14 evening

by Abhinav Singhal

RECOMMENDATION SYSTEM USING FREQUENT ITEM DATA SETS

A.Singhal[1,*], B.Pant[2]

[1] Pursuing B-Tech in Computer Science and Engineering
Graphic Era deemed To Be University, Dehradun, Uttrakhand
E-mail: abhinavsinghal024@mail.com .

[2] Department of Computer Science and Engineering
Graphic Era deemed To Be University,Dehradun,Uttrakhand.
E-mail: pantbhaskar2@mail.com

* *Corresponding Author*. E-mail : abhinavsinghal024@gmail.com

Abstract. The recommender system is the most successful application of information filtering. It helps the user to find items of there interest from huge data Sets in this paper we have proposed the Apriori Algorithm to find frequent interesting patterns and then the performance of the proposed method is evaluated on movie lens data set by considering minimum support as 20. The frequent patterns retrieved can then be used to make recommendations.

1 Introduction

The recommendation system is a prediction of ratings given by users for a particular item or it is a subclass of information filtering system. Recommender systems have become common in last few decades. The main significance of the Recommendation system is to get some useful recommendation of different items for user.Number of recommendations can be produced with the help of recommendation system.There are three techniques to recommend any data. And that are collaborative, content-based and hybrid recommendation technique.[1].
In collaborative Technique, a model is built from users past activity it has been noticed that similar decisions were made by new users. The items that are previously purchased or numerically rating are more liked by the users. Then the model is used for projecting items some examples- are movies, electronic gadgets and the products in general, so technically we can say recommendation system is evolve d from data analyzed and from new classes .knowledge making products recommendation during live customer dealing[2].

---

[1] Keywords: Recommendation,collaborative technique, content-based technique

In content-based technique, method are based on the description of item and the profile of users preferences in Content-based filtering keywords used to describe the items and users profile is built to indicate the type of item the users like. In other words, the algorithm tries to recommend items that are similar to the items that are liked by the user in his past activity.[3].

While the combination of collaborative technique and Content -based technique is known as Hybrid recommendation tehnique [4]. A hybrid approach can be applied in many ways by making collaborative and content-based work separately and then combining them[5].
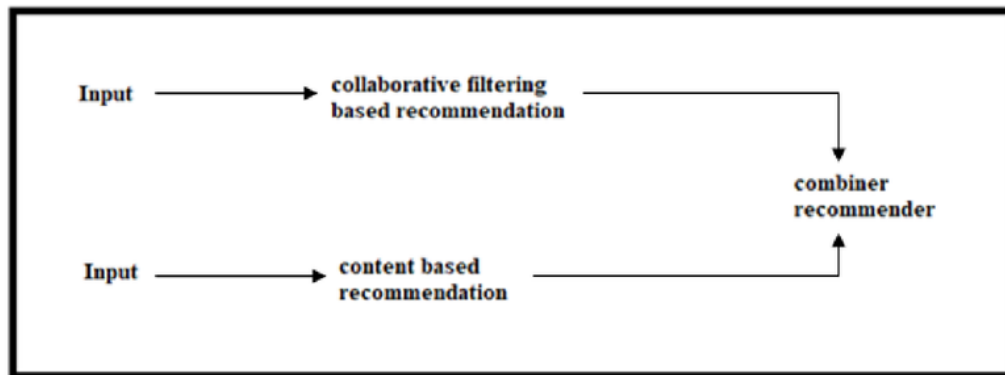

Fig. 1

Apriori Algorithm is an algorithm for mining of frequent item datasets and for learning of Associative rule . Agarwal and Srikant proposed Apriori in 1994. Bottom-Up Approach is used in Apriori Algorithm .Whereas one at a time frequency subsets are extended in this algo and Apriori is designed in such a way that it can be applied on transactional database example -customer brought number of items [6].Apriori also determined frequent itemsets. Apriori can be used to general trends in the database using association rule. Market basket analysis is one of its applications.Apriori algorithm uses large item-sets properly ,parallelization is very easy.It is easy to implement and generate frequent results[7].

