Collaborative Recommender System for Online Book Retailing by using Associative Rule Mining Approach

Thet Mon Phyo; Nhinn Minn Oo
Computer University (Maubin)
E-mail: snowwhite.yuya@gmail.com, hninminoo@gmail.com

Abstract

Recommender systems help the users to find and evaluate items of interest. They connect user with items to consume by associating the opinions of other individuals with the consuming user's actions or opinions. Such system has become powerful tool in e-commerce sites to facilitate the buying process. In this paper, we present the application of recommender system for online book retailing. We use a hybrid model of Association Rule Mining approach to find the frequent book-itemsets and Collaborative filtering techniques to generate recommendations base on the past buying behavior of the consumers and the items in the rules observed by association rule mining approach. We use the Apriori algorithm as the underlying techniques of association rule mining approach.

Keywords
Recommendation, Online retailing, Association rules mining, Collaborative filtering,

1. Introduction

In online retailing environment, customer search and process information in order to choose which product or service should buy from the many available options. Recommender system guide people toward products they are more likely to find interesting. It uses the opinions of a community of users to help individuals in that community more effectively identify content of interest from a potentially overwhelming set of choices. One of the most successful technologies for recommender systems, called collaborative filtering, has been developed to obtain the powerful recommender system. It is a hybrid approach of association rule mining and collaborative filtering techniques. In association rule mining approach, Apriori algorithm is used to find the frequent-itemsets. A possible conjecture is that the number of item combinations resulted from the Apriori -based method is very much smaller to state the recommended items group. So, most of the customized recommendation systems currently operating in real-world online retailing environments are developed base on Collaborative Filtering methods. This method produces efficient recommendation by using data from user with similar preferences and frequent itemsets derived from the apriori methods in order to determine an active customer's preferences and the recommended item list. So it can be determined as the basket-based estimation process. We construct a transaction table with the data from daily transaction list of an e-book retailer site.

2. Related Work

There are many online recommendation services available, which span the areas of book, music, movie, web-page and restaurant recommendations. There are currently two underlying categories of recommendation techniques: Content-based and Collaborative based recommenders. A typical content-based recommender selects items for recommendation on the grounds that they are similar to items that the user has liked in the past. In contrast, collaborative filtering approaches to recommendation adopt a very different standpoint, under the assumption that content-based descriptions may not be available. Instead they rely on ratings-based user profiles containing items that the user has previously rated according to their appeal or relevance. Recommendations are produced by locating users with similar ratings histories and by selecting items from these profiles that have been highly rated but that are absent from the target user's profile. But collaborative recommendation is more popular than content-based recommendation, mainly because in many domains (such as music, restaurants) it is hard to extract useful features from articles, which is generally a step required for content-based recommendation. Many research efforts are invested in exploiting efficient algorithms for collaborative recommendation.

Association rules are quite appropriate for the collaborative recommender systems and they can achieve good performance. So, this system combines both association rules mining to extract the associations between product categories and collaborative filtering techniques to identify the specific products for recommendations.

3. Recommender System

Recommender systems from a specific type of information filtering technique that attempts to present information items that are likely interested to the user. Recommender system can be utilized to efficiently provide personalized services on most of E-commerce domains, benefiting both the customer and the merchant. Recommender systems will benefit the customer by making him suggestions on items that he is assumedly like. At the same time, the business will be benefited by the increased of sales which will normally occur when the customer is presented with more appealing items [3]. There are two basic entities concerned in a recommender system. First, user is a person utilizes the recommender system providing his opinion and receiving recommendation about items. Second, item is being rated by users and the data of them are collected by the system.

There are three main processes in recommender systems algorithm [2]. They are representation, neighborhood formation and recommendation generation. **In representation process**, the input data is defined as a collection of numerical ratings of m users on n items, expressed by the m*n user-item table. The term of this user-item table of the input data set is representation or transaction. **In neighborhood formation**, similarity between users in the user-item table and the active users will be calculated. **In recommendation process**, it produces a recommendation, which will be expressed as a list of the top-N items that the active user will appreciate more. In both cases, the result should be based on the neighborhood of users.

4. Data Set Representation

We construct the **Sales** dataset with the tables such as Customer-Behavior, Customers, Products, Sales, Sales-Detail. Transaction table is Sales-Detail and it has the relations to the rests. All of the data are obtained from infibeam.com book-retailer site. It is a web-based recommender system and hundreds of users visit infibeam to rate and receive recommendation for books. We randomly selected enough users to obtain 896 tuples (i.e., 896 users) and 23 columns (i.e., 23 attributes concerned with the customer buying behavior).

