best service strategy*; and depend on the marketing preference the suppliers adopted, $\beta$ can be given by any values between 0 and 1 as *win–win strategy*. Note that in Model (12), $\mathbf{c}''$ is further normalized from $\mathbf{c}'$ into [0, 1] to match the same scale with $\mathbf{a}^{j'}$.

$$\begin{aligned} &\text{Maximize} \quad \sum_{j=1}^{J} \sum_{f^j=1}^{F^j} \mathbf{c}'\mathbf{x}^{f^{j'}} \\ &\text{Maximize} \quad \sum_{j=1}^{J} \sum_{f^j=1}^{F^j} \mathbf{a}^{j'}\mathbf{x}^{f^{j'}} \\ &\text{Subject to} \quad \mathbf{s}'\mathbf{x}^{f^{j'}} \leq B^{f^j}, \quad f^j = 1^j, 2^j, \ldots, F^j, j = 1,2,\ldots J \\ &\sum_{i=1}^{I} \sum_{d^{i'}=1}^{D^i} x_{id^{i'}}^{f^j} \geq 1, \quad f^j = 1^j, 2^j, \ldots, F^j, j = 1,2,\ldots J, x_{id^{i'}}^{f^j} \in \{0,1\} \end{aligned} \tag{11}$$

$$\begin{aligned} &\text{Maximize} \quad \beta\left(\sum_{j=1}^{J} \sum_{f^j=1}^{F^j} \mathbf{c}''\mathbf{x}^{f^{j'}}\right) + (1-\beta)\sum_{j=1}^{J} \sum_{f^j=1}^{F^j} \mathbf{a}^{j'}\mathbf{x}^{f^{j'}} \\ &\text{Subject to} \quad \mathbf{s}'\mathbf{x}^{f^{j'}} \leq B^{f^j}, \quad f^j = 1^j, 2^j, \ldots, F^j, \quad j = 1,2,\ldots J \\ &\sum_{i=1}^{I} \sum_{d^{i'}=1}^{D^i} x_{id^{i'}}^{f^j} \geq 1, \quad f^j = 1^j, 2^j, \ldots, F^j, j = 1,2,\ldots J, x_{id^{i'}}^{f^j} \in \{0,1\}, \end{aligned} \tag{12}$$

### 3.3.2. *Measures of recommendation performance*

To evaluate the performance of information retrieval, three measures of recall, precision, and F1 are usually employed [12,47]. They are defined as follows and will be used to evaluate our recommendation system as well.

$$\textbf{Recall} = |S(\text{user}) \cap Rec(\text{user})| / |Rec(\text{user})|, \tag{13}$$

$$\textbf{Precision} = |S(\text{user}) \cap Rec(\text{user})| / |S(\text{user})|, \tag{14}$$

$$\textbf{FI} = 2 \times \text{Recall} \times \text{Precision} / \text{Recall} + \text{Precision}, \tag{15}$$

where $S$(user) is the actual basket for the compared user; $Rec$(user) is the recommendation item set. Recall is the ratio of items successfully recommended, whereas precision measures the user's satisfactory degree. F1 is a leverage measure when recall and precision conflict with each other.

### 3.3.3. *Summary of offline and online operation procedures*

After introducing the individual sub-modules of the proposed RS, we would summarize the operation procedures for the proposed RS. The procedures are categorized into offline and online operations.

#### 3.3.3.1. *Offline operation procedures.*

Step 1. Construct user-groups through user's demographic features and item-groups by item attributes to obtain $U^j = \{u_{f^j}(\omega_g) | f^j = 1^j, 2^j, \ldots, F^j, j = 1,2,\ldots J\}$ and $P^i = \{p_{d^i}(\alpha_k) | d^i = 1^i, 2^i, \ldots, D^i, i = 1,2,\ldots,I\}$.

Step 2. Compute relative purchase priorities ($w_i^j$) between user-groups and item-groups by Eq. (3).

Step 3. Compute similarity measures between user-groups. Similarity function is used from common item set (Eq. (6)) or non-common item set (Eq. (8)).

Step 4. Derive out-of-clique probability measures by Eq. (4).

#### 3.3.3.2. Online operation procedures.

Step 1. Set up parameters on in-clique effects ($\theta$) and profit consideration ($\beta$) if adopt
 1.1. Maximal profit strategy, set $\beta=1$;
 1.2. Win–win strategy, set a $\beta \in (0,1)$;
 1.3. Best service strategy, set $\beta=0$.
Step 2. On-line inquiry of target users' profiles of demographic features ($u_{fj}(\omega_g)$), binary basket data ($C_{u_{f^\tau},p_{di}}$); the desired satisfaction level ($b^{f^j}$), and the budget limit ($B^{f^j}$).
Step 3. Classify target user into proper user-group by $U^j=\{u_{fj}(\omega_g)|f^j=1^j,2^j,\ldots,F^j,\quad j=1,2,\ldots J\}$.
 3.1. A *historical user* with basket data ($0<\theta\leq 1$)—compute purchase probabilities on each item with *CECF-C* (Eq. (6)) or *CECF-NC* (Eq. (8)).
 3.2. A *new user* without basket data ($\theta=0$)—retrieve out-of-clique probability measures as purchase probability on each item.
Step 4. Derive metadata from purchase probabilities (Eq. (5)) and user's request as input to Step 5.
Step 5. Run the analytical model and yield recommendation list.

