Target Advertising via Association Rule Mining

Asmita Joshi[1]
Amity, Noida
India

Dr. J.S.Sodhi[2]
Head-IT & CIO (AVP), Amity Group
India

Abstract: In data mining, association rule is a popular method for discovering interesting relations between variables in large databases. Association rule mining helps to find out the frequent interesting pattern for internet user. In data mining, Apriori is a classic algorithm for learning association rules. Apriori is designed to find the frequent patterns of interest. Our Proposed system can target the interest of consumer for improve the scalability of web advertising. In the proposed approach we find out the frequent patterns of interestingness of user and it will support add publisher's to publish the ads according to the preferences of user. Through our proposed idea publisher can publish the advertisement according to found patterns, proposed system improves the target marketing of products on internet.

Keywords: Apriori, Association rule, Target marketing, web advertising.

I. INTRODUCTION

Data mining is the technique that supports to extract relevant result for decision making from large amount of data. Data mining is the key step in the knowledge discovery process. We can divide Data mining tasks in two categories: Predictive and Descriptive. In the predictive tasks we predict the value of a specific attribute based on the values of other attributes, but in the descriptive task, we extract unknown and useful information like patterns, associations, changes, anomalies and significant structures, from large databases.

II. ROLE OF ASSOCIATION RULE MINING IN OUR APPROACH

Data mining supports different techniques of knowledge extraction for various kinds of predictions like clustering, classification, association rule mining, sequential pattern discovery and analysis. Now data mining play an important role in the field of researches, Business, analysis, Predictions. We use association rule mining technique to develop our application. It help us in describing and analyzing of data and presenting strong rules discovered in databases using different measures of interestingness. Agrawal [2] introduced association rules for showing regularities between products in large transaction data For example, the rule{diaper}=>{rash cream} It indicates if people are buying diaper than they can also buy rash cream. We are using association rule mining to show the relation between different items. It can find a correlation among products from the analysis of a large set of data. The association rule mining is also known as a market basket analysis. It generally works on to know the customers buying pattern. Market basket analysis shows the different combination of items in the customer's basket. Many Supermarkets use this technique. It also helps us to introduce new products in market.

Two basic concepts for association rules are:

Support -> Support helps us to measure how many transactions have such item sets that match both sides of the implication in the association rule [1].

Confidence -> Confidence indicates the certainty of the rule. This parameter count how often a transaction's item set matches with the left side of the implication with the right side. The item set that does not satisfies the above condition can be discarded [1].

We are using association rule mining in our proposed approach. We know that internet is a great source of marketing. The web pages have a specific area for marketing or advertising. Web advertisement is the field of research that always looking for the interest of customer. In our research we want to develop a system that introduces advertising with patterns.

III. RELATED WORK

Rakesh Agrawal, Tomasz Imielinski, Arun Swami [1] they proposed algorithm for discovering association rules between items in large databases of sales transactions. It's an efficient algorithm that generates all significant association rules between items in the database. The algorithm incorporates buffer management and novel estimation and pruning techniques. They also present results of applying this algorithm to sales data obtained from a large retailing company, which shows the effectiveness of the algorithm

Qiang Yang, Tianyili Ke Wang [3] proposed the approach for web servers to keep track of web users browsing behavior in web logs. They study prediction models that predict the user's next requests based on web-log data. The result of accurate prediction can be used for recommending products to the customers, suggesting useful links, as well as pre-sending, pre-fetching and caching of web pages for reducing access latency. They also study for build sequential classifiers from association rules obtained through data mining on large web log data.

Farah Hanna AL-Zawaidah, **Yosef Hasan Jbara [5]** present association rule mining approach that can efficiently discover the association rules in large databases. Their approach is derived from the conventional Apriori approach with some more features to improve data mining performance. They performed extensive experiments and compared the performance of their proposed algorithm with existing algorithms. Experimental results show that their approach outperforms then other existing approaches and it can quickly discover frequent item sets and effectively mine potential association rules.

