

A Modified Approach towards Tourism Recommendation System with Collaborative Filtering and Association Rule Mining

Monali Gandhi
CE Department, Parul Institute Of
Engineering & Technology,
Vadodara, Gujarat, India,

Khushali Mistry
CE Department, Parul Institute Of
Engineering & Technology,
Vadodara, Gujarat, India,

Mukesh Patel
IT Department, Sarvajnik College
of Engineering & Technology,
Surat, Gujarat, India,

ABSTRACT

In the tourism recommendation system, the number of users and items is very large. But traditional recommendation system uses partial information for identifying similar characteristics of users. Collaborative filtering is the primary approach of any recommendation system. It provides a recommendation which is easy to understand. It is based on similarities of user opinions like rating or likes and dislikes. So the recommendation provided by collaborative cannot be considered as quality recommendation. Recommendation after association rule mining is having high support and confidence level. So that will be considered as strong recommendation. The hybridization of both collaborative filtering and association rule mining can produce strong and quality recommendation even when sufficient data are not available. This paper combines recommendation for tourism application by using a hybridization of traditional collaborative filtering technique and data mining techniques.

KEYWORDS

Collaborative filtering, Association rule mining, tourism, recommendation system.

1. INTRODUCTION

Data Mining is the method of identifying valid, novel, and useful patterns from huge amount of data. It is also refers as the process of extracting or “mining” knowledge from large amounts of data. It functionalities includes Data characterization, Data discrimination, Association analysis, Classification, prediction, Cluster analysis, Outlier analysis, Evolution analysis etc.; Discovering patterns from the data via Association rule mining techniques are widely used in numerous applications such as pattern recognition, market research, image processing and biological data analysis[1].

Second section of this paper we proposes a hybrid method of recommendation system. Third section gives the need for the new system. Fourth section gives the related work done for this recommendation system. In fifth section we had focused on proposed system. Finally in sixth section we had given comparison of the different methods of recommendation system by using datamining techniques.

Recommendation system is used to provide recommendations of interesting items in a wide variety of application domains such as web page recommendation, digital news, movie recommendation, travel agent and many others. A variety of approaches has been used to perform recommendations in the domains which includes collaborative, content-based, demographic and knowledge-based.

In this research paper, tourism recommendation system applies both content-based and collaborative approaches. TRS conducts personalized travel recommendation by considering specific user profiles or attributes (eg. Age, gender, race, personal, professional) as well as travel group types (eg. Family group, couple). The system provides information about tourist places based on their similarity.

2. RECOMMENDATION TECHNIQUES

Collaborative filtering (CF) [3][8] is the most common method in the recommendation system. CF uses customer profile and user preferences for extracting the users according to their similarity. This approach is based on theory that “Users with common interests in the past will have similar behaviors in the future” [2].

CF method can generate recommendation based on the following information like **Customers’ rating**[9]. In this approach, customer priorities are extracted from their behavioral patterns, navigations and their purchases are collected. The customer preferences in purchases are determined numerically.

The only disadvantage of CF method is that it is dependent on human ratings only. More than one user must evaluate each item even than new items cannot be recommended until one user has taken time to evaluate them. This is known as Scarcity problem.

Content-Based filtering[10] makes suggestion according to the past interest of the customer. Therefore, the items will be suggested to the customers which are very similar in contents and characters to his or her favorites items[3].

Demographic Based filtering uses user profile information such as age, gender, material status, personal details, professional details, postal-code and hence so forth[1]. The only drawback of this method is that it is time consuming process and if the users do not provide the personal information, it is not possible to build any profile for them[3].

Social-Based filtering propose a social recommendation that incorporates a social information into a user based collaborative filtering model. This recommendation system uses same formula as user-based technique. It combines social ratings, tagging and demographic information etc. It obtains good performance and results at the expense of low user coverage. Their performance is higher. Many authors proposed different algorithms related to different techniques.

Context-Aware filtering predicts a user’s preferences in different context situations based on past user’s experience.

The system uses that the other users had done in similar context to predict the user preferences towards an item in the current context. The current context has to be captured each item when the users make some choice. It has two problem first how to manage the context time to the users and how to measure the similarities between the contexts.

Hybrid Approach includes the recommendation of the all above methods which provides a novel approach towards dynamic hybrid recommendation system. This approach improve the shortcomings of each and every filtering technique.

3. NEED FOR A NEW SYSTEM

Association rule mining is a popular method for discovering interesting relations between variables in a large database. It is intended to identify strong rules found in the database. It is used to discover regularity between products in large –scale transaction data. It is very important rule.