## 4. Case study: laptops RS of a 3C retailer

3C industries of Taiwan have the most advanced technologies in the world. Among various electronic products, the experiments of our proposed RS are conducted specifically with laptops because of three reasons. (1) Laptop transactions are usually fewer than those of other electronic products so introducing an RS would be meaningful to attract the users; (2) fewer transactions are difficult to exploit when introducing the RS, so our proposed RS aims to solve this situation by incorporating clique effects; and (3) laptops are all highly priced so that the profit consideration would be more applicable.

Following the provided data of a 3C retailer, the prototype of the system was established and evaluated in this section by first describing the given database; the laptop data set contains 915 market baskets including 227 customers and 192 items. The types of items in the basket are ranged from two to eight for each user. The user's information is revealed by user types (defined with users' demographic features by the 3C retailer) and five user-groups are yielded ($U^1$, $U^2$, $U^3$, $U^4$, $U^5$). The item attributes ($k$) are denoted as: (1) central processing unit (CPU), (2) random-access memory (RAM), (3) brand, (4) storage capacity, and (5) weight. By our classification rules with $K=5$, the item-groups consist of 32 exclusive groups. Due to incomplete data, there are only 16 non-empty item-groups.

### 4.1. A case study with experiments

In the experiments, 227 customers are divided randomly into 20%/80% as testing and training data in an echo. We would conduct three experiments with different goals. In the first experiment, we shall compare the recommendation performance of conventional *CF* with our proposed recommendation approach *CECF* in two cases of *CECF-C* and *CECF-NC*, and the three schemes are all with a fixed neighborhood size of 20. In the second experiment, we would compare the recommendation performance as well as the profit gained with respect to supplier's market strategies as: (1) $\beta=1$ yields *maximal profit strategy* and (2) $\beta \in (0,1)$ yields the *win–win strategy*, (3) $\beta=0$ emphasizes the customer's benefit of the *best service strategy*. In the third experiment, we compare the sensitive F1 values with respect to the neighborhood sizes (3, 5, 7, 10, and 20) under three schemes of *CF*, *CECF-NC* with profit consideration ($\beta=0.2$) and *CECF-NC* with non-profit consideration.

Three measures of recall, precision and F1 will be used for evaluation. Different values of parameters were chosen to demonstrate their impacts as sensitivity analysis. We pick one of the echoes for illustration in the following section. All experimental procedures would be shown in compliance with the procedures proposed in Section 3.3.3 (Table 4–6).

*Offline operation procedures* (*training data*)

**Table 5**
Purchase probabilities of new users by *CECF-NC* ($\theta=0$).

| | $p_1^2$ | $p_2^2$ | $p_{13}^2$ | $p_2^4$ | $p_3^4$ | $p_5^{16}$ | $p_6^{16}$ |
|---|---|---|---|---|---|---|---|
| $u_1^1$ | 0.0026 | 0.0026 | 0.0026 | 0.0070 | 0.0070 | 0.0008 | 0.0008 |
| $u_2^1$ | 0.0027 | 0.0027 | 0.0027 | 0.0078 | 0.0078 | 0.0007 | 0.0007 |
| $u_3^1$ | 0.0028 | 0.0028 | 0.0028 | 0.0088 | 0.0088 | 0.0004 | 0.0004 |
| $u_4^1$ | 0.0031 | 0.0031 | 0.0031 | 0.0087 | 0.0087 | 0.0007 | 0.0007 |
| $u_5^1$ | 0.0028 | 0.0028 | 0.0028 | 0.0084 | 0.0084 | 0.0006 | 0.0006 |

**Table 6**
Purchase probabilities of new users by *CECF-C* ($\theta=0$).

| | $p_1^2$ | $p_2^2$ | $p_{13}^2$ | $p_2^4$ | $p_3^4$ | $p_5^{16}$ | $p_6^{16}$ |
|---|---|---|---|---|---|---|---|
| $u_1^1$ | 0.0025 | 0.0025 | 0.0025 | 0.0069 | 0.0069 | 0.0008 | 0.0008 |
| $u_2^1$ | 0.0024 | 0.0024 | 0.0024 | 0.0068 | 0.0068 | 0.0007 | 0.0007 |
| $u_3^1$ | 0.0025 | 0.0025 | 0.0025 | 0.0079 | 0.0079 | 0.0004 | 0.0004 |
| $u_4^1$ | 0.0028 | 0.0028 | 0.0028 | 0.0076 | 0.0076 | 0.0008 | 0.0008 |
| $u_5^1$ | 0.0024 | 0.0024 | 0.0024 | 0.0073 | 0.0073 | 0.0005 | 0.0005 |

