Book Recommendation System Based on Collaborative Filtering and Association Rule Mining for College Students

Anand Shanker Tewari
CSE department
NIT Patna
Patna, India
e-mail: anand@nitp.ac.in

Kumari Priyanka
CSE department
NIT Patna
Patna, India
e-mail:prnka0910@gmail.com

***Abstract*— Recommendation systems are tools in e-commerce websites which helps user to find the appropriate products. With the rapid development of internet technologies the number of online book selling websites has increased which enhanced the competition among them. This paper presents online book recommendation system for students reading textbooks. The main motive of this paper is to develop the technique which recommends most suitable books to the students according to their price range and publisher's name. This is based on the combined features of classification, user based collaborative filtering and association rule mining.**

Keywords—Association rule, Classification, Recommendation system, User based collaborative filtering.

I. INTRODUCTION

Now-a-days, the changing trends in technology and rapid development in internet has affected almost every aspect of the life. People depend on internet for various things, online shopping is one of them. Purchasing books on the internet from the huge collection of book is very difficult work for various academician and students [1]. From the huge number of books, it is really difficult to choose a particular textbook. So, the recommendation system technique plays very important role and helps user to get books according to their need and interest.

Basically this paper explains the new approach for recommending books to the students and academician. Here the features of classification are considered to classify the books of proper price range and authors. The features of user based collaborative filtering and association rule are considered for selecting the highly rated top N- books. Thus the combine feature of classification, user based collaborative filtering and association rule mining has considered for recommending books to buyers.

The rest of the paper is organized as follows. Section 2 describes classification; Section 3 describes user based collaborative filtering; Section 4 describes major point of association rule mining; Section 5 discusses our novel approach for recommending books to the college students according to price range and publishers.

II. CLASSIFICATION

Classification techniques provide a way to extract rules and patterns from data that can be used for predictions. Classification techniques are used to classify data records into one among a set of predefined classes. In classification, analysis of a particular data set is considered, to generate a set of grouping rules which can be used to classify future data [2].Classification is basically, one of the data mining fields that helps in interference learning, knowledge mining, prediction and decision making. A learning set is developed, in classification cases which results a classifier algorithm. Each sample in learning set contains some variable fields and a class number which specifies its class. After completing the learning and test phase, classifier gives the predicted class of incoming new data[8]. In this paper we used classification technique; We have defined three classes for publishers and declare the variable *Pb*. Three classes are *Reputed* with *Pb* value 0.5, *Average* with *Pb* value 0.4 and *New* with *Pb* value 0.2.

III. USER BASED COLLABORATIVE FILTERING

Collaborative filtering (CF) provides a way to do recommendation on the web [11]-[12]. Collaborative filtering creates a database of preferences for items by users [9]. In the design of the user preference data set, different attributes are considered to describe the various degrees of detail of user's profile [4]. User based collaborative filtering is a method of personalized recommendations from a dataset, to a user based on similarity between the interest profile of a user and those of other users [12]. In other words, it can be explained as a method to predict the ratings for a product by a target user with the help of other similar user's rating behavior and buying behavior.

Basically CF recommendation system finds the aggregation of nearest neighbor. Suppose there are k users who are nearest similar to target user these users will be considered as k

978-1-4799-6629-5/14/$31.00

nearest neighbors [5],[6],[7]. There are two ways to calculate the similarity: correlation and cosine. This paper takes correlation method (Pearson's correlation) to calculate the similarity of users[3]. Pearson's correlation, as following formula, measures the linear correlation between two vectors of ratings.

$$Sim\ (m,n) = \frac{\sum_{a \in I_{mn}} (R_{m,a} - A_m)(R_{n,a} - A_n)}{\sqrt{\sum_{a \in I_{mn}} (R_{m,a} - A_m)^2 \sum_{a \in I_{mn}} (R_{n,a} - A_n)^2}} \quad (1)$$

Where $R_{m,a}$ is the rating of the item a by user *m,* A_m is the average rating of user m for all co-rated items, and I_{mn} is the items set both rating by user m and user *n*.

Select all the neighbors who will be considered as recommenders. For the selection of neighbors, Threshold-based selection technique is used. According to this technique, users whose similarity exceeds a certain threshold value are considered as neighbors of the target user.

