IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. VII (Jul.-Aug. 2017), PP 12-19
www.iosrjournals.org

A Hotel Recommendation System Based on Data Mining Techniques

Ravindra Ramawat, Mr. B. L. Pal
Dept. of Computer Science & Engineering Mewar University,Chittorgarh(Raj.), India
Dept. of Computer Science & Engineering Mewar University,Chittorgarh(Raj.), India
Corresponding Author: Ravindra Ramawat

Abstract: *The recommendation system is a software tool to recommend user about the quality of items or services to be use by the user. Due to increase of tourism industry as well as growth IT industry tourist or employee visiting from one place to another place frequently and staying in hotel for short period. So, numbers of researchers have design and developed hotel recommendation system for these types of visitors using different – different type of algorithms like memory based algorithm, model based algorithm and Association rule. Each and every algorithm has their own prose and cons. That is why we have design and developed a new recommendation system for the hotel booking using K-Means algorithm, Apriori algorithm and association rule.*
Keywords: *K-means Clustering, Apriori Algorithm, hotel, Association rule, visitors etc.*

Date of Submission: 01-07-2017 Date of acceptance: 26-08-2017

I. Introduction

In the present era life is fast and competitive. Each and every one taking suggestion from others to know the relevant item in day to day life i.e. also a one type recommendation system also. But, in computer era, each and every one has smart phone and taking help of search engine to know the actual facts and figure. Due to overloaded information regarding any item or service provided by any agency, we require web based recommendation system. There are three paradigm of web based recommendation system namely personalized recommendation, collaborative recommendation, content based recommendation, knowledge recommendation and hybrid based recommendation. All these recommendation systems using some types of recommendation system to recommend user like K-means algorithm, Association rule, Apriori algorithm etc.

Recommendation Algorithms

There are three types of recommendation algorithm as given below:

- Association Rule Based Recommendation System
- Model Based Recommendation System
- Memory Based Recommendation System

Association Rule Based Recommendation System

Association rules try to connect the causal relationships between items. An association rule essentially is of the form A1, A2, A3, ... => B1, B2, B3, ... It attempts to show how a series of items can determine another series of items.

Model Based Recommendation System

Model-based recommendation systems are based on the dataset of ratings. we extract some information from the dataset, and use that as a "model" to make recommendations without having to use the complete dataset every time. This approach potentially offers the benefits of both speed and scalability.

Memory Based Recommendation System

Memory-based algorithms approach the collaborative filtering problem by using the entire database.it tries to find users that are similar to the active user and uses their preferences to predict ratings for the active user.

II. Literature Review

- A Meenal Moghe, Abhishek Raghuvanshi, 2015 has proposed, "A Novel Collaborative Filtering Recommendation System Algorithm" to recommend item based on rating to solve the problems of sparsity [1]
- Qi Wang, Wei Cao and Yun Liu, 2014 has proposed "A Novel Clustering Based Collaborative Filtering Recommendation System Algorithm" using K-means clustering method and a rating matrix is constructed for the unrated items based on formula.[2].
- Mohammad Hamidi Esfahani, Farid Khosh Alhan, 2013 proposed "New Hybrid Recommendation System Based On C-Means Clustering Method" by combining the content based and collaborative based approach to recommend the new item [3].
- Hazem Hajj, Wassim El-Hajj, Lama Nachman, IEEE 2013 proposed "A Hybrid Approach with Collaborative Filtering for Recommender Systems" to estimate the ratings for the unrated items based on the weighted combination of user-based collaborative filtering and item-based collaborative filtering[4].
- Haojun Sun, 2012 proposed "A New Item Clustering based Collaborative Filtering Approach"to recommend the new items based on similar features of the users [5].
- Maria Teresa Andrade, Fábio Almeida, IEEE 2013 proposed "Novel Hybrid Approach to Content Recommendation based on Predicted Profiles" to overcome the over-specialization problem[6].
- Jens Grivolla, Toni Badia, Diego Campo, Miquel Sonsona, Jose-Miguel Pulido, 2014 proposed" A hybrid Recommender Combining User item and interaction data" which is very strong when little interaction data is available [7].
- Ahmad A. Kardan, Mahnaz Ebrahimi, 2012 proposed "A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups" which improves the accuracy of useful posts recommended to the users in comparison to content-based and the collaborative filtering techniques as well [8].
- Rahul Katarya, Dr. Om Prakash Verma ,Ivy Jain, 2013 proposed "User Behaviour Analysis in Context-aware Recommender System using Hybrid Filtering Approach" an algorithm for CARS for the movie recommender application. Which reduces the sparsity problem in the pre-filter method and at the same time it is quicker than the post-filter method[9].
- Qing Li, Byeong Man Kim, 2003 proposed "Clustering Approach for Hybrid Recommender System" which improves the correctness of collaborative predication [10].