**Table 4**
Out-of-clique probability measures.

| *NC* | $P^1$ | $P^2$ | $P^3$ | $P^4$ | $P^5$ | $P^6$ | $P^7$ | $P^8$ | $P^9$ | $P^{11}$ | $P^{12}$ | $P^{13}$ | $P^{14}$ | $P^{15}$ | $P^{16}$ | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $U^1$ | 0.003 | 0.082 | 0.005 | 0.222 | 0.005 | 0.060 | 0.365 | 0.001 | 0.054 | 0.029 | 0.029 | 0.054 | 0.026 | 0.007 | 0.034 | 1 |
| $U^2$ | 0.003 | 0.085 | 0.004 | 0.244 | 0.019 | 0.053 | 0.335 | 0.000 | 0.060 | 0.024 | 0.020 | 0.072 | 0.023 | 0.006 | 0.030 | 1 |
| $U^3$ | 0.003 | 0.087 | 0.005 | 0.278 | 0.021 | 0.062 | 0.326 | 0.001 | 0.057 | 0.022 | 0.022 | 0.072 | 0.012 | 0.000 | 0.021 | 1 |
| $U^4$ | 0.004 | 0.095 | 0.001 | 0.270 | 0.018 | 0.062 | 0.323 | 0.001 | 0.050 | 0.024 | 0.024 | 0.063 | 0.016 | 0.006 | 0.022 | 1 |
| $U^5$ | 0.000 | 0.088 | 0.006 | 0.264 | 0.016 | 0.058 | 0.331 | 0.001 | 0.055 | 0.017 | 0.020 | 0.070 | 0.022 | 0.006 | 0.029 | 1 |
| *C* | $P^1$ | $P^2$ | $P^3$ | $P^4$ | $P^5$ | $P^6$ | $P^7$ | $P^8$ | $P^9$ | $P^{11}$ | $P^{12}$ | $P^{13}$ | $P^{14}$ | $P^{15}$ | $P^{16}$ | Total |
| $U^1$ | 0.003 | 0.081 | 0.005 | 0.219 | 0.005 | 0.060 | 0.366 | 0.001 | 0.055 | 0.029 | 0.029 | 0.055 | 0.025 | 0.006 | 0.035 | 1 |
| $U^2$ | 0.004 | 0.077 | 0.005 | 0.216 | 0.013 | 0.054 | 0.359 | 0.000 | 0.063 | 0.031 | 0.025 | 0.065 | 0.025 | 0.005 | 0.034 | 1 |
| $U^3$ | 0.004 | 0.081 | 0.006 | 0.254 | 0.015 | 0.062 | 0.347 | 0.001 | 0.058 | 0.027 | 0.026 | 0.066 | 0.015 | 0.000 | 0.025 | 1 |
| $U^4$ | 0.007 | 0.088 | 0.001 | 0.239 | 0.012 | 0.063 | 0.349 | 0.001 | 0.053 | 0.034 | 0.031 | 0.054 | 0.016 | 0.004 | 0.024 | 1 |
| $U^5$ | 0.000 | 0.077 | 0.009 | 0.233 | 0.008 | 0.057 | 0.360 | 0.001 | 0.059 | 0.022 | 0.023 | 0.067 | 0.026 | 0.004 | 0.036 | 1 |

Step 1.

(1) Construct user-groups, $U^j$, $j=1, 2, \ldots, 5$, $|\cup_j U^j| = 182$. (227*0.8 $\cong$ 182)

(2) Construct item-groups, $P^i$, $i=1, 2, \ldots, 16$, $|\cup_j P^i| = 192$

Step 2. Compute relative purchase priority $w_i^j$.

Step 3. Compute similarity measures between user-groups by Common item set function i.e. Eq. (6) and Non-Common item set function i.e. Eq. (8).

Step 4. Derive out-of-clique probability measures by Eq. (4) as shown in Table 4. Note that the probability measures in each row are normalized and ensured that they sum up to 1.

*Online operation procedures (testing data)*

Step 1. Set up parameters on in-clique effects ($\theta$) and profit consideration ($\beta$), respectively. For implementation, the system could set up $\theta$ and $\beta$ as arbitrary values. In the experiments, we set up $\theta$ to be 0, 0.1, 0.2,…,1 and $\beta$ to be 0, 0.2, 0.4,…,1 for testing.

Step 2. The users are tested as *new users* or *historical users* by setting $\theta=0$ or $0<\theta\leq 1$, respectively. Satisfaction levels ($b^{f^j}$) are also defined to be 0.7, 0.8, 0.9 for experiments. Budget limits ($B^{f^j}$) are set arbitrary values that are lower than the summation of all items' prices.

Step 3. Classify target user into one user-group by $U^j = \{u_{f^j}(\omega_g) | f^j = 1^j, 2^j, \ldots, F^j, j = 1,2,\ldots,5\}$.

Step 3.1. The situation is simulated in a manner where some historical users are recommended when we set $0<\theta\leq 1$.

Step 3.2. The situation is simulated wherein some new users $(u_1^1, u_2^1, \ldots)$ are recommended by *CECF-NC* or *CECF-C* respectively when we set $\theta=0$, which is shown in Tables 5 and 6. Note that when a target user is regarded as a new user, the probability measures for him/her could be only derived from out-of-clique measures. For instance, in Table 4, the probability of $U^1$ to $P^{16}$ is 0.025, which shall be the same with that of $u_1^1$ to $p_5^{16}$ and $p_6^{16}$ in Table 5. The value is 0.0008 instead of 0.025 due to normalization.

