An Enhanced Approach for Tourism Recommendation System using Hybrid Filtering and Association Rule Mining

Monali Gandhi
CE/IT Department, Chottubhai
Gopalbhai Patel Institute of
Technology, Gujarat, India,
monali.gandhi@utu.ac.in

Sonali Gandhi
CS department, Pace junior
Science Collage, Borivalli,
Maharashtra, India,
sonali.gandhi6@gmail.com

*A*bstract-- 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 and content based 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 and content based filtering is used to provide opinion for the new users profile. So the recommendation provided by collaborative and content cannot be considered as quality recommendation. Recommendation after association rule mining is having high support and confidence level. So that it will be considered as strong recommendation. The hybridization of both hybrid 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 and content filtering techniques and data mining techniques.

KEYWORDS
Collaborative filtering, content based 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 Associationrule mining techniques are widely used in numerous applications such as pattern recognition, marketresearch, image processing and biological dataanalysis[1].

In second section of this paper we had proposed on relative work of recommendation system. Third section focuses on the CBF algorithm which is the existing approach..Fourth section gives the brief description of the proposed method which is named as Recommendation based on Collaborative Filtering Association(RCFA).Finally in fifth section we had given comparison of the CBF method

with RCFA method of recommendation system by using data mining 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 proposed work, tourism recommendation system applies collaborative filtering and content based filtering approach with association rule mining as data mining techniques.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. Related work

In [2] MasoumehMohammadnezzhad 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 are 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]MasoumehMohammadnezzhad,,MehreganMahdaviand 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, DongheeYoo, Gunwookim and YongmooSuh 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.

3. CBF algorithm

CBF algorithm, Content-Based filtering algorithm uses similarity of user behavior and profile similarity for providing recommendations to the user by using Apriori algorithm as the data mining technique. This requires the pre-processing step for carrying out the efforts for input data. To avoid the pre-processing step, this algorithm also includes discretization and fuzzification process for continuous attributes. It is recommended to set a high value for a confidence and the low value for the support. Here all the rules are generated which is used to classify the new user while the rule ordering scheme is not taken into account. Finally, the content-based filtering process is applied regarding only to the current user or visitors of the system. The recommendation is provided by using the top N places and by implementing the nearest neighbor algorithm.

Definition 1.(Class association rule configuration).
Condset ➤ y. Here condset is the set of conditions having descriptive attributes in the dataset and y is the condition related to the values of the label attribute in all the possible cases .An example for this can be given by using the following definition as: {(att1=a) **AND** (att2=b)} (class=C1), where "a" and "b" are the instances of the given attributes "att1" and "att2".

Definition 2. (Confidence measure).
Conf(A,B)=support(A,B)/support A.
where is the measure that access the frequency that the items of rule occurs in the dataset. Confidence is the measure that expresses the correspondence between items framing the rules. It is expressed by the occurrence of frequency of the rule among all the transactions containing the predecessor part.

CBF algorithm(D, C_K, $Freq_k$)

$C_K = 0$

For each unvisited place P in dataset D

    Mark P as visited

  Freq1 = {large 1-itemsets};
    CR1 = genRules(Freq1);
    prCR1 = pruneRules(CR1);
k = 2;
while {Freqk-1 / ø)
  Ck = canclidateItemsetsGen(Freqk_1);
for all (data case Di such that Di _ Df)
   Cd = ruleSubset(Ck, d);
for all (candidateItemset Ci such that Ci _ Cd)
  if (Di.class = Ci.class)
for all (attribute a such that a e Di)
  lineSupport = lineSupport _ a.support;
end for
Ci.rulesupCount = Ci .rulesupCount + lineSupport;
    end if
       end for
end for
Freqk = {c e Ck|c.rulesupCount > minsup};
CRk = genRules (Freqk);
prCRk = pruneRules(CRk);
k++
end while
CRs = [k CRk];
prCRs = [k PrCRk];

The time complexity of the CBF method is $O(n^2)$, where n is the number of places in the dataset; With the support of spatial access methods such as R-tree, its time complexity can be reduced to $O(n \log n)$.

4. Proposed System

In addition to CBF, the following definitions are required in RCFA (Recommendation based on Collaborative Filtering Association) which provides qualitive and strong recommendation to the visitors based on the past history of the visitor. Lastly, the collaborative filtering process is applied by considering the rating of the current user as well as of the past visitor.

