Recommendation based on Clustering and Association Rules

Jaimeel .M.Shah[1], Lokesh Sahu[2]

[1] *Student, Computer Department, Parul Institute, Gujarat, India*
[2] *Asst.Professor, CSE Department ,Parul Institute, Gujarat, India*

ABSTRACT

Abstract: Recommender systems play an important role in filtering and customizing the desired information. Recommender system are divided into 3 categories i.e collaborative filtering , content-based filtering, and hybrid filtering and they are the most adopted techniques being utilized in recommender systems. The paper mainly describe about the issues of recommendation system. The main aim of paper is to recommend the suitable items to the user, so for recommending the suitable items a better rule extraction is needed. Thus for better rule extraction Association mining is applied .The clustering method is also applied here to cluster the data based on similar characteristics .The propose methods try to eliminate certain problems such as sparsity, cold-start problem. So to overcome the certain problem association mining over clustering is used.

Keyword: - *K-means clustering, Eclat Algorithm , Hybrid recommendation* **.**

---

1. Introduction

In the current age, there is large amount of information available in web. It leads to information overload. This rapidly growth is mainly because of online world and Electronic Commerce. There are many e-commerce sites such as amazon.com, Netflix.com, Movielens.com etc. These sites suggest millions of products such as movies, books etc. E-commerce provides core business process of buying, selling goods, services and information over the internet. Its usage is increasing day-by- day. They help manufacturers in advertisement of their items also. Traditionally they advertise their product through TVs, radios, newspapers etc. These media takes more cost and feedback of people to their product cannot be known [1]. These sites help in sales of products. Due to this, users enjoy shopping from their home. Users approach those websites to get items to which they are more interested. All categories of items can be found. Finding products of user's interest is known as Personalization Recommendation system[2].

1.1 Application of Recommendation system

Recommendation system in Music domain: While surfing over internet, user will find many websites related to music recommendation. This application is going to discuss about the recommendation process about Last.Fm. User has to give a favorites artist which results in a list of audio stream of similar music.

Recommendation system in E-Commerce: E-Commerce application utilized recommendation system to suggest a list of recommendation to the user on the basic of their choices. Many applications involve the usage of explicit ratings given by users while other use the implicit ratings along with the demographic information, items most liked etc[4].

Recommendation system In Social Network: With the growth of the internet, user use social networks to communicate with each other and thus sharing their choices. A social network defines the concept of psychological and social relationship between each user or group in the form of graphs utilizing actors and relations.

1.2 Challenges And Issues

A. Cold-Start Problem: It's difficult to give recommendations to new users because his profile is almost empty and he has not rated any items so the taste of user remains unknown to the system so this type of the problem is called as coldstart problem. In some recommender systems this problem can be solve with the survey at the time of creating a profile. Another problem is when user has not rated before when new to the system . Both of these problems can be overcome with the help of hybrid approaches[2].

B. Data-Sparsity: Sparsity is the problem of lack of information. Suppose we have a huge amount of users and items but user have rated only few items. If a user has evaluated only few items then its difficult to determine the taste of the user.so to overcome this we use collaborative and hybrid approach to create neighborhoods of users based on their profiles.

C. Scalability: Due to increase of numbers of users and items, the system needs more resources for processing information and forming recommendations. Majority of resources is consumed with the purpose of determining users with similar tastes, and goods with similar descriptions. This problem is also solved by the combination of various types of filters and physical improvement of systems.

D. Gray Sheep: Gray sheep problem means where user does not consistently agree or disagree to the group of the people and due to this reason for such user recommendation seems to be difficult[7] .

2. Propose Framework

We propose a hybrid recommendation framework that integrates association rule mining with a cluster-based approach, based on an assumption that the (User × Item) space has a large number (e.g., larger than 1000) of users but a small number (e.g., less than 50) of items. We use the Eclat algorithm [5] to generate a set of association rules on clustering data. The Eclat algorithm mines over the frequent sets to discover association rules. The most important parameters in the Eclat algorithm are support count and minimum confidence [10]. Generated association rules play an important role in our proposed recommendation framework.