Step 4 and Step 5.

In the two steps, the target user's metadata is obtained and fed to the analytical model, and the output of recommendations is then yielded. We skip the list of the recommendation results and directly compare the performance of the proposed operation module by the following experiments.

**Experiment 1.** The performance of the recommendation results on *CECF-C*, *CECF-NC*, and *CF* is shown in Table 7, with evaluation of recall, precision, and F1 under sample values $\theta$, with a neighborhood size of 20. Note that when $\theta=1$, *CECF-C* and *CECF-NC* both become the *CF* since out-of-clique effects no longer exist. In Table 7, the results of an average performance show that *CECF-C* and *CECF-NC* perform better than *CF* except $\theta=0$. In addition, it could be also observed that *CECF-NC* performs slightly better than *CECF-C*. In Fig. 4, *CECF-NC* has been compared with *CF*; in the figure, the *CECF-NC* performs much better than *CF* in recall and F1 ($p$-value $<$ 0.001, 95% confidence level), and slightly better in precision.

**Table 7**
Average performance of *CECF-C*, *CECF-NC* and *CF*.

| $\theta$ | *CECF-NC* | | | *CECF-C* | | |
|---|---|---|---|---|---|---|
| | Recall | Precision | F1 | Recall | Precision | F1 |
| 0 | 0.297 | 0.458 | 0.325 | 0.297 | 0.458 | 0.325 |
| 0.1 | 0.877 | 0.928 | 0.939 | 0.867 | 0.925 | 0.932 |
| 0.2 | 0.900 | 0.942 | 0.962 | 0.893 | 0.938 | 0.955 |
| 0.4 | 0.908 | 0.943 | 0.963 | 0.903 | 0.942 | 0.959 |
| 0.5 | 0.910 | 0.945 | 0.968 | 0.910 | 0.945 | 0.968 |
| 0.6 | 0.911 | 0.945 | 0.968 | 0.907 | 0.945 | 0.967 |
| 0.7 | 0.910 | 0.945 | 0.968 | 0.910 | 0.945 | 0.968 |
| 0.8 | 0.910 | 0.945 | 0.968 | 0.910 | 0.945 | 0.968 |
| 0.9 | 0.910 | 0.945 | 0.968 | 0.910 | 0.945 | 0.968 |
| 1(*CF*) | 0.457 | 0.930 | 0.569 | 0.457 | 0.930 | 0.569 |

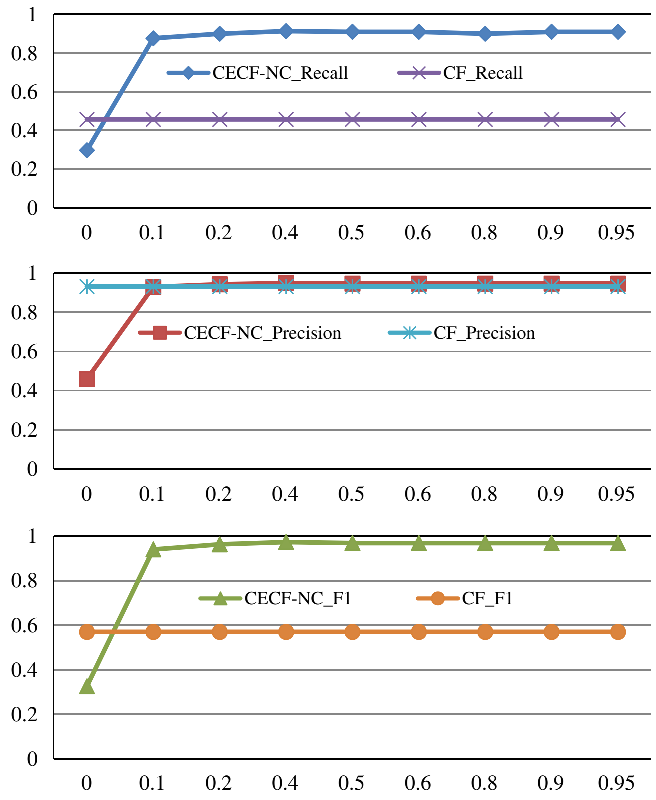

**Fig. 4.** Comparison of *CECF-NC* and *CF*.

**Experiment 2.** In Experiment 1, the average performance is better and more stable when *CECF-NC* and $\theta=0.6$ are used. Therefore, we set up $\theta$ to be 0.6 and continue experimenting on the analytical model by introducing $\beta$ to be 0, 0.2, 0.4,…,1 and satisfaction level ($b^{f^j}$) to be 0.7, 0.8, 0.9 under users' budget limits. We compare the *CECF-NC* with profit consideration as well as non-profit consideration in terms of recall, precision and F1 as shown in Fig. 5; and the difference of profit gained in the two cases are presented in Fig. 6. In Fig. 5, the results show that even when profit consideration is introduced, the recommendation performance would not be poorer ($p$-value $<$ 0.05, 95% confidence level). In Fig. 6, the results show that profit increases along $\beta$ increases from 0 to 1.

