



An Integrated Package for Market Segmentation in E Commerce

Ms Ashwini M

PG Scholar, Department of computer science and Engineering, Maharaja Institute of Technology (MIT), Mysore, Karnataka, India

Ms Kavya Ravishankar

Assistant professor, Department of computer Science and Engineering, Maharaja Institute of Technology (MIT), Mysore, Karnataka, India

Abstract—

Electronic Commerce is process of doing business through computer networks. A person sitting on his chair in front of a computer can access all the facilities of the Internet to buy or sell the products.

Unlike traditional commerce that is carried out physically with effort of a person to go & get products, ecommerce has made it easier for human to reduce physical work and to save time. Customer satisfaction is one of the challenges facing e-commerce today & there is still a lot of advancement made in the field.

One way to do this is to target customers with the particular offers most likely to attract them back to the store and to spend more time and money on their next visit. Demographic market segmentation is an approach to segmenting markets. A company divides the larger market into groups based on several defined criteria. Age, gender, marital status, occupation, education, locality and income are among the commonly considered demographics segmentation criteria.

Association rule mining was used as a method for identifying customers buying patterns and as a result customer profiles were determined.

Keywords—Demographic segmentation; Association Rules; customer segmentation;

I. INTRODUCTION

E-Commercial is a transaction of user, over internet. Electronic commerce is trading in products

or services using computer networks, such as Internet.

Today, most supermarkets record sales and collect customers' shopping details via a card dedicated to customer who holds customer's personal information (e.g. age, gender, job, income). Data mining helps using this huge amount of data in an efficient way and provides statistical information, thus predicting future customer behavior. One of the most important data mining methods is, also used in this study, association rule mining. The main purpose of this method is to determine correlations among the sales of items using a set of customer transactions on items. Association rule mining is also known market basket analysis. Market basket analysis helps to understand about the sets of items that are likely to be purchase together. In this paper, some questions were explored such as "Which products are commonly purchase together?" Generally it is being sought the dependency between two products X and Y. This information will be gathered by processing given a transaction database which is huge in size. This information will be used for store layout, promotions, discounts, catalog design, etc. These results will be analyzed whether they are related to customers' data (age, gender, income, marital status, locality etc.). According to the surveys made by the companies which use recommendation system, product recommendation is a milestone which helps the company to increase revenue by 15-40%. According to a Forrester study, 15% of the visitors declare to

buying recommended products. According to a recent survey, 74% of consumers get annoyed when the websites. They visit feature content, offers, and ads that don't match their interests. The main objective is meant for market segmentation that gives services according to the customer's interest/Profile. The prediction is based upon customer's age, gender, marital status, occupation, education and income.

II. EASE OF USE

This part shows all steps (Fig1) from start to end of data mining process, including gathering raw data, normalizing, preparing raw data to be processed, as well as processing data with data mining software. After data is completely processed, results are shown in specific sub-part. Prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

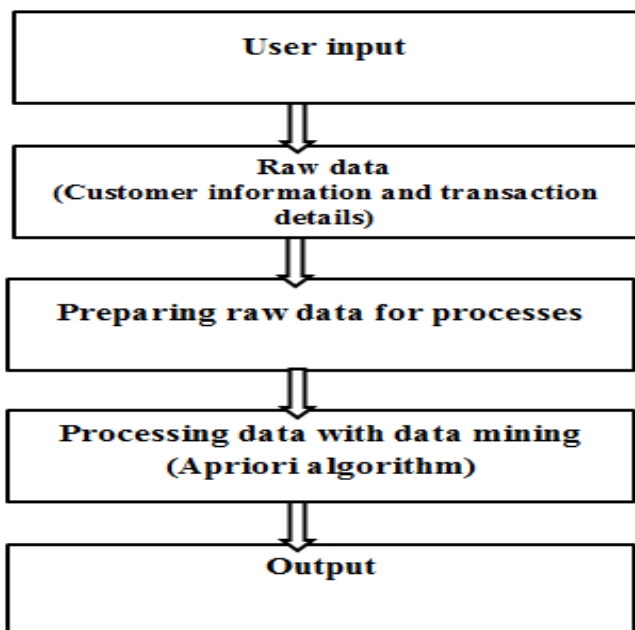


Fig1. Overall process

III. RELATED WORK

A. Using Association Rules for Course Recommendation

This paper is on recommending system based on Association Rules (RARE) In this paper it is used to recommend the academic courses by using former students experience and current students rating. To find unrated items there are many techniques like content based, collaborative and hybrid filtering. Another approach to recommendation is data mining based recommendations

Six steps followed are business understanding, data understanding, data preparation[selection, cleaning, construction, integration], modeling, evaluation, deployment from this steps we can avoid cold start problem.

Different databases used for this course recommendation they are: Former student base-collection of data which is collected using data mining technique process. Rule base- collection of rules which are generated by data mining process. User base - student information.

