Applied Mechanics and Materials Vol 39 (2011) pp 540-544 *Online: 2010-11-11*

doi:10.4028/www.scientific.net/AMM.39.540

Personalized Recommendation System Based on Association Rules Mining and Collaborative Filtering

Songjie Gong

Zhejiang Business Technology Institute, Ningbo 315012, China

Email: songjie_gong@sina.com

Keywords: personalized service, recommender systems, association rules mining, collaborative filtering, sparsity

Abstract. With the rapidly growing amount of information available, the problem of information overload is always growing acute. Personalized recommendations are an effective way to get user recommendations for unseen elements within the enormous volume of information based on their preferences. The personalized recommendation system commonly used methods are content-based filtering, collaborative filtering and association rule mining. Unfortunately, each method has its drawbacks. This paper presented a personalized recommendation method combining the association rules mining and collaborative filtering. It used the association rules mining to fill the vacant where necessary. And then, the presented approach utilizes the user based collaborative filtering to produce the recommendations. The recommendation method combining association rules mining and collaborative filtering can alleviate the data sparsity problem in the recommender systems.

Introduction

With the development of the Internet, the problem of overload to increase with information seriously. We all feel overwhelmed by the experienced number of new books, articles, and the process coming out of each year. Many researchers can more attention on building a suitable tool to help users get personalized resources. Personalized recommendations are such a software tool in the information filtering, collaborative filtering and data mining techniques are used to help users get recommendations for invisible elements on their preferences[1].

Most systems personal recommendation adopted three types of techniques, content filtering, collaborative filtering and association rule methods to recommend their products to customers[2,3]. The first content-based filtering, attempts are similar to those elements has a specific users recommended in the past you. It is based on a comparison between their content and a user profile. The second approach is called collaborative filtering, as a user whose tastes are similar to those of the specified user and recommends items they wanted. As a number of products, users can express their opinions of the objects they have tried before. The recommender can then compare the user's ratings with those of other users may have to "search for similar most" users on some criteria of similarity and then to recommend that users of similar products in the past. Scores for unseen items based on a combination of the values from the nearest neighbors predicted known. The third method of Association Rules, you get the recommendation by the association rules mining.

A personalized recommendation method combining the association rules mining and collaborative filtering was presented in this paper. It used the association rules mining to fill the vacant where necessary. And then, the presented approach utilizes the user based collaborative filtering to produce the recommendations. The recommendation method combining association rules mining and collaborative filtering can alleviate the data sparsity problem in the recommender systems.

Employing Association Rules Mining to Smoothing

1. Association rules mining

Association rules mining is one of the most well studied mining methods in data mining. It serves as a useful tool for discovering correlations among items in a big database. It explores the likelihood

that when certain items are present, which other items also present in the same dealings. An association rule is a condition of the form X=>Y where X and Y are two sets of items. An interpretation of the association rule in a commerce trade situation is when a customer buys items in X, the customer will also buy items in Y.

The apriori is the important algorithm in the algorithms of association rules mining. The main idea of the apriori is scanning the database repeatedly. The most important step in mining association is the generation of frequent item sets. In apriori algorithm, most time is consumed for scanning the database repeatedly[4].

Let I = { i1 , i2 , ···, im } be a set of all items, where an item is an object with some predefined attributes. A transaction T = < tid , It > is a duple, where tid is the identifier of the transaction. A transaction database T consist s of a set of transactions. An itemset is a subset of the set of items.

Definition 1: An association rule takes the form X =>Y where X < I , Y < I , and X ∩Y = O. The support of the rule X =>Y in the transaction database is :

support (X =>Y) = | { T ：X ∪ Y ∪ T , T ∈ D} | / | D |

Definition 2: The confidence of the rule X =>Y in transaction database is :

confidence (X => Y) = | { T ：X ∪ Y ∪ T , T ∈ D} | / | { T ：X < T , T ∈ D} |

2. Mapping user-item Matrixes to Transactions

Collaborative filtering user-item ratings data are usually represented as a preference matrices. They are changed in transactional databases for association rules mining. Each transaction is a transaction identity and content. The transaction identity TID, the user is the user ID belongs to the transaction. The content is the item ID and evaluations of the elements that have been rated by that user.

3. Algorithm

The Apriori algorithm calculates the frequent item sets in a database using many repeated iterations. All the frequent item sets calculated in the ith iteration are called k item sets. Each iteration consists of two steps: generating the candidate item sets, and calculating and choosing the candidate item sets. Its kernel thought is as follows[5]:

(1) L1 = {Large 1-Item };
(2) for (k = 2; kk - 1 ! = 0; k + +)
(3) Ck = Apriori-gen (Lk - 1) ;
(4) for all transaction t∈D do begin
(5) Ct = SubSet (Ck , t) ;
(6) for all candidates c∈Ct do
(7) c. count + +;
(8) end
(9) Lk = { c∈Ck };
(10) end
(11) UkLk
(12) end

The essence of Apriori algorithm is that all the non empty subitems of frequent itemsets must be frequent. It covers two steps: conjunction and pruning.