III. Problem Statement

Due to lack of time or due to their attitude, each person using web based recommendation system instead of taking suggestion from family members or the expert of the respective field to purchase an item or to avail the services provided by some agencies. Another problem of present era is life is fast and competitive around the world. Person is frequently moving from one place to another far from family members to earn money or to visit some place. So, they require a place like guest house or hotel to stay there with home like environment and facilities. So, they require a recommendation system to overcome these types of problem.

IV. Objective

Main objectives are:

- To collect data regarding hotel rating.
- To prepare dataset for clustering using K-Means algorithm.
- To generate association rule using Apriori algorithm.
- To design and develop recommendation system.

V. Proposed Work

A. Proposed System

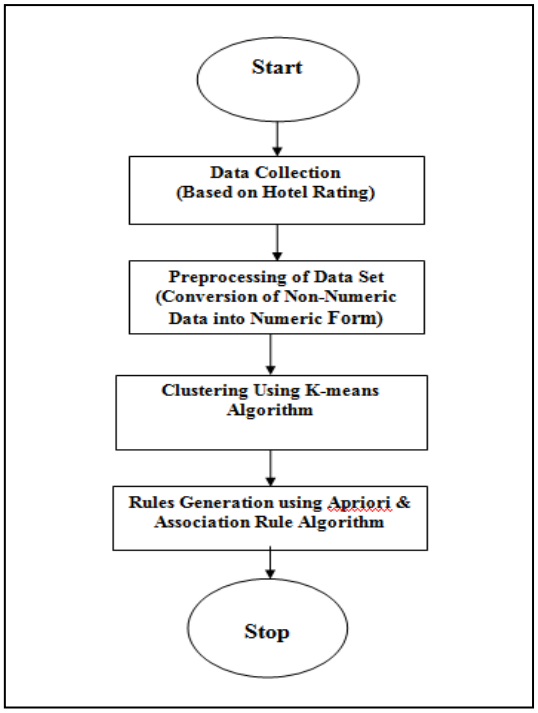

Fig 1: Proposed System Flow Chart

First step is to select the appropriate data set the dataset should be preprocessed in such a way that we have to convert the non-numeric data into the numeric data and convert it into the matrix form Applying K-means clustering algorithm on the matrix form pre-processed data set to generate the clusters and applying Apriori algorithm on the clustered data to generate the best rules for recommendation

B. Theoretical Work

STEP1 (Data Collection) First step is to collect dataset regarding hotel rating. Let us suppose there are 10 users and 5 hotels as given below in table1.