B. Recommendation with Association Rules: A web mining application

In this paper it shows how to build recommendation models using association rules and describe how to build recommendation models in order to improve the usability of the site. Our recommendation models are sets of association rules. Then it measures the performance of the models with different metrics on different levels of detail of the dataset.

One of the methods to build recommendation models that can create recommendations to every user, as user navigates the site, according to the user visits in a session. The idea to construct the recommendation model by producing association rules from the web data. Then user visited to the page is matched with the creator of the rules. The consequents of the like rules with the maximum confidence become the Recommendations.

Data preparation: Preparing the data has required by the application. Standard data system is used for collection and preparation of data.

Although association rules are not the greatest common recommendation system, in past have been used and adopted for the work described. The performance of the Association Rules-based recommendation models on the datasets result is acceptable, in the sense that they depart consider the random recommendation. In terms of the application as a complete, data preparation has been an exact laborious task particularly because of the dynamics of the website structure and the trouble in obtaining certain answers to our business accepting and data understanding questions.

IV. DATA MINING TECHNIQUES

Most of the literature survey papers of Data Mining have made use of various classification techniques in order to extract hidden information. It can be categorized into four groups:

1. Decision trees – The algorithms used are ID3, CART, C4.5, Rep tree, Random tree, CHIAD. Decision tree Authors and Affiliations techniques are simple to understand and interpret, It allows addition of new possible scenarios, It helps to determine worst, best and expected values for different scenarios, It can be combined with other decision tree techniques to generate rules easily. This technique has many disadvantages as the number of training data increases like over fitting. It does not handle numeric data and pruning may become cumbersome.

2. Bayesian Classifier- It includes Naive Bayes algorithm and its variants. This technique is simple and easy to understand, requires small amount of training data to estimate the parameters, Fast and Space efficient, Insensitive to irrelevant features and handles both real and discrete data well. The major disadvantage of this technique is that it assumes conditional independence among the attributes.

3. Neural Networks- It includes algorithm like Multi-layer Perceptron. This technique is a generalized method, works well with noise. But it does not scale well from small research system to large real-time system. It is computationally expensive and it does not guarantee a solution. It mainly depends on the algorithm used to train the system.

4. Clustering Techniques- It includes algorithms like K-means, Nearest Neighbour etc. This technique is relatively scalable and simple. But they are highly sensitive to noise and outliers.

The proposed system is the real time, web based application which has to take any number of training data as an input. So Bayesian classifiers are best suited as it easy to implement and insensitive to irrelevant features.

V. DATA MINING PROCESSES

Data mining is defined as a process of discovering hidden valuable knowledge by analyzing large amounts of data, which is stored in databases or data warehouse, using various data mining techniques such as machine learning, artificial intelligence(AI) and statistical.

Data understanding: Review the data that you have, documented it; identify data management and data quality issues.

Data preparation: Get your data ready to use for modeling

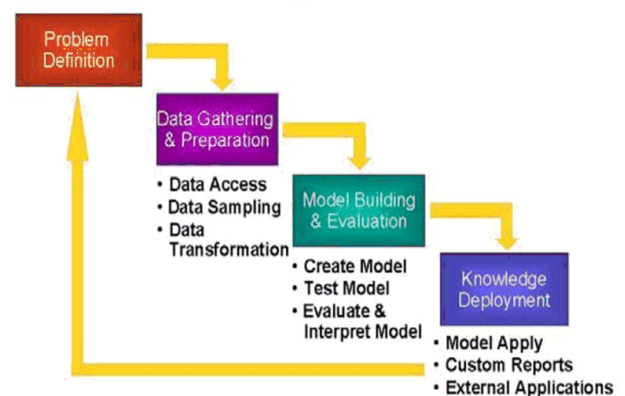


Fig2. Data mining Process

Modeling: Use mathematical techniques to identify patterns within your data.

Evaluation: Review the patterns you have discovered and assess their potential for business use.

Deployment: Put your discoveries to work in everyday business.

VI. APRIORI ALGORITHM

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

Apriori is designed to operate on database containing transactions (for example, collections of items bought by customers, or details of a website frequentation).

Frequent Item sets: All the sets which contain the item with the minimum support (denoted by L_i for i^{th} item set). Apriori Property: Any subset of frequent item set must be frequent. Join Operation: To find L_k , a set of candidate k -item sets is generated by joining L_{k-1} with itself. Is explained in Fig 4.