5. Neighborhood Formation

At this stage, similarity between users in the user-item table and the active user will be calculated. Users similar to the active user will form a proximity-based neighborhood with the active user. To do this stage, we need to evaluate two approaches: Association Rule mining and Collaborative Filtering.

5.1. Association Rule Mining

Association Rules describe the association between the items in a specific transaction database. Each transaction is a set of items and an association rule is an expression of the form X=>Y where X and Y are a set of items. In discovering association rule, it needs two parameter values: minimum confidence of the rule (c%) and minimum support of the rule (s%). Here, c% of transactions that contains X and also contains Y; s% of all transactions that contain both of these items [1]. Association rule mining process contains the following process:

1. **Find all frequent itemsets :** Each support of these frequent itemsets will at least equal to predetermined minimum support count.
2. **Generate strong association rules from the frequent itemsets:** These rules must be frequent itemsets and must satisfy minimum support count and minimum confidence

There are many methods to provide association rule generation. In our system, we use Apriori algorithm for mining frequent itemset. It is a level-wise search and process is taken by the following steps [4].

1. Scan the transaction database to get the support count of each itemset, Compare these support count with minimum support count and get a set of frequent itemsets, L1.
2. Use L_{k-1} join L_{k-1} to generate a set of candidate k-itemsets and use Apriori property to prune the unfrequented k-itemsets from this set.
3. Scan the transaction database to get the support S of each candidate k-itemset in the final set, compare S with minimum support and get a set of frequent k-itemsets, L_k.
4. If the candidate set is equal to null then go to step 5, otherwise go to step 2.
5. For each frequent itemset l, generate all nonempty subsets of l.

6. For every nonempty subsets s of l, output the rule "s=> (l-s)" if confidence of the rule "s=> (l-s)" (support S of l/support S of s) is equal to minimum confidence.

To evaluate e-book recommender system, items in the derived rules are collected as the group and find who purchased these items. And then what other items they purchased. This is the concept of collaborative filtering technique.

5.2 Collaborative filtering technique

Collaborative filtering technique filtered information based on the preferences of each user. The functionality of collaborative filtering is based on the collection of user rating that is used to identify the users that have similar preferences. Such users become members of the same group and Collaborative filtering recommends to people in the same group the products that their co-members prefer [5]. Collaborative filtering system usually takes two steps:

1. Look for users who share the same rating patterns with the active user (i.e. the user whom the prediction is for).
2. Use the rating from those like-minded users found in step1 to calculate a prediction for the active user.

The items filtered by collaborative filtering process are specified as recommended items and to show these items to the user, they are sent to the recommendation process.

6. Recommendation Process

Recommendation process is a final step in recommender system and it produces the recommended list that guides the users to find items they want to buy from a business. In our system, we present the two forms of recommend. If the user has not collected items they prefer, system shows 15 top selling products of all transactions as first recommended list. Otherwise, system shows at most 15 top selling products that are associated with the user's preferable items as second recommended list.

7. Proposed System Design

In this paper, we present e-book retailing system shown in figure 1. It will generate recommendations for book items by applying the association rule mining to select book category and collaborative filtering techniques to identify the specific books that it should propose.

When a user visits on e-book retailer site, this site shows the book items to the user. If the user wants to buy any book, he must log in to this site. If he has not had a user-id, he made registration on this site. After registration has been completed, customer profile table is updated with the registration data.

Once the user logs in on the website of e-book retailer, webpage shows 15 top selling products of all transactions as first recommended list. When the user selects any one prefer book, the books associated with this book are identify by using the association rule mining process. Here, the interestingness measures of association rule mining process are minimum support threshold 2% and minimum confidence threshold 70%. And then, Collaborative filtering technique identify the specific books that it should proposed base on the book items derived from association rule mining approach and the book items that are preferred by the co-members of the active user. The books derived from the collaborative filtering technique are shown as second recommended list. Here, the target rate of book items is at most 15.