Ankit R Kharwar, Viral Kapadia ,Nilesh Prajapati, Premal Patel [6] uses association rule mining for generally to know the supermarket or inventory management strategy. They use association rule mining for web usage mining. Web usage mining technique can identify the identity or origin of Web users along with their browsing behavior at a Website. They Combine web usage mining with the association rule mining to optimize the content of the serve log data. They used Apriori algorithm and generate association Rule from server log which are useful in many application like cache for web page, Marketing, Targeted Advertising etc.

Pramod Prasad, Dr. Latesh Mali [7] Elaborates the use of association rule mining in extracting patterns that occur frequently within a dataset and shows the implementation of Apriori algorithm in mining for transactional data set of super market.

IV. PROBLEM DEFINITION

We proposed the framework that will extract the noun or entity from the user input. The collection of extracted nouns stored in the manually maintained dictionary. This dictionary Contains the set of related nouns that will treat as a data source for apriori algorithm To find the strong rules .This strong rules help us to target the most frequent pattern for display through advertisement on internet.

Problem Definition can be divided into two parts:

1 First we have to find the customer needs by tracing there search statements, for this we extract the nouns from the user input from search statements.

2. In the second part we find the frequent patterns of relative product with the help of dictionary that maintain the set of extracted nouns and apply the Apriori algorithm for generation of strong rules.

<div align="center">V. METHODOLOGY USED</div>

(1) Extract the noun from Statements.

(2) Store the entities at publisher Database.

(3) Generate the Strong rules for related entities for display.

<div align="center">VI. PARSING (EXTRACTION OF NOUNS)</div>

Noun Phrase Extraction (NPE) one of the most critical components of task in NLP. Almost all the information retrieval systems use noun phrase extraction for entity identification. It is very necessary task for natural language processing. In our proposed approach we are using Parsing for extracting the noun from the specific search statement given by the user. Noun phrase extraction is considered as a base of many Natural Language Processing. A noun phrase is can be accompanied by a set of modifiers. Modifiers include determiners, adjectives, prepositional phrases, and relative clauses.

Examples of some noun phrases in simple sentences:

"The (**DET**) red (**ADJ**) ball (**NN**)"

"The (**DET**) books (**NN**)

Nouns are those for whom we are talking about for example **'books', 'ball' .**We are using Penn tree bank approach for parsing for extraction of nouns from the statements. This Technique based on machine learning technique. Tree bank approach uses tag set for making difference b/w difference noun and modifiers of statement.

<div align="center">VII. BASIC ALGORITHM</div>

The Apriori algorithm was proposed by R. Agrawal and R. Srikant in 1994 [1] for mining frequent item sets to obtain strong association rules. A frequent item set is a set of transactions that occurs with a minimum specified support. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set properties. A strong rule is one that satisfies both minimum support and minimum confidence. Apriori algorithm uses scanning of database in iterative manner , where existing item sets (an item set that contains k items) are used to explore $k+1$ item sets.[1]

<div align="center">VIII. OUR PROPOSED APPROACH</div>

1. We take a search Statement as input for our system.

2. Now, Extracting noun (N_L) from the search statements.

3. We put it (**LHS= N_L**)

4. We find the relation with joining the existing nouns in the dictionary.

5. (N_{LU} **Nr**) remove the products N_k which don't have correlation with others.

6. N_{LU} **Nr** = P_i ^{is} the set of correlative products.

7. Display the ad's from the set of Pi.

Here we are using concept of sequential association rules. We put the referenced noun in LHS and finding the corresponding pattern from the right hand side.