**Experiment 3.** In this experiment, we compare three recommendation schemes of *CF*, *CECF-NC* with profit consideration ($\beta=0.2$) and *CECF-NC* with non-profit consideration in terms of their F1 measures. Fig. 7 shows that the F1 values increase as the neighborhood size increases from 3, 5, 7, 10, to 20. In addition, Fig. 7 showed consistent results we obtained from the previous two experiments, that is, the *CECF-NC* with profit/non-profit consideration outperforms conventional *CF*.

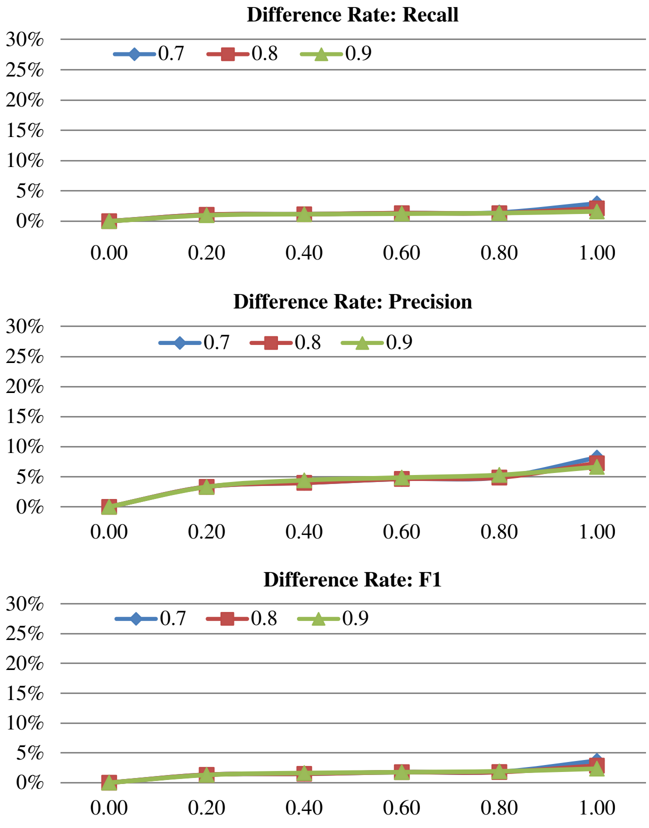

**Fig. 5.** The difference rates while introducing profit consideration.

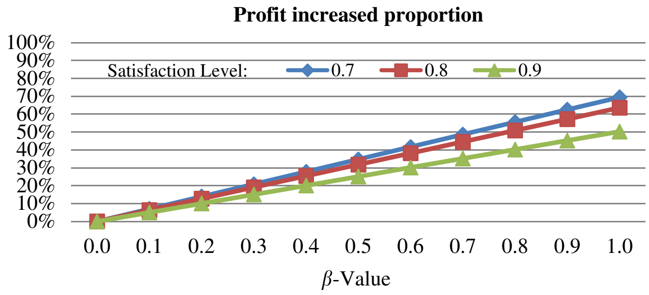

**Fig. 6.** Profit gained under different $\beta$-values.

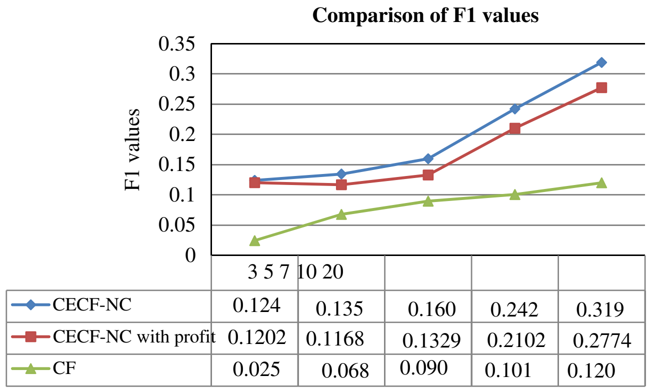

**Fig. 7.** Comparison of F1 values with respect to the neighborhood sizes.

### 4.2. Summary and remarks of experiments

We have conducted three experiments in the case study for the proposed strategy-oriented operation module of a 3C retailer in Taiwan. In the first experiment, we compared the performances of *CECF-C*, *CECF-NC*, and *CF* by three measures of Recall, Precision, and F1. In Table 7 and Fig. 4, it showed that the proposed *CECF-NC* and *CECF-C* perform better than conventional *CF* except for $\theta=0$, which was the situation in which new user recommendations was simulated. It has been mentioned that *CF* could not recommend while the target user is without basket data. Besides, the results in Experiment 1 also showed slightly better performance of *CECF-NC* as compared to *CECF-C*. The reason is that in the relatively sparse user's basket data, the non-common item set would show additional information for recommendation. In Table 7, we could observe that when $\theta=0.7$, 0.8, 0.9, the performances are almost the same; when $\theta=0.6$, the recommendation performance reaches the highest level. This phenomenon, which would be data-specific, tells that the effects from non-neighbor groups would not enhance but maintain the performance while $\theta \geq 0.6$.