User Name (User ID)	Hotel Name (Hotel ID)	Rating
Ram (u1)	Meera (R1)	***
Ram (u1)	Tirupati (R2)	***
Ram (u1)	Guru Kripa (R3)	?
Ram (u1)	Chopal sagar (R4)	**
Ram (u1)	Padmavati (R5)	?
Ajay (u2)	Meera (R1)	*****
Ajay (u2)	Tirupati (R2)	?
Ajay (u2)	Guru Kripa (R3)	***
Ajay (u2)	Chopal sagar (R4)	*
Ajay (u2)	Padmavati (R5)	?
Yash (u3)	Meera (R1)	?
Yash (u3)	Tirupati (R2)	***
Yash (u3)	Guru Kripa (R3)	?
Yash (u3)	Chopal sagar (R4)	?
Yash (u3)	Padmavati (R5)	*****
Pravin (u4)	Meera (R1)	?
Pravin (u4)	Tirupati (R2)	****
Pravin (u4)	Guru Kripa (R3)	****
Pravin (u4)	Chopal sagar (R4)	****
Pravin (u4)	Padmavati (R5)	****
Shyam (u5)	Meera (R1)	*****
Shyam (u5)	Tirupati (R2)	***
Shyam (u5)	Guru Kripa (R3)	***
Shyam (u5)	Chopal sagar (R4)	**
Shyam (u5)	Padmavati (R5)	*****
Narendra (u6)	Meera (R1)	*****
Narendra (u6)	Tirupati (R2)	?
Narendra (u6)	Guru Kripa (R3)	***
Narendra (u6)	Chopal sagar (R4)	**
Narendra (u6)	Padmavati (R5)	*****
Sita (u7)	Meera (R1)	****
Sita (u7)	Tirupati (R2)	****
Sita (u7)	Guru Kripa (R3)	****
Sita (u7)	Chopal sagar (R4)	***
Sita (u7)	Padmavati (R5)	?
Madhav (u8)	Meera (R1)	***
Madhav (u8)	Tirupati (R2)	*****
Madhav (u8)	Guru Kripa (R3)	*****
Madhav (u8)	Chopal sagar (R4)	?
Madhav (u8)	Padmavati (R5)	?
Gayatri (u9)	Meera (R1)	?
Gayatri (u9)	Tirupati (R2)	?
Gayatri (u9)	Guru Kripa (R3)	***
Gayatri (u9)	Chopal sagar (R4)	?
Gayatri (u9)	Padmavati (R5)	?
Karun (u10)	Meera (R1)	****
Karun (u10)	Tirupati (R2)	?
Karun (u10)	Guru Kripa (R3)	?
Karun (u10)	Chopal sagar (R4)	***
Karun (u10)	Padmavati (R5)	*****

Table 1: Non Numeric data

STEP2 (Data Preprocessing): The dataset should be preprocessed in such a way that we have it will be converted the non-numeric data into numeric data in the form of matrix as shown in table4.

User Name (User ID)	Hotel Name (Hotel ID)	Rating
u1	R1	3
u1	R2	3
u1	R3	?
u1	R4	2
u1	R5	?
u2	R1	5
u2	R2	?
u2	R3	3
u2	R4	1
u2	R5	?
u3	R1	?
u3	R2	3
u3	R3	?
u3	R4	?
u3	R5	5
u4	R1	?
u4	R2	4
u4	R3	4
u4	R4	4
u4	R5	4
u5	R1	5
u5	R2	3
u5	R3	3
u5	R4	2
u5	R5	5
u6	R1	5
u6	R2	?
u6	R3	3
u6	R4	2
u6	R5	5
u7	R1	4
u7	R2	4
u7	R3	4
u7	R4	3
u7	R5	?
u8	R1	3
u8	R2	5
u8	R3	5
u8	R4	?
u8	R5	?
u9	R1	?
u9	R2	?
u9	R3	3
u9	R4	?
u9	R5	?
u10	R1	4
u10	R2	?
u10	R3	?
u10	R4	3
u10	R5	5

Table2: Conversion of table1

Fig.2: converting data set into Matrix form

User ID	Hotel Rating				
	R1	R2	R3	R4	R5
u1	5	3	?	2	?
u2	5	?	3	1	?
u3	?	3	?	?	5
u4	?	4	4	4	4
u5	5	3	3	2	5
u6	5	?	3	2	5
u7	4	4	4	3	?
u8	3	5	5	?	?
u9	?	?	3	?	?
u10	4	?	?	3	5

Table 3: Matrix Form of table2

Replacing the missing data which is represented by '?' with "0". So we have assign a "0" to the missing data.

