Frequent Item Set Based Recommendation using Apriori

Abhishek Saxena, Navneet K Gaur

***Abstract*—Exponential growth of the world-wide-web and emergence of e-commerce as a platform within reach of common customers had led to the development of numerous recommender systems that defines a personalized information retrieval technique used to identify set(s) of items that will be of interest to a certain user.Most of these researches relate item recommender on the basisof user profile and item-oriented recommendation. But here we are exploring scope of frequent item set based recommendation by implementing Apriori algorithm. Apriori is mainly used to find frequently purchased items/products. The key idea behind this recommendation is that any item set that occurs frequently together must have each item (or any subset) occur at least as frequently.**

***Index Terms*—Recommendation System,Apriori Algorithm,Association Rule, Frequent Item set.**

I. INTRODUCTION

Recommendation systems have become extremely common in recent years. This systemis applied in a variety of applications. In definition, goal of a Recommender System is to generate significant recommendations to a collection of users for items or different products. Recommendation systems usually produce a list of recommendations in one of two ways - through collaborative filtering or content-based filtering.In Collaborative filtering, it approaches building a model from a user's past activities (items that are previously purchased and/or numerical ratings given to those items) as well as similar decisions made by other users; then use that model to projection items (or ratings for items) that the user may have an concern in.

The most popular ones are probably books, research articles, search queries, movies, music,news, social tags, and products in general. There are also recommendation systems for life insurance companies, jokes,experts, restaurants,financial services,and Twitter followers.

Technically a recommendation system is software evolved from a new class of data analysis which applies knowledge discovery techniques to the misfortune of making product recommendations during live customer dealings.

In this work, we are dealing of frequent item set based recommendation using Apriori Algorithm which works on concept of eliminating most large sets as candidates by looking first at smaller sets and recognizing that a large set cannot be frequent unless all its subsets are.

II. EXISITING RECOMMENDATION SYSTEM

The majority of existing approaches to recommender systems focus on recommending the most relevant items to individual users without taking consideration of any contextual information, such as time, place and the company of other people (e.g., for watching movies or dining out). In other words, traditionally recommender systems deal with applications having only two types of entities, users and items, and do not put them into a context when providing recommendation.

It also providesrecommendations that are based on the user's area of interests, customer searches and also suggests products based on it. For e.g. Amazon uses user view data.If any customer is searching a product from a particular category the system suggests a product form the same category. It is also based on the current search by the user, the site recommends products. [7]

E-commerce recommendation algorithms often operate in a challenging environment. For example: [3]

- A large retailer might have huge amounts of data, tens of millions of customers and millions of distinct catalog items.
- Many applications require the results set to be returned in real-time, in no more than half a second, while still producing high-quality recommendations.
- Older customers can have a glut of information, based on thousands of purchases and ratings.
- Customer data is volatile: Each interaction provides valuable customer data, and the algorithm must respond immediately to new information.

However, in many applications, such as recommending a vacation package, personalized content on a Web site, or a movie, it may not be sufficient to consider only users and items – it is also important to incorporate the contextual information into the recommendation process in order to recommend items to users in certain circumstances. For example, using the temporal context, a travel recommender system would provide a vacation recommendation in the winter that can be very different from the one in the summer. Similarly, in the case of personalized content delivery on a Web site, it is important to determine what content needs to be delivered (recommended) to a customer and when. Every user who visits the site may not buy a product. They can just go through it and based on those search results the site recommends a product.

III. ASSOCIATION RULES

To extract the frequent sets of items from large amount of data and to gather this information used the association rules. The form of an association rule is I → j, where I is a set of items and j is an item. The implication of this association rule is that if all of the items in I appear in some basket, then j is "likely" to appear in that basket as well.

IV. PROPOSED RECOMMENDATION SYSTEM

In our proposed system, we are planning to use a methodology dealing of frequent item set based

ISSN: 2278 – 7798 All Rights Reserved © 2015 IJSETR

recommendation using Apriori Algorithm. Here we are using a "bottom up" approach, where frequent subsets are extended one item at a time and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Especially important are pairs or larger sets of items that occur much more frequently than would be expected were the items bought independently.

The whole point of the algorithm (and data mining, in general) is to extract useful information from large amounts of data. The algorithm aims to find the rules which satisfy both a minimum support threshold and a minimum confidence threshold (Strong Rules).