2 Material Methods And Working:

The data is fetched from (movieslens.com-https://grouplens.org/datasets/movielens) which contains some necessary supplies such as User Id or (we can say number of users), Movie Id, Movie Name, categories in which user can comment (Action, Adventure, Animation, children, comedy, crime, drama, Fantastic, mystery, music, Horror, romantic, Documentary, thrilling, war, western) i.e genre and the ratings that are done from 1to 5 ,we have taken review from 40 users and then applying algorithm of frequent item data set. Basically, there are 16 categories in which 40 users can give there comment. so when we make the combination of 2 items i.e 7C2=21 possible combination are formed out of which 20 passes the threshold value i.e 20. Now we have applied our algorithm for a combination of three items i.e 7C3=35 possible combination are formed out of which 22 items pass the threshold value. Then we have moved further and make the combination of 4 items i.e 7C4=35.now out of these 35

possible combinations, 7 items pass the threshold value and reaches to the next level. Now I have tried to make the combination of 5 items i.e 4 possible combinations are formed and no one has passed the threshold value.So (ACTION, ADVENTURE, COMEDY, DRAMA, CRIME, THRILLER, RO-MANCE) are the items which are most liked by the users from the box given below.

AC TI ON	ADV ENT URE	ANI MAT ION	CHI LDR EN	CO ME DY	CR IM E	DR A M A	FA NT ASY	MY STE RY	MU SIC AL	HO RR OR	RO MA NCE	DOCU MENT ARY	TH RIL LER	W A R	WE STE RN
33	30	16	19	36	31	36	16	13	9	8	26	2	32	1 3	6

Fig. 2

2.1 Apriori Algorithm:

According to apropri property all non empty subsets of frequent itemset must be frequent.whereas apriori assumes that if all subsets of frequent itemset must be frequent.if a item set is infrequent than all its superset is infrequent To count candidate Item-Set efficiently Apriori algorithm uses hash tree structure and breadth-first search technique.It generates candidate item set of length k. From its set of length K-1. Pseudo code is given for a transaction database T. Support threshold of E . Ck is the candidate set of level k. Count[c] accesses a field of data structure that represents data set c which is initialized by 0. [8].

$$\text{Apriori}(T, \epsilon)$$
$$L_1 \leftarrow \{\text{large } 1 - \text{itemsets}\}$$
$$k \leftarrow 2$$
$$\textbf{while } L_{k-1} \neq \emptyset$$
$$\quad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \land b \notin a\} - \{c \mid \{s \mid s \subseteq c \land |s| = k - 1\} \not\subseteq L_{k-1}\}$$
$$\quad \textbf{for } \text{transactions } t \in T$$
$$\quad\quad D_t \leftarrow \{c \mid c \in C_k \land c \subseteq t\}$$
$$\quad\quad \textbf{for } \text{candidates } c \in D_t$$
$$\quad\quad\quad count[c] \leftarrow count[c] + 1$$
$$\quad L_k \leftarrow \{c \mid c \in C_k \land count[c] \geq \epsilon\}$$
$$\quad k \leftarrow k + 1$$
$$\textbf{return } \bigcup_k L_k$$