But alone collaborative filtering cannot provide effective solution to the generated frequent item-sets. However association rule suffers from the shortcoming that the number of rules generated are based on the number of itemsets in a frequent itemsets.

To solve the above difficulties, this research uses hybridization of collaborative filtering and association rule mining is used to improve the quality of recommendations. It applies collaborative filtering after the generation of rules before providing recommendation to the user.

4. LITERATURE SURVEY

In [2] Masoumeh Mohammadnezzhad and Mehregan had reviewed different papers. They had proposed a method for recommendation system which uses only collaborative filtering. The data mining techniques which they used are clustering and association rule mining. The number of clusters is created by K-means algorithm. Recency, Frequency and Measure parameters were not used for collaborative filtering technique which does not give any accurate suggestion to the customers. The precision of the recommendation was also very low.

In [3] Masoumeh Mohammadnezzhad,, Mehregan Mahdaviand Guilan has proposed a recommendation method for the large number of users and items for identifying the similar users. The objective of this paper is to improve the quality of recommendation and to provide strong recommendation to the users. In this article they had presented two methods of recommendation which is collaborative filtering and content-based filtering. They had used data mining techniques such as clustering and association rule mining. This model has four phases, at first

tourists are clustered based on their location. In second phase a two level graph model is used to show the similarity between the tourists interests and the similarity of the tours. Finally, recency, frequency and measure parameters are used to provide suggestion to the users. According to the experimental result, the standard F-measure indicates that the quality of the recommendation is higher than the traditional approaches.

In [4] Keunho Choi, Donghee Yoo, Gunwookim and Yongmoo Suh has proposed a method by taking an example of online shopping mall in which explicit rating information is not available. This poses a problem in providing recommendation services using collaborative filtering techniques for their users. Sequential pattern analysis provides recommendation to the users with less accuracy. This article proposed a scheme for providing implicit rating that can be applied to the online transaction. The combined approach of CF and SPA can be used to provide quality recommendation to the customer by using explicit rating and the hybrid approach proves to be better one.

In [5] Yan-Ying Chen, An-Jung Cheng, and Winston H. Hsu had proposed a method to personalized a travel recommendation method by using specific user profile or attributes like age, gender and race as well as travel group types like family, friends and couple. They had exploited the detected people attributes and travel group types in photo contents. They had used probabilistic Bayesian learning framework which is used as a part of mobile recommendation on the spot. They had conducted experiment on more than 10 million photos. The experiments has confirm that people attributes of individuals and groups are promising and orthogonal to prior works using travel logs only and can further improve prior travel recommendation methods especially for difficult predictions by further leveraging user contexts via mobile devices.

In [6] Joel P. Lucas , Nuno Luz , María N. Moreno Ricardo Anacleto , Ana Almeida Figueiredo , Constantino Martins proposed a recommendation method by using the hybrid approach of collaborative filtering and content-based filtering .The data mining techniques which is used is classification based on association is applied. It is also known as associative classification method which can combine the concept of classification and association [7] They had shown comparison method which had improved the quality of recommendation instead of using only eitherassociation or classification. The datasets which they had used are the fuzzy logic datasets ranging between 0 to 1. Memory based technique which is also known as nearest neighbor technique has been introduced for the first time to select the top n items. The comparison of all the techniques is shown in the Table 2.

Table 2: Comparison of different techniques

Author	Advantages	Disadvantages
MasoumehMohhamma and Mehregan[2]	<ul style="list-style-type: none"> Effective solution to tourists interest and preferences. Cold start problem can be addressed. 	<ul style="list-style-type: none"> RFM parameters are not used. Not precision recommendation is provided and also no graph model is provided.
Guilan[3]	<ul style="list-style-type: none"> Reliable to high extent as it improves RFM parameters. 	<ul style="list-style-type: none"> Time consuming for large databases.
Keumho Choi and Yongmoosuh[4]	<ul style="list-style-type: none"> Item rate technique is used Provide implicit knowledge 	<ul style="list-style-type: none"> Less efficiency in terms of support and confidence.
Winston Shu and Yan-ying Chen[5]	<ul style="list-style-type: none"> Travel group type predictions can be made. 	<ul style="list-style-type: none"> Very high false probability for high recommendations.
Joel Lucas, Nuno and Maria[6]	<ul style="list-style-type: none"> Users can make recommendation on the basis of the past history as well as user's behaviour. 	<ul style="list-style-type: none"> False negative and false positive for low reliability.

5. PROPOSED SYSTEM

Association rule mining leads to the discovery of patterns and correlations among items in large transactional or relational data sets. With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining such patterns from their databases. Association rule mining consists of the following step [1]:

Step 1: Find all frequent itemsets.