User ID	Hotel Rating				
	R1	R2	R3	R4	R5
u1	5	3	0	2	0
u2	5	0	3	1	0
u3	0	3	0	0	5
u4	0	4	4	4	4
u5	5	3	3	2	5
u6	5	0	3	2	5
u7	4	4	4	3	0
u8	3	5	5	0	0
u9	0	0	3	0	0
u10	4	0	0	3	5

Table 4: Preprocessed data in matrix form

STEP3(Clustering): Applying K-means clustering algorithm on the pre-processed data set. Firstly K-means algorithm randomly selects K of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is most similar, based on the Euclidean distance between object and the cluster mean. Now Applying K-Means with 2 cluster where center**C1=(0,4,4,4,4), C2=(2,0,3,2,2)** where **U1= (5,3,0,2,0)** Calculating the distance between C1 and U1 and distance between C2 and U1

$$D_{11} = \sqrt{(5.0-0)^2+(3.0-4.0)^2+(0.0-4.0)^2+(2.0-4.0)^2+(0.0-4.0)^2} = 7.87$$

$$D_{12} = \sqrt{(5.0-2.0)^2+(3.0-0.0)^2+(0.0-3.0)^2+(2.0-2.0)^2+(0.0-2.0)^2} = 5.56$$

As the distance of D12 is smaller as compare to D_{11} so the U1 belongs to **cluster2.**Similarly we will calculate the distance all other items. The result of K-means clustering is

Cluster-1

User	R1	R2	R3	R4	R5
u2	5	0	3	1	0
u4	0	4	4	4	4
u7	4	4	4	3	0
u10	4	0	0	3	5

Table 5: Cluster1

Cluster-2

User	R1	R2	R3	R4	R5
u1	5	3	0	2	0
u3	0	3	0	0	5
u5	5	3	3	2	5
u6	5	0	3	2	5
u8	3	5	5	0	0
u9	0	0	3	0	0

Table 6: Cluster2

STEP4 (Rule Generation): The results of K-means algorithm are given to Association mining algorithm. With the help of the Éclat algorithm we can find the frequent item sets. Based on frequent item sets the strong association rules will be generated. Here converting the data into the Boolean form i.e. true and false. Here we have considered the minimum support =2 so based on that frequent item set that are generated are as follows

Fig 3: Cconverting Clustered file into Boolean form

User	R1	R2	R3	R4	R5
u2	True	False	True	True	False
u4	False	True	True	True	True
u7	True	True	True	True	False
u10	True	False	False	True	True

Table 7: Boolean form of cluster_1

Frequent 1-Hotel Rating set

R1=True	{u2,u7,u10}
R2=True	{u4,u7}
R3=True	{u2,u4,u7}
R4=True	{u2,u4,u7,u10}
R5=True	{u4,u10}
R1=False	{u4}
R2=False	{u2,u10}
R3=False	{u10}
R4=False	{Ø}
R5=False	{u2,u7}

Table 8: Organizing table7 into True and false group

Frequent 2- Hotel Rating set

R1=True,R2=False	{u2,u10}
R1=True,R3=True	{u2,u7}
R1=True,R5=False	{u2,u7}
R1=False,R5=True	{u4,u7}
R2=True,R3=True	{u4,u7}
R2= True,R4=True	{u4,u10}
R3=True,R4=True	{u2,u4,u7}
R3= True,R5=False	{u2,u7}
R4=True, R5=False	{u2,u7}
R4= True,R5=True	{u4,u10}

Table 9: Union operation of R1_ R2 from table7

Frequent-3 Hotel Rating set

R1=True,R3=True,R4=True	{u2,u7}
R1=True,R3=True,R5=False	{u2,u10}
R2=True,R3=True,R4=True	{u4,u7}
R1=True,R4=True,R5= False	{u2,u7}
R3=True,R4=True,R5= False	{u2,u7}

Table10: Union operation of R1_ R3_R4_R5 from table7

So based on the frequent item set the rules that are generated are as follows that satisfy the minimum support and minimum confidence:

- R1 = True,R3 =True -> R4 =False
- R3 =True,R4 = True->R1=False
- R1=True,R4 =True->R3 =False

STEP-5 On the basis of strong rules the average will be calculated for the particular rules and if it satisfies the threshold rules then we can recommend to the user.