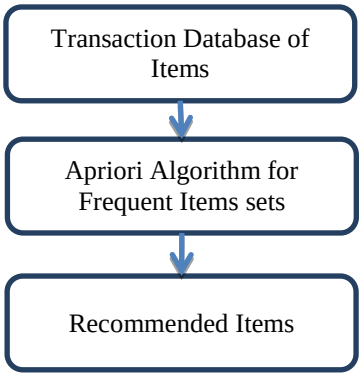

Figure-1: Proposed System

Here we are describing the Apriori Algorithm for finding frequent item sets. The key idea behind this algorithm is that any item set that occurs frequently together must have each item (or we can say any subset) occur at least as frequently. [6]

First Pass. In this algorithm, firstly we make one pass on all the tuples and retain a count for all the n items. Here we canuse a Hash Table. We set a threshold t and then only keep items that occur at least tn time (that is in at lead t percent of the tuples). For any frequent item set that occurs in at least 100t% of the tuples, must have each item also occur in at least 100t% of the tuples.

Second Pass. After the first pass, we make a second pass over all tuples. On this pass, we examine for frequent pairs of items, specifically for those items which occur in at least at-fraction of all baskets. Both items must have been found in the first pass. So we need to deliberate only $\binom{n1}{2} \approx n_1^2/2$ pairs of counters for these pairs of elements.

After this pass, we can again reject all pairs which occur less than at-fraction of all tuples. After this remaining set is expected far less than $n_1^2/2$. And these remaining pairs are already relatively interesting. They record all pairs that co-occur in more than at-fraction of purchases. And obviously include those pairs also which are occurring together even more frequently.

Further Passes. On the i^{th} pass, we can find sets of i items that occur together frequently (above at- threshold). For example, on the third pass we only need to consider triples were all sub-pairs occur at least at-fraction of times themselves. These triples can be found as follows: First sort all pairs (p, q) by their smaller indexed item (let smaller indexed be p). Then for each smaller indexed item p, consider all completions of this pair q. Now look at the pairs (q, r) with smaller item as q. For each of these pairs, check if the pair (p, r) also remains. Only triples (p, q, r) which pass all of these tests are given counters in the third pass.

Consider the following dataset where I want to find all item sets that occur in at least 1/3 of all tuples (at least 4 times):

TID	List of Items
T1	{A, B, C, D, E}
T2	{B, F, G, I}
T3	{A, C, E, F}
T4	{B, F, I}
T5	{G, H}
T6	{A, B, F}
T7	{J, C, E, F}
T8	{J, B, D}
T9	{B, D}
T10	{F, G, I}
T11	{C, F, I}
T12	{F, G, H}

Table- 1

After the first pass 1 have the following counters:

J	A	**B**	**C**	D	E	**F**	**G**	H	**I**
2	3	**6**	**4**	3	3	**8**	**4**	2	**4**

Table-2

So only n1 = 5 items survive {B, C, F, G, I}. In pass 2 we consider (n1/ 2) = 10 pairs: {(B, C),(B, F),(B, G),(B, I),(C, F),(C, G),(C, I),(F, G),(F, I),(G, I)}.

(B,C)	1
(B,F)	3
(B,G)	1
(B, I)	2
(C,F)	3
(C,G)	0
(C,I)	1
(F,G)	3
(F,I)	**4**
(G,I)	2

Table-3

We find that the only item set pair that occurs in at least 1/3 of all baskets is (F, I). Thus there can be no item set triple (or larger grouping) which occurs in all 1/3 of all tuples since then all of its pairs would need to be in 1/3 of all baskets, but there is only one such pair that satisfies that property. We can now examine the association rules. And see that the count of item F is quite large 8, and is much bigger than that of item I which is only 4. Since there are 4 pairs (F, I), then every time I occurs, F also occurs.

This methodology eliminates the problem which is making recommendations for new customer. As this proposed system is using transaction oriented data, it won't require the customer's profile to recommend products.