Producing Recommendation Employing User-based Collaborative Filtering

1. Similarity weighting

There are many similarity algorithms that have been used in the collaborative filtering recommendation systems [6,7].

Pearson's correlation, as following formula, measures the linear correlation between two vectors of ratings.

$$sim(x,y)=\frac{\sum_{c\in I_{xy}}(R_{xc}-A_x)(R_{yc}-A_y)}{\sqrt{\sum_{c\in I_{xy}}(R_{xc}-A_x)^2\sum_{c\in I_{xy}}(R_{yc}-A_y)^2}}$$

Where Rxc is the rating of the item c by user x, Ryc is the rating of the item c by user y, Ax is the average rating of user x for all the co-rated items, and Ixy is the items set both rating by user x and user y.

The cosine measure, as following formula, looks at the angle between two vectors of ratings where a smaller angle is regarded as implying greater similarity.

$$sim(x,y)=\frac{\sum_{k=1}^{n}R_{xk}R_{yk}}{\sqrt{\sum_{k=1}^{n}R_{xk}^2\sum_{k=1}^{n}R_{yk}^2}}$$

Where Rxk is the rating of the item k by user x, Ryk is the rating of the item k by user y and n is the number of items co-rated by both users.

The adjusted cosine, as following formula, is used in some collaborative filtering methods for similarity among users where the difference in each user's use of the rating scale is taken into account.

$$sim(x,y)=\frac{\sum_{c\in I_{xy}}(R_{xc}-A_c)(R_{yc}-A_c)}{\sqrt{\sum_{c\in I_{xy}}(R_{xc}-A_c)^2*\sum_{c\in I_{xy}}(R_{yc}-A_c)^2}}$$

Where Rxc is the rating of the item c by user x, Ryc is the rating of the item c by user y, Ac is the average rating of user x for all the co-rated items, and Ixy is the items set both rating by user x and user y.

2. Selecting neighborhoods

Select of the neighbors who will serve as recommenders. We employ the top-n technique in which a predefined number of n-best neighbors selected.

3. Producing a prediction

The rating of the target user p to the target item q is as following:

$$P_{pq}=A_p+\frac{\sum_{i=1}^{c}(R_{iq}-A_i)*sim(p,i)}{\sum_{i=1}^{c}sim(p,i)}$$

Where Ap is the average rating of the target user p to all items which rated, Riq is the rating of the neighbour user i to the target item q, Ai is the average rating of the neighbour user i to the items, sim(p, i) is the similarity of the target user p and the neighbour user i, and c is the number of the neighbours.

Dataset and Metrics

1. Dataset

MovieLens records collected by the GroupLens research project at the University of Minnesota [8]. The historical data set consists of 100,000 ratings from 943 users on 1682 movies by any user with at least 20 reviews and simple demographic information for users is included. Therefore, the lowest level of the sparseness of the tests is defined as 1 - 100000 / 943 * 1682 = 0.937.

The ratings are on a numeric five-point scale with 1 and 2 representing negative ratings, 4 and 5 representing positive ratings, and 3 indicating ambivalence. We randomly divided 20% of the experiment data set as test data set and the rest were set as training data set.

2. Metrics

The metrics for evaluating the accuracy of the prediction algorithm can be divided into two main categories [9,10]: statistical accuracy metrics and decision-support metrics. Statistical accuracy metrics evaluate the accuracy of a prediction by comparing predicted values with user provided values. Decision-support accuracy measures, as well predictions help users select high-quality products.

Decision support accuracy metrics evaluate how effective a prediction engine to help a user select high-quality items from the set of all elements. The Receiver Operating Characteristic (ROC) sensitivity is an example of the decision to support accuracy metrics. The metric indicates how effectively the system, the user in the direction of highly valued and to steer away from products with lower-rated ones. Suppose x1, x2, x3, ..., xn the prediction of user ratings, reviews and the corresponding real data set of users y1, y2, y3, ... yn. View the ROC-4 definition as follows:

$$ROC\text{-}4 = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} v_i}$$

$$u_i = \begin{cases} 1, & x_i \geq 4 \ \ and \ \ y_i \geq 4 \\ 0, & otherwise \end{cases}$$

$$v_i = \begin{cases} 1, & x_i \geq 4 \\ 0, & otherwise \end{cases}$$

The larger the ROC-4, the more accurate the predictions would be, allowing for better recommendations to be formulated.