Definition 1:-As the implicit rating method is used to provide recommendation, the absolute performance of

visitor v on places p, AP(v,p) is defined by the following equation:

AP(v,p)=In(The number of transaction of visitor v including place p/ The number of transaction of visitor of visitor v.) (1)

Definition 2:- Support(A,B)=P(A U B)

Here support refers to the probability of A and B which appears in the data set D at the same time. Support is divided by the total number of transactions which is called as support of the item sets. (2)

Input : Data set D
Minimum places required to provide recommendation
Rating is required to find the actual confidence level
Output: No of recommendation of places
Algorithm RCFA (D,C_K, $Freq_K$)

1. Initially all places are read from dataset
2. For each place find the maximum item set
3. If(ouput.rules.count>0)
4.   For each (apriori.rule in output rule)
5.   Add max items(rule X+ rule Y)
6. End if
7. If strong rules >0
8. For each (apriori.rule in strong rule)
9. Double rating =0.0
10.   If (rule.Y.length==1)
11.   Double newconf=(rule.conf*100+rating)
12.   Else
13.   Double new conf=(rule.conf*100+rating)/2
14.   End if
15.   Generate filtering(output)
16.   Fill recommend()

5. Performance Analysis

To judge against the performance of the proposed algorithm, we have also implemented the well acknowledged CBF algorithm as well as RCFA algorithm. .NET is used as a language to implement the algorithms. The performance of the above two algorithms are evaluated by using the 2-Dimensional synthetic dataset in .xml file format. The 2-Dimensional synthetic dataset is containing varying visitors and places in 2-Dimensional plane. We had compared the confidence level of both algorithms as well as the filtering process of both algorithms which is used to recommend places.

Table 1. Confidence level of CBF & RCFA

Dataset	Visitors	Places Encoded	CBF	RCFA
			Level of Confidence	
	13	13	50	60
TRS Small	26	13	55	60
	50	26	67.5	77.5
TRS	100	26	68.5	77.5

Medium				
	300	26	78.5	80
TRS Large	500	52	79.5	80

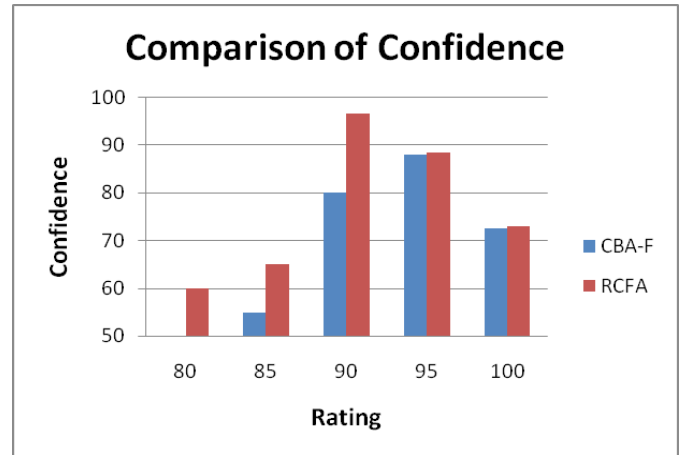**Figure shows the comparison of CBF & RCFA**

Table 2. Recommended places- CBF & RCFA

Dataset	Visitors	Places Encoded	CBF	RCFA
			Filtering for providing recommendation	
	13	13	4	2
TRS Small	26	13	3	3
TRS Medium	50	26	5	3
	100	26	6	4
	300	26	6	4
TRS Large	500	52	6	4

6. Conclusion and Future work

In this paper we have described two different methods of recommendation. We had also discussed the two different algorithms for recommendation system . Here we have aimed to improve the quality of recommendation and to provide strong recommendation to the users.. These two techniques have their own advantages and disadvantages . 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.

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 had compared the results obtained with the results of the previous techniques to differentiate the performance of our proposed solution and the existing solutions.

7. References

[1] Jiawei Han and MichelineKamber ,"Data Mining Concepts & Techniiques",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 2012Tavel, 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 RetailingCustomers*, In proceeding ICIE '10 "Proceedings of the 2010 WASE International Conference onInformation 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 WorldProblems, Warangal, India.

[11] Salter, J., Antonopoulus, N." *CinemaScreen recommender agent: Combining collaborative filtering and content-based filtering"*, IEEE Intelligent Systems, Vol. 21, No. 1, pp.35-41, IJITCS 2013.

[12] Shaw, Geva, S. "*Investigating the use of association rules in improving the recommender system*" Proc. 14th Australasian Document Computing, Sydney, Aussssstralia.

[13] Delic A, Neidharat J,Nguyen TN. "Research methods for group recommendations system" Proc. Recommendation tour 2016, Vol 1685.

[14] Imen Akermi, Mohand Boughanmen Rim Faiz Une approach de recommendation proactive dans unenvironment mobile, 2016.