The rating(P_{xk}) of the target user *x* to the target item *k* is as following-

$$P_{xk} = A_x + \frac{\sum_{m=1}^{c} (R_{mk} - A_m)^* Sim(x,m)}{\sum_{m=1}^{c} Sim(x,m)} \quad (2)$$

Where A_x is the average rating of the target user *x* to the items, R_{mk} is the rating of the neighbor user m to the target item k, A_m is the average rating of the neighbor user m to the items , *Sim(x,m)* is the similarity of the target user *x* and the neighbor user *m*, and *c* is the number of the neighbors.

IV. Association Rule Mining

Association rule mining discovers interesting association and correlation relationship among large data set of items. The discovery of interesting correlation among large amount of business transaction data helps in various business decision making processes, such as customer shopping behaviour analysis [10]. One of the typical examples of association rule mining is Market basket analysis. In market basket analysis customer buying habit is analysed for finding association between different items customer keeps together in their shopping cart.

Let I = {i_1, i_2….. ,i_m} be a set of items. An association rule can be represented by this form A→B, where A ⊆ I, B ⊆ I and A ∩ B=∅ [13]. Association rule extract the pattern from the database based on the two measures minimum support and the minimum confidence. The support and confidence measures are described as stated in [10].

Support: The minimum percentage of instances in the database that contain all items listed in a given association rule. The rule A→B holds in the transaction set D with support *s*, where *s* is the percentage of the transactions in D containing A∪B.

$$Support\ (A \cup B) = P\ (A \cup B) \quad (3)$$

Confidence: Given a rule of the form "if A then B", confidence is defined as the conditional probability that B is true when A is known to be true. Thus, the rule (A→B) has *confidence c* in the transaction set D, where *c* is the percentage of transactions in D containing A that also contain B.

$$Confidence\ (A \rightarrow B) = P\ (B|A) \quad (4)$$

In general association rule mining can be thought of as a two step process:

(i) User defined minimum support must be considered to generate all item sets which is having *support* greater than or equal to it.

(ii) User defined minimum confidence must be considered to generate all the rules having the *confidence* factor greater than or equal to it.

V. Book Recommendation System

Purpose of this book recommendation system is to recommend books to the student according to their comfort price range and publishers. This recommendation system stores recommendation in student's web profile and works when user remains offline. This system has following steps:

1. From buyer profile records find out the category of the book that the buyer has bought earlier like CSE, ELECTRICAL, ECE, CIVIL etc.
2. Find out the subcategory of the book if there is any in the step 1 found category like for CSE subcategory will be C, C++, Data structure, Operating System etc.
3. Extract the profile of all buyers whose year of joining is less than target user's year of joining but not lesser than 5 years (senior students of target user). From the transaction database find all those transactions done by the above buyers and whose category and subcategory (if there is any) is same as found in step1 and step2.

   Apply association rule on those transactions and find out the books that the target buyer can buy afterward. Adjust the support and confidence parameters to get the stronger rules. This step gives the name of the subject which user is expected to buy and also gives the list of all books of given subject.
4. With the help of classification techniques find out the class of publisher's name and price range for each book, present in the List of books (output of step 3). The output of this step gives

publisher value(*Pb*) and price(*Pr*) for each book.

5. Perform user based collaborative filtering to find the rating(*R*) of the target user to the books found in step 3. For selecting the nearest neighbors to the target user, consider the profile of all buyers whose year of joining is less than the target user's year of joining but not lesser than 5 years (senior students of target user). This step will give the predicted rating(*R*) by target user to the books (list of books from output of step 3).
6. Now, apply mathematical formula to get list of books for recommendation.

$$D = {}^{10}/_{Pr} + Pb + R \quad (5)$$

Where, *Pr* implies price of the book, *Pb* implies the publisher value, *R* implies the predicted rating for this book. D is a variable. We, recommend top 5 books according to the descending values of *D*. Book having highest value with *D* will be recommended first.

From table 1, we can see values of *D* for different books and Book_1 will be placed in first rank for recommendation. Similarly this calculation can be done for many numbers of books.

This book recommendation system is represented by block diagram in the Fig. 1.