Summary

As the rapidly increasing amount of information available, the problem of information overload is always growing acute. Personalized recommendations are an effective way to get user recommendations for unseen elements within the enormous volume of information based on their preferences. The personalized recommendation system commonly used methods are content-based filtering, collaborative filtering and association rule mining. Unfortunately, each method has its drawbacks. In this work we presented a method combining a personal recommendation for the Association Rules mining and collaborative filtering. It used to fill the empty mining association rules where necessary. And then uses the presented approach, the user based collaborative filtering to produce the recommendations. The recommended method combines Association Rules mining and collaborative filtering can alleviate the data sparsity problem in recommender systems.

Acknowledgment

A Project Supported by Scientific Research Fund of Zhejiang Provincial Education Department (Grant No. Y200909659).

References

[1] Gao Fengrong, Xing Chunxiao, Du Xiaoyong, Wang Shan, Personalized Service System Based on Hybrid Filtering for Digital Library, Tsinghua Science and Technology, Volume 12, Number 1, February 2007,1-8.

[2] Yi-Fan Wang, Yu-Liang Chuang, Mei-Hua Hsu, Huan-Chao Keh. A personalized recommender system for the cosmetic business. Expert Systems with Applications 26 (2004) 427–434.

[3] George Lekakos, George M. Giaglis, Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors, Interacting with Computers 18 (2006) 410–431.

[4] LI Pingxiang, CHEN Jiangping, BIAN Fuling, A Developed Algorithm of Apriori Based on Association Analysis, Geo-spatial Information Science, Volume 7, Issue 2, 2004

[5] TAN Ying, YIN Guofu, LI Guibing, CHEN Jianying, Mining Compatibility Rules from Irregular Chinese Traditional Medicine Database by Apriori Agorithm. Journal of Southwest Jiaotong University (English Edition) Vol.15, No.4, 2007

[6] M.G. Vozalis, K.G. Margaritis, Using SVD and demographic data for the enhancement of generalized Collaborative Filtering, Information Sciences 177 (2007) 3017–3037.

[7] Yu Li, Liu Lu, Li Xuefeng, A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce, Expert Systems with Applications 28 (2005) 67–77.

[8] George Lekakos, George M. Giaglis, A hybrid approach for improving predictive accuracy of collaborative filtering algorithms, User Model User-Adap Inter (2007) 17:5–40.

[9] Breese J, Hecherman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98). 1998. 43~52.

[10] Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information tapestry. Communications of the ACM, 1992,35(12):61~70.

Quantum, Nano, Micro and Information Technologies
10.4028/www.scientific.net/AMM.39

Personalized Recommendation System Based on Association Rules Mining and Collaborative Filtering
10.4028/www.scientific.net/AMM.39.540

DOI References

[2] Yi-Fan Wang, Yu-Liang Chuang, Mei-Hua Hsu, Huan-Chao Keh. A personalized recommender ystem for the cosmetic business. Expert Systems with Applications 26 (2004) 427–434.
doi:10.1016/j.eswa.2003.10.001
[3] George Lekakos, George M. Giaglis, Improving the prediction accuracy of recommendation lgorithms: Approaches anchored on human factors, Interacting with Computers 18 (2006) 410–431.
doi:10.1016/j.intcom.2005.11.004
[6] M.G. Vozalis, K.G. Margaritis, Using SVD and demographic data for the enhancement of eneralized Collaborative Filtering, Information Sciences 177 (2007) 3017–3037.
doi:10.1016/j.ins.2007.02.036
[7] Yu Li, Liu Lu, Li Xuefeng, A hybrid collaborative filtering method for multiple-interests and ultiple-content recommendation in E-Commerce, Expert Systems with Applications 28 (2005) 7–77.
doi:10.1016/j.eswa.2004.08.013
[8] George Lekakos, George M. Giaglis, A hybrid approach for improving predictive accuracy of ollaborative filtering algorithms, User Model User-Adap Inter (2007) 17:5–40.
doi:10.1007/s11257-006-9019-0
[2] Yi-Fan Wang, Yu-Liang Chuang, Mei-Hua Hsu, Huan-Chao Keh. A personalized recommender system for the cosmetic business. Expert Systems with Applications 26 (2004) 427–434.
doi:10.1016/j.eswa.2003.10.001
[3] George Lekakos, George M. Giaglis, Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors, Interacting with Computers 18 (2006) 410–431.
doi:10.1016/j.intcom.2005.11.004
[6] M.G. Vozalis, K.G. Margaritis, Using SVD and demographic data for the enhancement of generalized Collaborative Filtering, Information Sciences 177 (2007) 3017–3037.
doi:10.1016/j.ins.2007.02.036
[7] Yu Li, Liu Lu, Li Xuefeng, A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce, Expert Systems with Applications 28 (2005) 67–77.
doi:10.1016/j.eswa.2004.08.013
[8] George Lekakos, George M. Giaglis, A hybrid approach for improving predictive accuracy of collaborative filtering algorithms, User Model User-Adap Inter (2007) 17:5–40.
doi:10.1007/s11257-006-9019-0