 All Rights Reserved © 2015 IJSETR

V. ILLUSTRATION OF ALGORITHM

Apriori algorithm is a classic algorithm for learning association rules. We are applying Apriori on a database that contains the transaction (e.g. a collection of items purchased by customers etc.)[7]

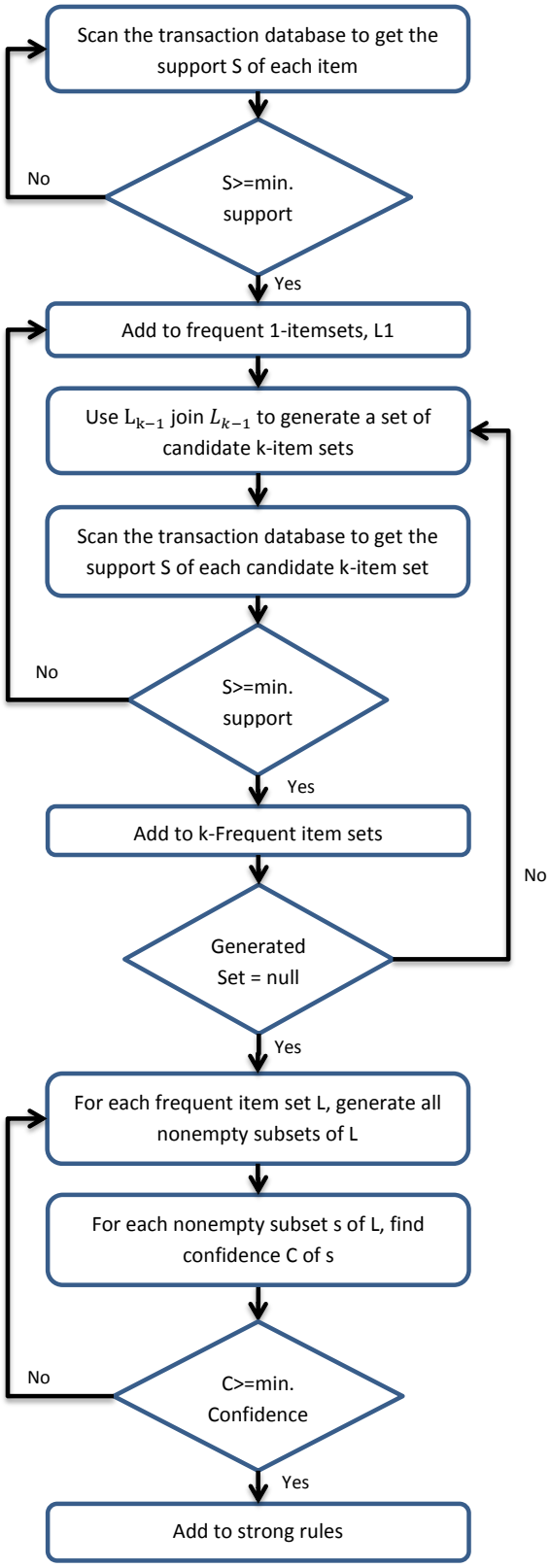

Figure-2: Apriori illustration to achieve frequent items-set

Apriori Algorithm is used to generate all frequent itemset[4]

Pass 1

I. Generate the candidate itemsets in C_1
II. *Save the frequent* itemsets in L_1

Pass k

I. Generate the candidate itemsets in C_k from the frequent itemsets in L_{k-1}
   i. Join L_{k-1}p with L_{k-1}q, as follows:
   insert into C_k
   select p.$item_1$, p.$item_2$, . . . , p.$item_{k-1}$, q.$item_{k-1}$
   from L_{k-1}p, L_{k-1}q
   where p.$item_1$ = q.$item_1$, . . .p.$item_{k-2}$ = q.$item_{k-2}$, p.$item_{k-1}$<q.$item_{k-1}$
   ii. Generate all (k-1)-subsets from the candidate itemsets in C_k
   iii. Prune all candidate itemsets from C_k where some (k-1)-subset of the candidate itemset is not in the frequent itemset L_{k-1}
II. Scan the transaction database to determine the support for each candidate itemset in C_k
III. Save the frequent itemsets in L_k

Algorithm in this illustration is plotted to mine all frequent item sets in database. Multiple iterations are performed to searches in database for findingfrequent item sets where k-item sets are used to generate k+1-itemsets. Figure-2 illustratesimplementationof Apriori Algorithmto achieve frequent items set.