Fig. 3

3 Result and Discussion:

After applying algorithm of frequent item datasets with minimum support as 10 the result obtained is very large and time taking .Then we have taken minimum support as 20 and applied the algorithm in dataset. Basically there are 5 tables each table shows us the different combinations of items and the rating of the items. we find top 5 movie categories liked by users (ACTION, ADVENTURE, COMEDY, DRAMA, CRIME). So after looking at these results, one can predict the liking and disliking of peoples.
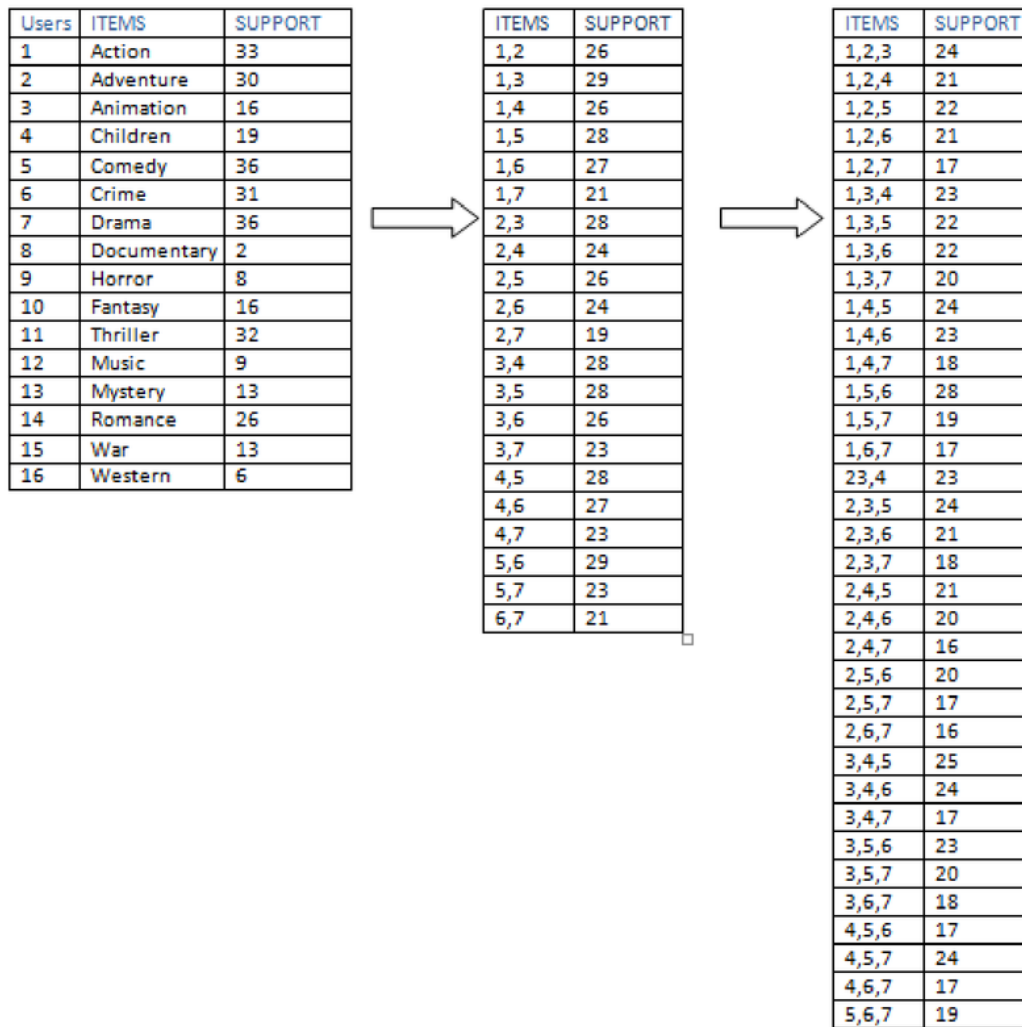
Users	ITEMS	SUPPORT
1	Action	33
2	Adventure	30
3	Animation	16
4	Children	19
5	Comedy	36
6	Crime	31
7	Drama	36
8	Documentary	2
9	Horror	8
10	Fantasy	16
11	Thriller	32
12	Music	9
13	Mystery	13
14	Romance	26
15	War	13
16	Western	6

ITEMS	SUPPORT
1,2	26
1,3	29
1,4	26
1,5	28
1,6	27
1,7	21
2,3	28
2,4	24
2,5	26
2,6	24
2,7	19
3,4	28
3,5	28
3,6	26
3,7	23
4,5	28
4,6	27
4,7	23
5,6	29
5,7	23
6,7	21

ITEMS	SUPPORT
1,2,3	24
1,2,4	21
1,2,5	22
1,2,6	21
1,2,7	17
1,3,4	23
1,3,5	22
1,3,6	22
1,3,7	20
1,4,5	24
1,4,6	23
1,4,7	18
1,5,6	28
1,5,7	19
1,6,7	17
23,4	23
2,3,5	24
2,3,6	21
2,3,7	18
2,4,5	21
2,4,6	20
2,4,7	16
2,5,6	20
2,5,7	17
2,6,7	16
3,4,5	25
3,4,6	24
3,4,7	17
3,5,6	23
3,5,7	20
3,6,7	18
4,5,6	17
4,5,7	24
4,6,7	17
5,6,7	19