Get frequent items:

- Items whose occurrence in database is greater than or equal to the min.support threshold.

Get frequent itemsets:

- Generate candidates from frequent items.
- Prune the results to find the frequent itemsets.

Step 2: Generate strong association rules from frequent itemsets. Rules which satisfy the min.support and min.confidence threshold.

This work also proposes the working of Collaborative Filtering. Collaborative Filtering consists of two major step:

Step 1: A user expresses his or her preferences by rating items (e.g. books, movies or CDs) of the system. These ratings can be viewed as an approximate representation of the user's interest in the corresponding domain.

Step 2: The system matches this user's ratings against other users' and finds the people with most "similar" tastes.

Step 3: With similar users, the system recommends items that the similar users have rated highly but not yet being rated by this user (presumably the absence of rating is often considered as the unfamiliarity of an item).

New model has been proposed as the final choice on recommender system approach depends on the information sources and objects-of-interest, which are to be used in the system. Some of these sources of information are easy to obtain and maintain, others involve more cost and effort. In fact, this choice is the main determining factor of a recommender system. As we have seen, all basic recommending approaches are applicable to the Tourism domain. Moreover, the heterogeneity of this domain favours the use of hybrid recommenders, which is another indication that all approaches can produce valuable contributions to the inference process.

The main modules of the work can be divided into four phases. In the first phase the database pre-selection is done. In the second phase data mining technique Association rule mining is used. In third phase Collaborative filtering and content-based filtering is done. Finally, select Top-N items algorithm is applied to provide recommendation to the user. The details of the module can be shown as:

Database pre-selection provides an initial selection of items, based on simple database interactions (similar to rule-based filtering using boolean logic). This yields a very efficient reduction of the number of items, which have to be processed, at a very early stage of the workflow.

Association Rule mining methods can now either provide initial item ratings (using fuzzy logic), or perform a further reduction of the item set size (using boolean logic). The usage of knowledge-based filtering allows developers to make use of explicit domain knowledge.

We can make use of **collaborative filtering method** or virtually any other form of item rating technique to obtain one or more numerical ratings for every item. This allows us to incorporate algorithmic or implicit domain knowledge.

By selecting the **N best scoring items**, we obtain a final set of recommendations as shown in figure 2 in my proposed work. The basic steps of the algorithm are as follows:

Step 1: Data generation which is an additional step to store the attributes of the customer profile.

Step 2: Apply adaptive Apriori algorithm for generating multistage itemset recommendation.

Step 3: Run complete Apriori for generating strong association rules in customer database having wide range of attributes which are essential or desirable.

Step 4: Every customer has to rate the package system on the basis of profile similarity as well as rating the system or ranking it.

Step 5: Now apply strong association rule again in customer database on the wide variety of attributes.

Step 6: Apply collaborative filtering algorithm for selecting the Top-N items.

Step 7: After applying collaborative filtering user opinions and user preferences will be considered and recommended to the another customer.

After implementation of proposed system it has expected that proposed model is certainly come up with good results. This new approach is expected to improve the performance of the system by providing strong recommendation to the customer on the basis of the customer profile and user similarity which are both essential and desirable.