We will check for R1=True,R4=True ->R3=True. So for those who have recommended R1,R4 and not rated R3 the R3 will be recommended to that user but we have to check the average of R1,R4 and R3. find out the

average ,suppose the average of R3 is 4 and average of R1,R4 is 3.5 so we can recommend R3 will be recommended to that user who have recommend R1,R4.

VI. Result Analysis

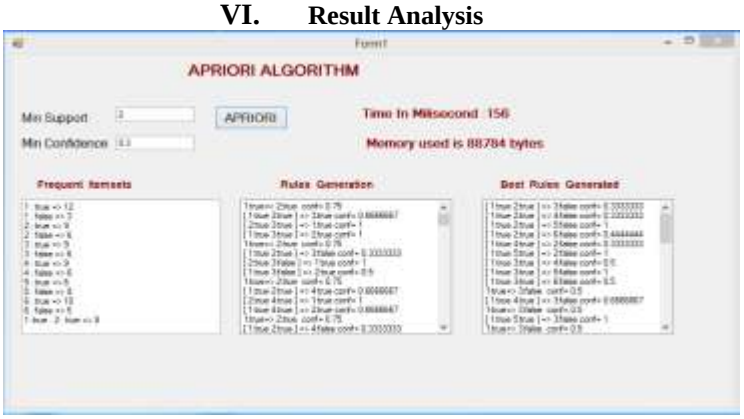

Fig 4: Best Rule Generated

A. Result Analysis

If we using the proposed system instead of existing system then number of best generated rule are less with respect to number of generated rules. So, end user will get recommendation in less amount of time.

VII. Conclusion

Proposed Recommendation System is better than the existing recommendation system because it is the need of present time due to healthy competition between us and fast life which take less time and provide better suggestion on the basis of better rule generation.

VIII. Future Work

The scope for the future enhancement is to implement C-means Clustering Algorithm and Principal Direction Divisive Partitioning (PDDP) algorithm for clustering.

References

[1] Meenal Moghe, Abhishek Raghuvanshi, 2015 has proposed, "A Novel Collaborative Filtering Recommendation System Algorithm"– *(IJETR)*-October 2015
[2] Qi Wang, Wei Cao and Yun Liu– "A Novel Clustering Based Collaborative Filtering Recommendation System Algorithm " – Springer 2014
[3] Mohammad Hamidi Esfahani, Farid Khosh Alhan - "New Hybrid Recommendation System Based On C-Means Clustering Method " -IEEE-2013 5th conference of information Knowledge
[4] Hazem Hajj, Wassim El-Hajj, Lama Nachman – "A Hybrid Approach with Collaborative Filtering for Recommender Systems" – IEEE 2013.
[5] Haojun Sun –"A New Item Clustering based Collaborative Filtering Approach." –IEEE 2012.
[6] Maria Teresa Andrade, Fábio Almeida, "Novel Hybrid Approach to Content Recommendation based on Predicted Profiles" IEEE 2013.
[7] Jens Grivolla, Toni Badia, Diego Campo, Miquel Sonsona, Jose-Miguel Pulido, "A hybrid Recommender Combining User item and interaction data"-IEEE 2014.
[8] Ahmad A. Kardan, Mahnaz Ebrahimi, "A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups" 2012.
[9] Rahul Katarya, Dr. Om Prakash Verma ,Ivy Jain, "User Behaviour Analysis in Context-aware Recommender System using Hybrid Filtering Approach"- (ICCCT) 2013.
[10] Qing Li, Byeong Man Kim, "Clustering Approach for Hybrid Recommender System" – IEEE 2003
[11] R.Srinivasa Raju I.Kali Pradeep I.Bhagyasri "Recommender Systems for E-commerce Novel parameters and Issues". - IJARCSSE 2013.
[12] Hridya Sobhanam, A.K. Mariappan, "A Hybrid Approach to Solve Cold Start Problem in Recommender System using Association Rules and Clustering Technique". – IJCA 2013

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Ravindra Ramawat. "A Hotel Recommendation System Based on Data Mining Techniques ." IOSR Journal of Computer Engineering (IOSR-JCE), vol. 19, no. 4, 2017, pp. 12–19.