Table -1: user X item matrix

User/item	I_1	I_2	I_3	….I_N
U1	3	0	2	……
U2	4	3	0	
U_n	…..	……	…….	…..

2.1 Hybrid Recommendation

Our proposed framework consists of three parts. The first part is to generate a set of clusters using the K-means algorithm. The second part is to apply Eclat algorithm for generating frequent item sets .The third part consist of generating association rules to recommend items for a user from frequent itemsets.

2.2 User Clustering

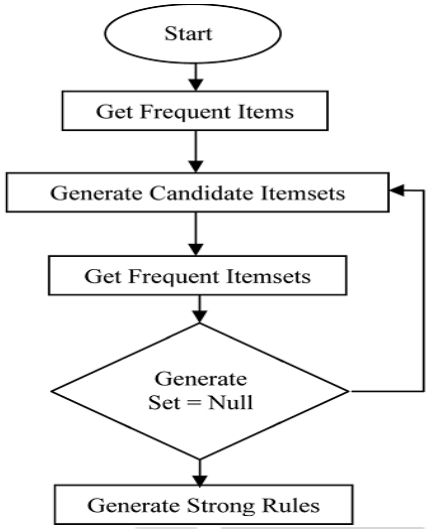
First of all to apply our proposed system we first need to cluster the data. The main aim of the clustering is to divide the clusters based on the similarity characteristics. The first step is user clustering, and clustering is a preliminary step for the subsequent step to gather those similar users. In this paper, we use K-means algorithm to cluster our user set in the user-item rating matrix. Furthermore, we use Euclidean distance to represent the distance between users. The clustering algorithm are as follows.

Input: k original user centroids

1.for each user vector U
2. for the kth centroid C
end
3. find the shortest distance—distance[i]
4. distance[k]=Euclidean(U,C);
5. assign user U to cluster i
6. end

2.3. Association rules

Secondly, to apply our proposed algorithm, we first need to obtain the required association rules via Eclat algorithm. The inputs of the Eclat algorithm are: the transactions file, minimum support, and minimum confidence. Eclat algorithm is mainly a depth-first search algorithm which uses set intersection. Vertical database layout is mainly used in Eclat i.e. instead of listing all transactions explicitly; each item is stored together with its cover (also called tid list) and uses the intersection based approach to compute the support of an itemset[9]. In this way, the support of an itemset X can be computed by simply intersecting the covers of any two subsets X, Z ⊆ Y, such that X U Z = Y. It says that, when the database is stored in the vertical layout, the support for set can be counted much easier by intersecting the covers of two of its subsets that together give the set itself



Algorithm 1 Eclat

Input: $E((i_1,t_1),\dots(i_n,t_n))|P), s_{min}$
Output: $F(E, s_{min})$
1: *for all* i_j *occuring in* E *do*
2: $P \coloneqq P \cup i_j$ // *add* i_j *to create a new prefix*
3: *init* (E') // *initialize a new equivalence class with the new prefix* P
4: *for all* i_k *occuring in* E *such that* $k > j$ *do*
5: $t_{tmp} = t_j \cap t_k$
6: *if* $|t_{tmp}| \geq s_{min}$ *then*
7: $E' \coloneqq E \cup (i_k, t_{tmp})$
8: $F = F \cup (i_k \cup P)$
9: *end if*
10: *end for*
11: *if* $E' \neq \{\}$ *then*
12: *Eclat* (E', s_{min})
13: *end if*
14: *end for*