Fig. 4

ITEMS	SUPPORT
1,2,3,4	20
1,2,3,5	19
1,2,3,6	19
1,2,3,7	17
1,2,4,5	20
1,2,4,6	16
1,2,4,7	14
1,2,5,6	18
1,2,5,7	17
1,2,6,7	15
1,3,4,5	21
1,3,4,6	21
1,3,4,7	17
1,3,5,6	19
1,3,5,7	19
1,3,6,7	18
1,4,5,6	21
1,4,5,7	18
1,4,6,7	14
1,5,6,7	19
2,3,4,5	20
2,3,4,6	19
2,3,4,7	17
2,3,5,6	17
2,3,5,7	16
2,4,5,6	19
2,4,5,7	15
2,4,6,7	18
2,5,6,7	19
3,4,5,6	21
3,4,5,7	17
3,4,6,7	16
3,5,6,7	17
4,5,6,7	18

ITEM	SUPPORT
(1,2,3,4,5)	17
(1,2,3,4,6)	18
(1,2,4,5,6)	15
(1,3,4,5,6)	18
(2,3,4,5,6)	15

Fig. 5

4 Conclusion And Future Scope:

In Apriori, algorithm scanning is done at every level. It can be very slow. (n+1) scans are required ,Where the n= length of the longest path. Number of database scans are required .Transaction database is memory resident in this.In Dynamic dataset counting new candidate itemsets only added when all their subsets are estimated to be frequent.While in Sampling , mining is done on a given subset of data.[9]. Recommendation system helps customers to figure out the best quality products which they must buy. The recommendation also increases the sales and increases the revenue. As the number of internet users is increasing day by day. So E-commerce is highly affected by recommendation technique there statistics have increased. Many companies have handled devices this will increase the scope of recommendation systems in coming years In Apriori, algorithm Scanning is done at every level. It can be very slow. (n+1) scans are required Where the n= length of the longest path.[10].

References

[1] Lahiru S Gallege and Rajeev R Raje. Parallel methods for evidence and trust based selection and recommendation of software apps from online marketplaces, 2017.

[2] Samuel Arleo and Joao Mendes-Moreira. Dynamic selection of recommender systems: an application to tourism. *Journal of Tourism and Hospitality*, 6(6), 2017.

[3] Ali Davar, David Robert Thompson, and Maziyar Hamdi. Evaluating service providers using a social network, June 5 2018. US Patent 9,990,609.

[4] CAN CHANGE. i can.

[5] Nazli Medghalchi and Ahmed Aljmiai. A literature review: Qualitative evaluation of recommendation systems for social networks based on data quality issues.

[6] Norhidayah Ali, Kamaruzaman Jusof, Syukriah Ali, Najah Mokhtar, and Azni Syafena Andin Salamat. The factors influencing studentsperformance at universiti teknologi mara kedah, malaysia. *Management Science and Engineering*, 3(4):81–90, 2009.

[7] Aarti Bansal. Comparative study of data mining algorithms for analyzing crimes against women-a review. 2015.

[8] Bhaskar N Patel, Satish G Prajapati, and Kamaljit I Lakhtaria. Efficient classification of data using decision tree. *Bonfring International Journal of Data Mining*, 2(1):06–12, 2012.

[9] Heena Sabnani, Mayur More, Prashant Kudale, and Surekha Janrao. Prediction of student enrolment using data mining techniques. 2018.

[10] Abhishek Saxena and Navneet K Gaur. Frequent item set based recommendation using apriori. *International Journal of Science, Engineering and Technology Research (IJSETR)*, 4(5), 2015.

14 evening

Student Paper

8   Submitted to Maharshi Dayanand University
    Student Paper   1%

9   Submitted to University of Westminster
    Student Paper   1%

10 scholar.oauife.edu.ng
    Internet Source   1%

Exclude quotes   Off   Exclude matches   Off
Exclude bibliography   On