In the second experiment, while introducing the profit parameter $\beta$ and user's satisfaction level ($b^{f^j}$), we set up the recommendation environment by $\theta=0.6$ for better and more stable performance. It is very important to note that while we introduce the profit parameter $\beta$ in the recommendation process, the recommendation performance with $\beta \in (0,1]$ would probably decrease since the goal of recommendation was no longer to emphasize user's benefits with $\beta=0$ only. Then the emphasis should be on whether the *service level decreases* while the *profit gain increases*. In Figs. 5 and 6, the results showed that while we increased $\beta$, the profit gained increases without losing recommendation performance. This phenomenon could be attributed to the analytical model since the recommended items are aligned with the user's satisfaction level. Therefore, even if we look forward to increasing the profit gains of the supplier, we would still maintain the recommendation performance for services.

In the third experiment, we tested the effect of neighborhood size on F1 measure. The three schemes all showed consistent results that the F1 values were positively related to the neighborhood sizes; and the *CECF-NC* outperformed *CF* in F1 measure. In particular, while the neighborhood sizes were small i.e. 3, 5, 7, the *CECF-NC* still reached higher F1 values than *CF*, which shows the advantage of using the clique effects to compensate for rare information.

## 5. Conclusion and further research

In the field of RSs, there have been numerous studies proposed in order to find the best recommendation to users. Among those studies, *CF* has been regarded as the most effective method for its recommendation accuracy and flexibility. However, in practice, it is confronted with the problem that target users with rare information could not get recommendations from the system. Although many approaches based on *CF* have been proposed to pursue better performance by increasing service levels and solving the problem of sparse data, the excessive emphasis on recommendation performance would lead to overlooking the profit consideration, which is also an essential concern for an EC company.

In addition, a systematic and comprehensive module for an RS is still lacking. In this regard, we have proposed a strategy-oriented operation module that could be comprehensively applied to EC Web sites as a decision support module so that the choice of various marketing strategies combining profit consideration for suppliers and users can be developed. Consequently, under the framework of the proposed recommender module, we also developed a method named *CECF* to predict users' purchase behavior.

In the experimental results, the proposed *CECF* performed better than *CF* and could provide a promising solution to the "new user" problem. Nevertheless, introducing profit consideration into an RS would cause a decrease in recommendation performance. The proposed module with marketing strategies also shows achievements not inferior to what is only based on recommendation performance. For the further research, the consideration may be placed on the different sources of users' profiles such as rating data. Lastly, updating a constructed database when online transactions increase will also be an impending issue.

## Acknowledgment

The authors acknowledge the financial support from National Science Council with project number (NSC97-2221-E007-092).

## Appendix A

Let us specify the phenomenon as follows. Consider three users $u_{f^1}$, $u_{f^2}$, and $u_{f^3}$, with their market basket as $S(u_{f^1})=\{\text{item 1,item 2,item 3,item 5}\}$; $S(u_{f^2})=\{\text{item 3,item 5,item 10,item 12}\}$; $S(u_{f^3})=\{\text{item 3,item 10,item 11}\}$ and universal item set $P$=item 1,…,item 1,000. Then the similarity measure specified in Eq. (7) is further revised as

$$sim(u_{f^j},u_{f^\tau})=\frac{|\overline{S}(u_{f^j})\cap\overline{S}(u_{f^\tau})|}{|\overline{S}(u_{f^j})\cup\overline{S}(u_{f^\tau})|}=\frac{|P-(S(u_{f^j})\cup S(u_{f^\tau}))|}{|P-(S(u_{f^j})\cap S(u_{f^\tau}))|}.$$

Therefore, the similarity measure for each pair is

$$sim(u_{f^1},u_{f^2})=\frac{|P-(S(u_{f^1})\cup S(u_{f^2}))|}{|P-(S(u_{f^1})\cap S(u_{f^2}))|}=\frac{1000-6}{1000-2}=0.996,$$

$$sim(u_{f^1},u_{f^3})=\frac{|P-(S(u_{f^1})\cup S(u_{f^3}))|}{|P-(S(u_{f^1})\cap S(u_{f^3}))|}=\frac{1000-6}{1000-1}=0.995,$$

$$sim(u_{f^2},u_{f^3})=\frac{|P-(S(u_{f^2})\cup S(u_{f^3}))|}{|P-(S(u_{f^2})\cap S(u_{f^3}))|}=\frac{1000-5}{1000-2}=0.997.$$

The instance indicates that in the large-scale basket data, $(|P-(S(u_{f^j})\cup S(u_{f^\tau}))|)/(|P-(S(u_{f^j})\cap S(u_{f^\tau}))|)$ would be always close to 1 since $|P|$ is much larger than $|(S(u_{f^j})\cap S(u_{f^\tau}))|$ or $|(S(u_{f^j})\cup S(u_{f^\tau}))|$. As a consequence, in order to narrow down the gaps between $|P|$ and $|(S(u_{f^j})\cap S(u_{f^\tau}))|$ as well as $|(S(u_{f^j})\cup S(u_{f^\tau}))|$, we suggest two directions: one is to decrease $|P|$ and the other is to increase $|(S(u_{f^j})\cap S(u_{f^\tau}))|$ and $|(S(u_{f^j})\cup S(u_{f^\tau}))|$. For $P$, the union set of all purchased item sets could be used instead of the universal set, as described by $\bigcup_{i=1}^{j} S(U^j)$. Alternatively, to increase the co-purchased and purchased items, that is, $|(S(u_{f^j})\cap S(u_{f^\tau}))|$ and $|(S(u_{f^j})\cup S(u_{f^\tau}))|$.