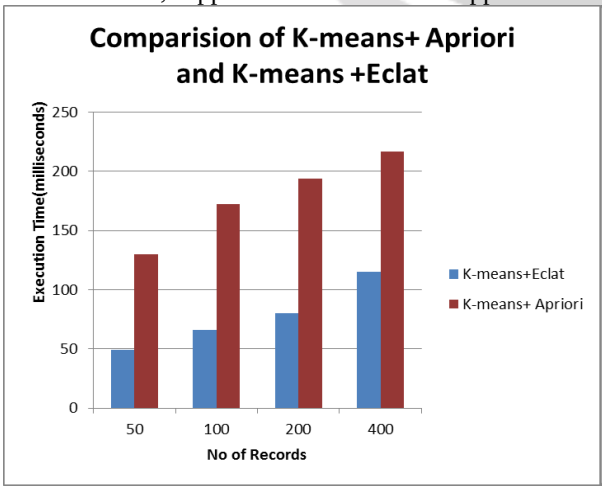
Chart -1: Flow chart of Eclat

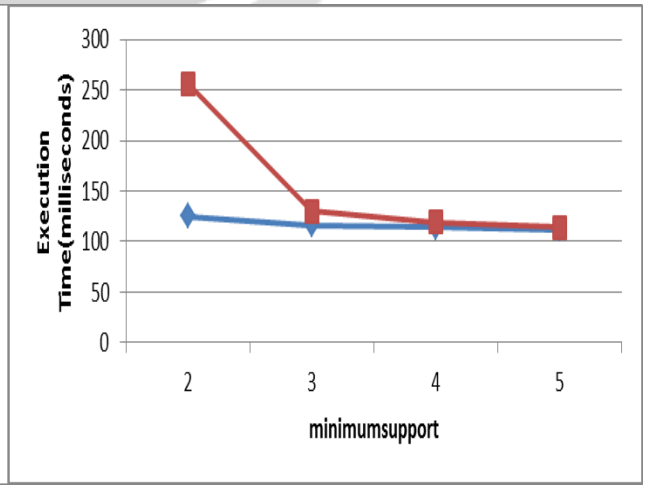
3. Experiments And Results

This section presents an experimental study of our proposed framework. It describes the experimental setup, presents the experiment results, and finally it summarizes our observation. Here we have taken a 100 users X 107 hotels, but after preprocessing we have selected a hotel where users get 50 ratings.

3.1 Execution Time

Here we have compare the eclat algorithm with the apriori algorithm based on the certain parameters such as execution time, support count. In 1st case support count remains fixed and in 2nd case no of records remains fixed.





4. CONCLUSIONS

Recommendation systems provide valuable suggestions to users with the help of user rating data. The main aim of the proposed system is to recommend the best suitable items to the user. The proposed system deals with the improving the accuracy and sparsity problem. The proposed system applies association mining over clustering. Initially the standard dataset is taken for evaluation by the system. The dataset is pre-processed and the pre-processed data is given to the clustering K-means algorithm and it will give clustered data. As a future work applying C-means algorithm are given to Association mining algorithm i.e. to generate the strong association rules from results. The optimized results from Association mining algorithm module are helpful in providing the better rules.

REFERENCES

1. Mohammad Hamidi Esfahani, Farid Khosh Alhan - "New Hybrid Recommendation System Based On C-Means Clustering Method " -IEEE-2013 5th conference of information Knowledge.
2. Qi Wang, Wei Cao and Yun Liu– "A Novel Clustering Based Collaborative Filtering Recommendation System Algorithm " –Springer 2014
3. Ahmed Mohammed K. Alsalama – "A Hybrid Recommendation System Based On Association Rules" – ISSR-2014
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. R.Srinivasa Raju I.Kali Pradeep I.Bhagyasri "Recommender Systems for E-commerce: Novel Parameters and Issues".-IJARCSSE 2013
7. Robin Burke "Hybrid Recommender Systems: Survey and Experiments"–ACM 2002
8. Ujwala H.Wanaskar, Sheetal R.Vij, Debajyoti Mukhopadhyay "A Hybrid Web Recommendation System based on the Improved Association Rule Mining Algorithm".-Arxiv journal 2013
9. S.Vijayarani,P.Sathya – "Mining Frequent Item Sets over Data Streams using Éclat Algorithm".
10. Soumi Ghosh, Sanjay Kumar - "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms".-(IJACSA) International Journal of Advanced Computer Science aSApplications, Vol. 4, No.4, 201