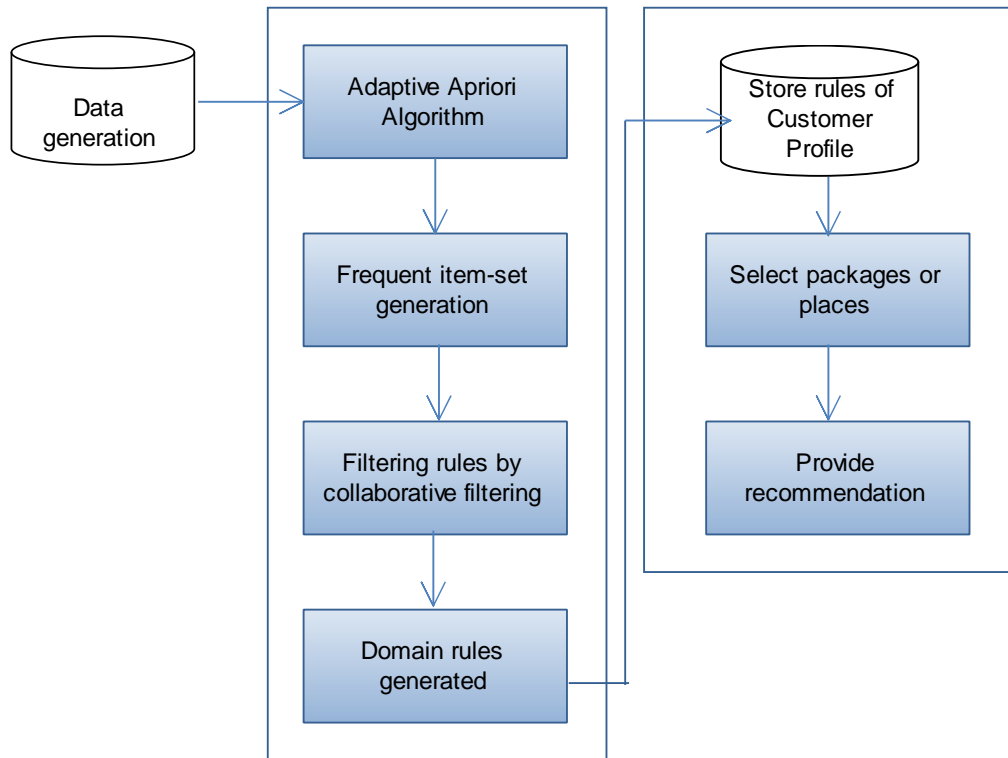


Figure 2: Architecture of proposed system

6. CONCLUSION

In this paper we have described the different methods of recommendation. We had also discussed the different algorithms for recommendation system and collaborative filtering. Here we have aimed to improve the quality of recommendation and to provide strong recommendation to the users.. All of these techniques have their own advantages and disadvantages which are mentioned in the paper. The solutions presented here target to one or other parameters to improve the efficiency. It's very necessary to further improve the performance of sparsity which would poses a hindrance to the collaborative filtering technique. Moreover, it can also help to design a new and more powerful hybrid architecture for providing strong recommendation to the users.

7. FUTURE WORK

The future work includes the experimentation on more people attributes and provide strong association domain rules for such diverse attributes. New lines of research will be developed for fields and aims such as proper combination of existing

recommendation method that uses different types of available information. Data mining from recommendation system database for non-recommendation uses areas such as market research, general trends, visualization of differential characteristic of demographic groups. We can also expand our model for more contexts such as travel duration and travelling season. Lastly we compare the results obtained with the results of the previous techniques to differentiate the performance of our proposed solution and the existing solutions.

8. REFERENCES

- [1] Jiawei Han and Micheline Kamber, "Data Mining Concepts &Techniques",Elsevier,2011.
- [2] Masoumeh Mohammad and Mehregan Mahdavi, IJITCS Intelligent Systems, Vol. 21, No. 1, pp.35-41, IJITCS 2012..
- [3] Adomavicius, G., Tuzhilin, A" Toward the next generation of recommender systems: asurvey of the state of-the-art and possible extensions", IEEE Transactions on

- Knowledge and Data Engineering, Vol.17, No. 6, pp. 734-749, IJITCS 2012 Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.
- [4] Aggarwal, C. C., Procopiuc, C., and Yu, P. S.” Finding localized associations in market basket data. IEEE Transactions on Knowledge and Data Engineering”, 14, 1, 2002 ,pp.51-62, ELSEVIER 2012. Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.
- [5] Yan Ying Chen, An-Jung Cheng and Winston H.Hsu” IEEE Transactions on Multimedia”, 15,6,pp,1283-1288,IEEE Oct 2013.
- [6] Lee, C.-H., Kim, Y.-H., & Rhee, P.-K..” Web personalization expert with combining collaborative filtering and association rule mining technique”. Expert Systems and Applications, 21(3), 131–137,ELSEVIER 2013.
- [7] Agrawal, R., Imielinski, T., & Swami, A.” Mining association rules between sets of items in large databases. In P. Buneman& S. Jajodia (Eds.),” pp.207–216, ELSEVIER 2013.
- [8] Ricci, F., Rokach, L., Shapira, B. (2011) *Recommender Systems Handbook*, Springer, ISBN 978-0-387-85819-7, pp. 1-184.
- [9] Liangxing, Y., Aihua, D. (2010) *Hybrid Product Recommender System for Apparel Retailing Customers*, In proceeding ICIE '10 Proceedings of the 2010 WASE International Conference on Information Engineering, Washington, DC, USA.
- [10] Banati, H., Mehta, S. (2010) *Memetic Collaborative filtering based Recommender System*, Second Vaagdevi International Conference on Information Technology for Real World Problems, Warangal, India.
- [11] Salter, J., Antonopoulos, N.” CinemaScreen recommender agent: Combining collaborative filtering and content-based filtering”, IEEE Intelligent Systems, Vol. 21, No. 1, pp.35-41, IJITCS 2012.