LHS (Diapers) $\longrightarrow$ RHS (Rash cream) = Obtained Prod. Set (Diapers, Rash cream)

If someone is looking for the babies diapers on internet then we can suggest our user for Rash cream also. With the help of above formula we can create some frequent product set.
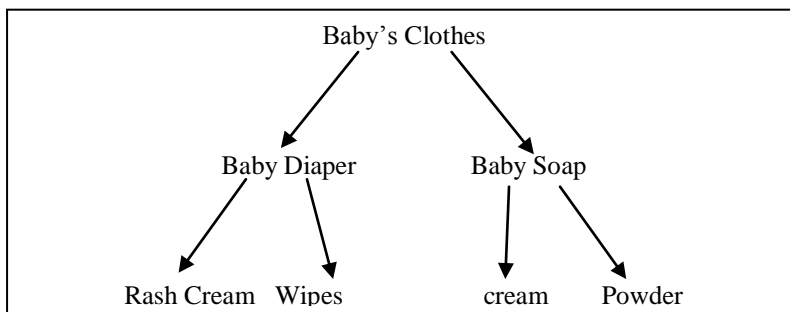
Fig 1: Correlative Product Set

We have prepared the Dictionary for storage of extracted nouns; it is a collection of different nouns those are related to each other. This structure supports the sequential pattern of preferences. Extracted Noun is put at the left side and existing nouns are put at right side then we create the set of items and remove those items that are not preferable in specific set. Our derived algorithm applied with following dictionary is prepared for context 'Baby Products'. Now we can find our strong rules for different Context and sub context like 'baby bathing products', 'baby toys', With the help of above figure we can justify our sequential rule of finding frequent pattern of preferable products.

Table 1: Correlative Products on Extracted Noun and Related Noun

Extracted Noun (N$_L$)	Related Noun (Nr)
Diaper	Shampoo, Rash cream
Rash cream	Diaper, Baby soap,
Baby Powder	Rash cream, baby shampoo
Baby Shampoo	Baby Soap, Diaper
Toys	Baby walker, doll
Baby walker	Doll

We implement the following data set of values in Weka. Weka is mining tool that can generate strong rules to show the sequential pattern of interest. We implement the working of the Apriori algorithm in Weka. We utilized a CSV file that contains the set of nouns.

Table 2: Support and Confidence obtain from transaction data

If Antecedent, Support then Consequent	Support	Confidence	Support X confidence
If buy deo then buy powder	6/14=42%	6 /7=85%	.36
If buy perfume then buy powder	5 / 14=35%	5 / 6=83%	.29
If buy perfumes then buy deo	4 / 14=28%	6 / 6=100%	.22
if buy shampoo then buy conditioner	4 / 14=28%	4 / 5=80%	.22
if buy conditioner then buy shampoo	4 /14=28%	4 / 5=80%	.22
if buy perfume and powder then buy deo	4 / 14=28%	4 /5=80%	.22
if buy perfume and deo then buy powder	4 /14=28%	4 / 5=80%	.22

IX. RESEARCH ANSWER

We proposed a system that can be effective to target large number of consumer in the market by giving them different correlative products. Our experiment results shows strong rule like **if A is true then B is true**. Above result justify the fact that if customer is looking for a then he/she may be interested in B Our Proposed system would increase the sale of product. Our system required observation of searching pattern of customer. The above related product may be known as a most frequent product of any purchasing list or they may be publicized as **sequential pattern of interest** at internet.

X. CONCLUSION

Our Proposed research is to make target marketing more reliable as we mentioned above Apriori algorithm is use to find the interest of customer in purchasing from super market or to predict the interest of user on internet. We use this approach with target web advertising. The publishers generally use traditional advertising.

References

1. Rakesh Agrawal, Tomasz Imielinski, Arun Swami, "Mining Association Rules between Sets of Items in Large Databases", Proceedings of the 1993 ACM SIGMOD Conference Washington DC, USA, May 1993

2. Agrawal, R. and Srikant, R. 1995. Mining sequential patterns. In Proceedings of the 1995 Int. Conf. Data Engineering, Taipei, Taiwan, pp. 3–14.