## Appendix B

The similarity measures indicated in the Appendix A could be computed as follows. (Here each user is regarded as single user-group to significantly identify differences):

$$sim(U^1,U^2)=\frac{|\bigcup_{j=1}^{3}S(U^j)-(S(U^1)\cup S(U^2))|}{|\bigcup_{j=1}^{3}S(U^j)-(S(U^1)\cap S(U^2))|}=\frac{7-6}{7-2}=0.20,$$

$$sim(U^1,U^3)=\frac{|\bigcup_{j=1}^{3}S(U^j)-(S(U^1)\cup S(U^3))|}{|\bigcup_{j=1}^{3}S(U^j)-(S(U^1)\cap S(U^3))|}=\frac{7-6}{7-1}=0.17,$$

$$sim(U^2,U^3)=\frac{|\bigcup_{j=1}^{3}S(U^j)-(S(U^2)\cup S(U^3))|}{|\bigcup_{j=1}^{3}S(U^j)-(S(U^2)\cap S(U^3))|}=\frac{7-5}{7-2}=0.4.$$

## References

[1] Adulbhan P, Tabucanon MT. Bicriterion linear programming. Computers & Operations Research 1977;4(2):147–53.

[2] Aggarwal CC, Wolf JL, Yu PS. A new method for similarity indexing of market basket data. In: SIGMOD conference, 1999. p. 407–18.

[3] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the conference management of data, ACM SIGMOD, Washington, DC, 1993. p. 207–16.

[4] Bazgan C, Hugot H, Vanderpooten D. Solving efficiently the 0–1 multi-objective knapsack problem. Computers & Operations Research 2009;36(1) 260–79.

[5] Bell RM, Koren Y, Volinsky C. Modeling relationships at multiple scales to improve accuracy of large recommender systems. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining; 2007.

[6] Breese JS, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th conference uncertainly in artificial intelligence, 1998. p. 43–52.

[7] Burke R. Hybrid recommender systems: survey and experiments. User Modeling and User-Adapted Interaction 2002;12:331–70.

[8] Captivo ME, Clímaco J, Figueira J, Martins E, Santos JL. Solving bicriteria 0–1 knapsack problems using a labeling algorithm. Computers & Operations Research 2003;30(12):1865–86.

[9] Changchien SW, Lee CF, Hsu YJ. On-line personalized sales promotion in electronic commerce. Expert Systems with Applications 2004;27:35–52.

[10] Chen LS, Hsu FH, Chen MC, Hsu YC. Developing recommender systems with the consideration of product profitability for sellers. Information Sciences 2008;178:1032–48.

[11] Chen L, Sycara K. Webmate: a personal agent for browsing and searching. In: Proceedings of 2nd conference on autonomous agents, 1998. p. 132–9.

[12] Cho YH, Kim JK. Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce. Expert Systems with Applications 2004;26:233–46.

[13] Fleszar K, Hindi KS. Fast, effective heuristics for the 0–1 multi-dimensional knapsack problem. Computers & Operations Research 2009;36(5):1602–7.

[14] Goldberg K, Roeder T, Guptra D, Perkins C. Eigentaste: a constant-time collaborative filtering algorithm. Information Retrieval Journal 2001;4: 133–151.

[15] Guo J. Business-to-business electronic market place selection. Enterprise Information Systems 2007;1(4):383–419.

[16] Guo J. Collaborative conceptualization: towards a conceptual foundation of interoperable electronic product catalogue system design. Enterprise Information Systems 2009;3(1):59–94.

[17] Hanani U, Shapira B, Shoval P. Information filtering: overview of issues, research and systems. User Modeling and User-Adapted Interaction 2001;11: 203–259.

[18] Herlocker JL, Konstan JA, Terveen LG, Riedl JT. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 2004;22(1):5–53.

[19] Hogg T. Inferring preference correlations from social networks. Electronic Commerce Research and Applications 2010;9(1):29–37.

[20] Hsu PY, Chen YL, Ling CC. Algorithms for mining association rules in bag databases. Information Sciences 2004;166:31–47.

[21] Karypis G. Evaluation of item-based Top-N recommendation algorithms. Technical Report CS-TR-00-46, Computer Science Department, University of Minnesota; 2000.

[22] Kaufman L, Rousseeuw PJ. Finding groups in data. An introduction to cluster analysis. New York: Wiley; 1990.