VI. CONCLUSION

Recommendation system is aninnovative interactive technology for fetching information that can provide a different platform for growth by assisting customer in searching items of their need on the basis of their environment and behavior. Recommendations on item set that occurs frequently will add new demission by providing associativity in item (or any subset) occur at least as frequently. Associated items in an item set will provide scope for recommending item set to a customer in place of individual items by helping customers to find products which they want to buy plus enabling then to pick product which they must buy. Conversely, they also help business by generating more sales, increasing their revenue.

VII. FUTURE SCOPE

With explosion of IT and increasing numbers of user with access to internet recommendation system holds a big future in E-commerce through web platform. In addition to this most of the e-commerce companies introduced their apps for handheld devices, this will increases the need and scope of recommended system as they will try to expose relevant products to their customers with minimum number of clicks. With years to come recommendation system will influence sales done through e-commerce platforms.

Consideration of Apriori algorithm for the implementation opens a new scope for item set bases recommendation

ISSN: 2278 – 7798 All Rights Reserved © 2015 IJSETR

system. Initial level of dry run of proposed system shows positive results. In addition to this, algorithm can be extended to combine logic to recommend product based on associativity on the base of usability of the product searched.

REFERENCES

[1] Agrawal, R., Imielinski, T., and Swami, A. N., "Mining association rules between sets of items in large databases." , In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, 1993.

[2] D. Burdick, M. Calimlim, and J. Gehrke., "MAFIA: a maximal frequent itemset algorithm for transactional databases. "In Intl. Conf. on Data Engineering, Apr. 2001.

[3] Greg Linden, Brent Smith, and Jeremy York, "Amazon.com Recommendations tem-to-Item Collaborative Filtering", Industry Report, Published by the IEEE Computer Society, pp.76-79, January & February 2003.

[4] Howard Hamilton, "Apriori Itemset Generation", Computer Science 831: Knowledge Discovery in Databases, 2012.

[5] J.B. Schafer, J.A. Konstan, and J. Reidl, "E-Commerce Recommendation Applications," Data Mining and Knowledge Discovery, Kluwer Academic, pp. 115-153, 2001.

[6] Jeff M. Phillips, "L13: Frequent Itemsets", Data Mining; Spring 2013, University of Utah

[7] Karandeep Singh Talwar, Abhishek Oraganti, Ninad Mahajan, PravinNarsale,"Recommendation System using Apriori Algorithm","IJSRD - International Journal for Scientific Research & Development, pp.183-185, Vol. 3, Issue 01, 2015.

[8] M. Boley and H. Grosskreutz. Approximating the number of frequent sets in dense data. Knowl. Inf. Syst., pages 65–89, 2009.

[9] Rakesh Agrawal and RamakrishnanSrikant, "Fast Algorithms for Mining Association Rules", IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120.

[10] Recommender Systems Handbook, Ricci, F.; Rokach, L.; Shapira, B.; Kantor, P.B. (Eds.), 2011, Springer.

[11] Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl,J., "Analysis of Recommendation Algorithms for E-Commerce. In Proceedings of the ACM EC", Conference. Minneapolis, MN. pp. 158-167, 2000.

[12] Takeaki Uno, Masashi Kiyomi, Hiroki Arimura, "Efficient Mining Algorithms for Frequent/Closed/Maximal Itemsets", Information Science and Technology, Hokkaido University, Japan.

ABHISHEK SAXENA has received Master of Science (M.S.) degree in Software Systems from Birla Institute of Technology & Science, Pilani and Bachelor of Technology degree in Computer Science & Engineering from UP Technical University. Currently, he is working as a Module Lead in Xavient Information Systems, Noida.

Navneet K Gaur has received M. Tech from Karnataka Open University and Master in Computer Application. Currently he is working as a Project Manager in Xavient InformationSystems, Noida. Some of the contributions are as below;

Dimensional Risk Assessment Matrix for ICT Risk Assessment, Conference on "Network Security Issues and Resource Optimization", GEU Dehradun, Sponsored by UCOST, Uttarakhand, October 28, 2012

Traditional Security and Cryptography cut offs for Secure and Trusted Distributed Pervasive Environment, Serials Publications, ISSN: 0974-5807, February 26, 2010

Addressing the Challenging Issues of Cryptography & Network Security, National conference on "Information Technology trends in Engineering Applications" Organized by Center for Post Graduate studies in Engineering, & IEEE - DSCE Student Branch Dayananda Sagar College Of Engineering, Bangalore, June 1, 2010

 All Rights Reserved © 2015 IJSETR