3. Qiang Yang ,Tianyili Ke Wang, "Building Association-Rule Based Sequential Classifiers for Web-Document Prediction"2004.

4. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter the WEKA Data Mining Software: An Update, Reutemann, IA H. Witten, SIGKDD Explorations, Vol.1, and Issue 1, 2009.

5. Farah Hanna AL-Zawaidah, Yosef Hasan Jbara "An Improved Algorithm for Mining Association Rules in Large Databases"2011

6. Ankit R Kharwar,Viral Kapadia ,Nilesh Prajapati, Premal Patel "Implementing APRIORI Algorithm on Web serve log" ,2011

7. Pramod Prasad, Dr. Latesh Mali "Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions"2011 Ana Azevedo and Manuel Filipe Santos, "A Perspective on Data Mining Integration with Business Intelligence", Information Science Reference, IGI, 2011.

8. Conceptual Model of Business Value of Business Intelligence Systems, Ales Popovic, Tomaz Turk, Jurij Jacklic, Journal of Management, Vol. 15, 2010.

9. Nizar Mabroukeh and C. Ezeiefe, "A Taxonomy of Sequential Pattern Mining Algorithms", ACM Computing Surveys, Vol. 43, No. 1, Article 3, Nov. 2010.

10. M. Tiwari, "Data Mining: A Competitive tool in Retail Industries", Global Journal of Enterprise Information System, vol. 2, Issue 2, December 2010.

11. E. W. T. Ngai, Li Xiu and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification", Elsevier, Expert Systems with Applications. vol. 36, p. 2592–2602, 2009.

12. The WEKA Data Mining Software: An Update, Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten, SIGKDD Explorations, Vol.1, Issue 1, 2009.

13. K. Shyamala and S. P. Rajagopalan, "Mining Essential and Interesting Rules for Efficient Prediction", AJIT, Vol. 6, 2009.

14. Rajesh Natarajan and B. Shekar, "Tightness: A Novel Heuristic and a Clustering Mechanism to Improve the Interpretation of Association Rules", IEEE IRI, Las Vegas, Nevada, USA, 2008

15. Tong Gang, Cui Kai and Song Bei, "The Research & Application of Business Intelligence System in Retail Industry", Proceedings of The IEEE International Conference on Automation and Logistics Qingdao, China September 2008.

16. Andreas Fink et al., "Advances in Data Analysis, Data Handling and Business Intelligence", Proceedings of the 32nd Annual Conference of the Gesellschaft für Klassifikation e.V., Joint Conference with the British Classification Society (BCS) and the Dutch/Flemish Classification Society (VOC), Helmut-Schmidt-University, Hamburg, 2008.

17. Chunhua Ju and Dongjun Ni, "Distributed Mining Model and Algorithm of Association Rules for Chain Retail Enterprise", ISECS International Colloquium on Computing, Communication, Control, and Management, 2008.

18. Hamid Rastegari and Mohd. Sap, "Data Mining and E-Commerce: Methods, Applications and Challenges", Journal of Information Management, Vol. 20, 2008.

19. Hongwei Liu, Bin Su and Bixi Zhang, "The Application of Association Rules in Retail Marketing Mix", Proceedings of the IEEE International Conference on Automation and Logistics, Jinan, China, 2007.

AUTHOR(S) PROFILE

Ms. Asmita Joshi is working as an Asst. professor in CSE dept. in CET-IILM-AHL Greater Noida, UP, India. She is pursuing M.Tech in CS from Amity University, Noida.

Asmita et al.,

International Journal of Advance Research in Computer Science and Management Studies
Volume 2, Issue 5, May 2014 pg. 256-261

Dr. J.S.Sodhi Currently working as Head-IT&CIO(Assistant Vice President) with AKC Data Systems (An Amity Group and AKC Group Company).Doctorate in Information Security Management and Certified Security Compliance Specialist (CSCS).His area of interest is IT Management.