[23] Kim BM, Li Q, Park CS, Kim SG, Kim JY. A new approach for combining content-based and collaborative filters. Journal of Intelligent Information System 2006;27:79–91.

[24] Kim D, Yum BJ. Collaborative filtering based on iterative principal component analysis. Expert Systems with Applications 2005;28(4):823–30.

[25] Kim HT, Ji AT, Kim HaI, Jo GS. Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation. Electronic Commerce Research and Applications 2010;9(1):73–83.

[26] Kohrs A, Meriadlo B. Creating user adapted websites by the use of collaborative filtering. Interacting with Computers 2001;13:695–716.

[27] Konstan JA, Riedl J, Borchers A, Herlocker JL. Recommender systems: a grouplens perspective. In: Recommender systems: papers from the 1998 workshop (AAAI technical report WS-98-08). Penlo Park, CA: AAAI Press, p. 60–4.

[28] Lam C. SNACK: incorporating social network information in automated collaborative filtering. In: Proceedings of the fifth ACM conference on electronic commerce (EC'04). New York, NY, USA: ACM Press; 2004. p. 254–5.

[29] Lee CH, Kim YH, Rhee PK. Web personalization expert with combining collaborative filtering and association rule mining technique. Expert Systems with Applications 2001;21:131–7.

[30] Li H, Wang H. A multi-agent-based model for a negotiation support system in electronic commerce. Enterprise Information Systems 2007;1(4): 457–472.

[31] Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing 2003;7:76–80.

[32] Liu DR, Shih YY. Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. The Journal of Systems and Software 2005;77:181–91.

[33] Liu DR, Shih YY. Integrating AHP and data mining for product recommendation based on customer lifetime value. Information and Management 2005;42:387–400.

[34] McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annual Review of Sociology 2001;27:415–44.

[35] Mild A, Reutterer T. Collaborative filtering methods for binary market basket data analysis. In: Lecture notes in computer science, vol. 2252. 2001. p. 302–13.

[36] Mild A, Reutterer T. An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data. Journal of Retailing and Consumer Services 2003;10:123–33.

[37] Montaner M, López B, Rosa JLDL. A taxonomy of recommender agents on the Internet. Artificial Intelligence Review 2003;19:285–330.

[38] Mooney RJ, Roy L. Content-based book recommending using learning for text categorization. In: Proceedings of 5th ACM conference on digital libraries, 2000. p. 195–204.

[39] Phillips C, Meeker M. The B2B Internet Report: Collaborative Commerce. Morgan Stanley Dean Witter Equity Research; 2000.

[40] Raghavan VV, Wong SKM. A critical analysis of vector space model for information retrieval. Journal of the American Society for Information Science 1986;37:279–87.

[41] Russell GJ, Petersen A. Analysis of cross category dependence in market basket selection. Journal of Retailing 2000;76:367–92.

[42] Sarwar BM, Karypis G, Konstan JA, Riedl J. Analysis of recommendation algorithms for E-commerce. In: Proceedings of 2nd ACM EC, 2000. p. 158–67.

[43] Sarwar BM, Karypis G, Konstan JA, Riedl J. Application of dimensionality reduction in recommender system—a case study. In: Proceedings of ACM WebKDD workshop. Boston, MA; 2000.

[44] Sarwar BM, Konstan JA, Borchers A, Herlocker J, Miller B, Riedl J. Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. In: Proceedings of CSCW. Seattle, Washington; 1998.

[45] Schafer JB, Konstan JA, Riedl J. Recommender systems in e-commerce. In: Proceedings of 1st ACM EC, 1999. p. 158–66.

[46] Schafer JB, Konstan JA, Riedl J. E-commerce recommendation applications. Data Mining and Knowledge Discovery 2001;5:115–53.

[47] Schein AI, Popescul A, Ungar LH, Pennock DM. Methods and metrics for cold-start recommendations. In: Proceedings of the 25th ACM SIGIR on research and development in information retrieval, Tampere, Finland; 2002.

[48] The Economist. Hands off the Internet. The Economist 1997; 344(8024): 15–6.

[49] Theodoridis S, Koutroumbas K. Pattern recognition. USA: Academic Press Elsevier (USA); 2003. ISBN: 0126858756, 9780126858754.

[50] Wang HF. Multicriteria decision analysis—from certainty to uncertainty. Taiwan: Ting Lung Book Co.; 2004. ISBN: 986-7777-55-7

[51] Wang HF, Wu CT. A mathematical model for product selection strategies in a recommender system. Expert Systems with Applications 2009;36(3): 7299–7308.

[52] Wang S, Archer NP. Electronic marketplace definition and classifications: literature review and clarifications. Enterprise Information Systems 2007;1(1):89–112.

[53] Xu X. Effects of two national environmental factors on e-commerce functionality adoption: a cross-country case study of a global bank. Enterprise Information Systems 2008;2(3):325–39.

[54] Yang HW, Pan ZG, Wang XZ, Xu B. A personalized products selection assistance based on e-commerce machine learning. In: Proceedings of 3rd conference on machine learning and cybernetics, 2004. p. 26–9.