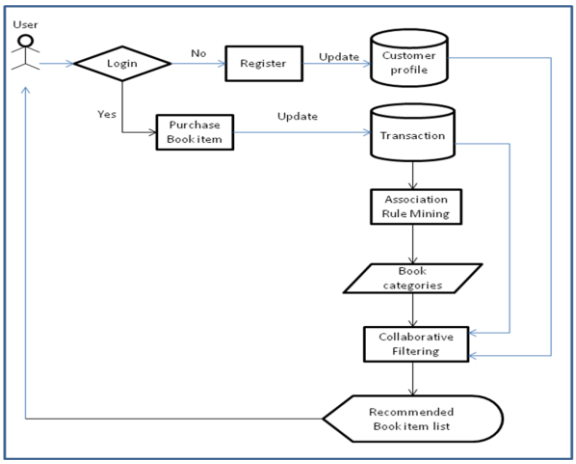

Figure 1: Proposed System Design

8. Evaluation Result

Collaborative filtering algorithms make use of simple ratings-based profiles and generate recommendations on the basis of similarity between the target user and neighboring profiles. On average, two users are unlikely to have rated many of the same items but there will be little direct overlap between their profiles. This is problematic when it comes to identifying similar profiles because it means that there is no direct way to measure the similarity between two profile cases unless we have access to similarity knowledge that allows us to compare non-identical profile items. In this case, we use of association rule mining techniques in order to discover this type of knowledge and Apriori based method is used as and underlying technique. For

example, we have transaction table as shown in Table 1.

Table 1: Transaction Table

TID	Items
T1	I1, I2, I3
T2	I2,I3, I4
T3	I3, I4, I5
T4	I2, I3, I4

After generating the support of each item by using Apriori-based method and pruning the itemsets that are not satisfy with minimum support count, we have the itemset l={I2, I3, I4}. The non-empty subsets of l are {I2}, {I3}, {I4}, {I2, I3}, {I2, I4}, {I3, I4}. Here we illustrate the correlation between the set {I2}, {I3} and {I4} as a sample and their support count are ({I2}: 3), ({I3}: 4), ({I4}: 3) respectively. Their confidence results are shown in Table 2.

In learning Item-Item Similarity Knowledge, there are many automated techniques that could be used to derive various sorts of similarity knowledge. The approach we have chosen is to apply data mining techniques, in particular the Apriori algorithm, to extract association rules between items in user-profile cases. To illustrate this we will use a sample of items transaction. For example, a person that likes I1 and I2 would not normally be comparable to a person that likes I3 and I4, but discovering a relationship between I2 and I3 would provide a basis for profile matching.

Table 2: Item Association Result

Rule	Support	Confidence
I3=>I2	3	100
I3=>I4	3	75
I2=>I4	2	66

Table 3: Similarity Matrix

	I3	I2	I4
I3	1	1	.75
I2	-	1	.66
I4	-	-	1

Treating user profiles as transactions and the item ratings therein as itemsets, the Apriori algorithm can be used to derive a set of rules and associated confidence levels between items. Table 2 shows some example rules that were generated by running Apriori on our sample data set. The confidence values are taken as similarity scores and used to fill in an item similarity matrix, as shown in Table 3.

The derivation of item similarity information naturally suggests two extensions to further elaborate the similarity matrix. First, the generated rules can be chained together to provide indirect relationships between items. And then we specify the minimum rate of similarities. Second, it chooses the rules above the minimum rate, for example, if the specified rate is 70, then the rule “likes I3=>likes I2” and “likes I3=>likes I4” are fired and it seems intuitive to think that a rule of the form “likes I2” => “likes I3” as reverse order of “likes I3=>likes I2”.

7. Conclusion

Recommender systems are a powerful new technology for extracting additional value for a business from its user databases. Recommender systems benefit users by enabling them to find items they like and help the business by generating more sales. E-book retailing system presented in this paper generates high quality recommendation because of the new collaborative filtering technique. Recommender systems are being stressed by huge volume of user data in existing corporate databases, and will be stressed even more by the increasing volume of user data available on the web. So, new technologies are needed that can dramatically improve the scalability of recommender systems.

10. References

[1] A. Savasere, E. Omiecinski, and S. Navathe, “Mining for strong negative associations in a large database of customer transactions”, Orlando, FL, Feb. 1998.

[2] B. Sarwar, G.Karypis, J.Konstan and J.Riedl, “Analysis of Recommendation Algorithms for E-Commerce”, Minneapoils, Minnesota.

[3] J.Breese, D.Heckerman and C.Kadie, “Empirical analysis of predictive algorithms for collaborative filtering”, In Proceedings of the Fourteenth Conference on University on Artificial Intelligence, Madison, WI, July 1998.

[4] J. Han, M. Kamber , “*Data Mining”,* Morgan Kaufmann Publishers, San Francisco, CA, 2009.

[5] Jerry Michalski, “Collaborative Filters”, Esther Dyson’s Report, 19 November 1996.