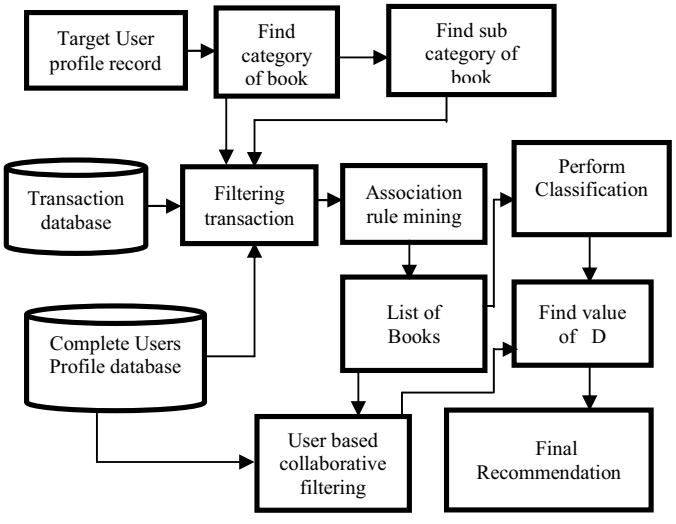

Fig.1. Block diagram of Book recommendation system

TABLE I. EXPLANATION WITH EXAMPLES

Name of book	Predicted rating(*R*)	Price (*Pr*)	Publisher value (*Pb*)	*D*
Book_1	5	500	0.5	5.52
Book_2	5	300	0.4	5.43
Book_3	5	600	0.2	5.22

VI. CONCLUSION

Recommender system has become a crucial tool in E - commerce on the web. The main job of recommender system is to help users find items they want to buy from a large volume of items. This book recommendation system is basically for students and academician who use textbooks. In this paper, we presented a book recommendation system which recommends books to the users according to their price range and preferred publishers, So that a user will get a book under his/her budget and familiar publisher. In this recommendation system we considered user based collaborative filtering which helped in getting the books of good quality, association rule mining which helped in filtering the transaction to find stronger recommendation and classification technique used in forming various classes of price range and publishers. This recommendation system works in offline mode, so it does not have any performance problem.

REFERENCES

[1] Shahab Saquib Sohail, Jamshed Siddiqui, Rashid Ali, "Book recommendation system using opinion mining technique", *International Conference on Advances in Computing, Communications and Informatics (ICACCI),2013.*

[2] Kesavaraj, G. ; Sukumaran, S. , " A study on classification techniques in data mining", *International Conference on Computing,Communications and Networking Technologies(ICCCNT),2013.*

[3] Li Ya Li ; Pan Shan Liang ; Xuang Xi " A Kind Of Web Service Recommendation Method Based On Improved Hybrid Collaborative Filtering", *International Conference on Cognitive Informatics & Cognitive Computing(ICCI*CC),2012.*

[4] Mu-HsingKuo , Liang-Chu Chen b, Chien-Wen Liang, "Building and evaluating a location-based service recommendation system with a preference adjustment mechanism", in *Expert system with applications* vol.36 ,issue 2 , part 2 pp.3543-3554 , march 2009.

[5] Guangyao, C.,"Interest mining and recommender system for local activities"*IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT '08), 2008.*

[6] Herlocker. Jonathan L., Konstan.Joseph A., et al,"Evaluating Collaborative Filtering Recommender Systems", ACM Transactions on Information System, New York, 22 (1),pp.5-53, Jan.2004.

[7] Hofmann T." Collaborative filtering via Gaussian robabilistic latent semantic analysis", Proceedings of the ACM Conference on Research and Development in Information Retrieval. Canada: ACM Press, pp.259-266,2003.

[8] Qiang Ding , Qin Ding, William Perrizo,"Decision Tree Classification of Spatial Data Streams Using Peano Count Trees" ,SAC 02:413-417,2002.

[9] Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. 2001."Item-based collaborative filtering recommendation algorithms". In *Proceedings of the 10th International Conference on World Wide Web(WWW'01)*. ACM, New York, NY, pp.285–295,2001.

[10] J. Han, M.Kamber, Data Mining : Concepts and Techniques, The Morgan Kaufmann Series, 2001.

[11] Resnick, P.,and Hal, R. V., 1997. Recommender Systems,Communications of the ACM, 40, 3, pp. 56-58,1997.

[12] Shardanand, U. ; Maes,P., Social information filtering:algorithms for automating "word of mouth". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems(CHI'95)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, pp. 210–217,1995.

[13] Agrawal, R., Imielinski, T., Swami, A. N. "Mining association rules between sets of items in large databases". *In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data,* pp.207-216, 1993.

.