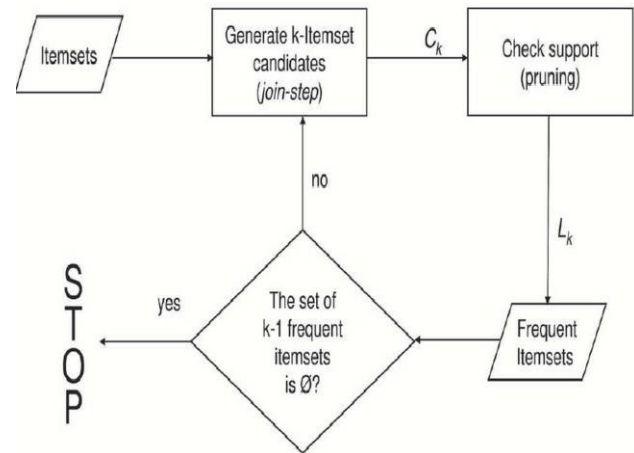


Fig3: Algorithm flowchart

The Apriori Algorithm(Fig 3) : Pseudo Code

1.Join Step: C_k is generated by joining L_{k-1} with itself

2.Prune Step: Any $(k-1)$ -itemset that is not frequent cannot be a subset of a frequent k -itemset

3.Pseudo-code : C_k : Candidate itemset of size k

L_k : frequent itemset of size k

4. $L_1 = \{ \text{frequent items} \};$

for $(k = 1; L_k \neq \emptyset; k++)$ **do begin**

$C_{k+1} =$ candidates generated from L_k ;

for each transaction t in database **do**

increment the count of all candidates in C_{k+1}

that are contained in t

$L_{k+1} =$ candidates in C_{k+1} with min_support

end

return $\cup_k L_k$;

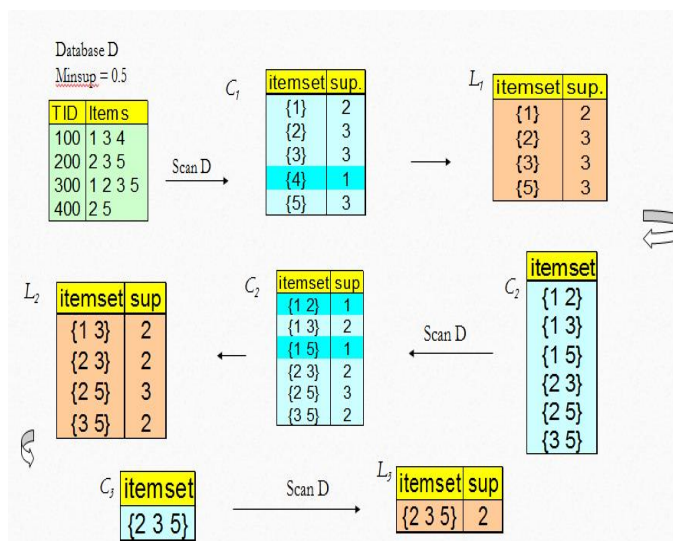


Fig 4: Algorithm example

Conclusion:

In common, today's commerce must continuously struggle to generate the next best thing that consumers need since consumers endure to wish their products, services etc. to constantly be better, faster, and cheaper. In this world of novel technology, businesses want to provide accommodations to the novel types of consumer requirements and developments because it will prove to be dynamic to their business' success and survival. The system "An integrated package for market segmentation in E Commerce" is developed and tested successfully and satisfies all the requirement of the client.

The goals that have been achieved by the developed system are: Maximize the Business Profit. Increase the customer satisfaction by not giving unrelated recommendation. The system gives services according to the customer interest/profile. It reduces the recommender system challenges like sparsity, scalability and cold start problem. As we use only last one month transaction for recommendation it helps to recommend product for change in customers' interest.

REFERENCES

- [1] Dr. Sankar Rajagopal, "Customer Data Clustering Using Data Mining Technique", International Journal of Database Management Systems (IJDBMS) Vol.3, No.4, November 2011.
- [2] Dr.K.Duraiswamy & B.Jayanthi, "A Novel preprocessing Algorithm for Frequent Pattern Mining in Multidatasets", International Journal of Data Engineering (IJDE), Volume (2) : Issue (3) : 2011.
- [3] Priyanka Asthana, Anju Singh, Diwakar Singh, "A Survey on Association Rule Mining Using Apriori Based Algorithm and Hash Based Methods", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013.
- [4] Jeetesh Kumar Jain, Nirupama Tiwari, Manoj Ramaiya, "A Survey: On Association Rule Mining", International Journal of Engineering Research and Applications (IJERA) www.ijera.com Vol. 3, Issue 1, January -February 2013.
- [5] T. Karthikeyan and N. Ravikumar, "A Survey on Association Rule Mining", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 1, January 2014.
- [6] Jochen Hipp, Ulrich Guntzer, Gholamreza Nakhaeizadeh, "Algorithms for Association Rule Mining – A General Survey and Comparison", ACM SIGKDD, Vol 3, Issue 1, July 2000.
- [7] Masoumeh Mohammadnezhad1 and Mehregan Mahdavi, "An Effective Model For Improving The Quality Of Recommender Systems In Mobile E-Tourism", International Journal of Computer Science & Information Technology (IJCSIT) Vol 4, No 1, Feb 2012.
- [8] Kyoung-jae Kim, Hyunchul Ahn, and Sangwon Jeong, "Context-aware Recommender Systems using Data Mining Technique", International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering Vol:4, No:4, 2010.
- [9] V.L. Miguéis, A.S. Camanho, João Falcão e Cunha, "Customer data mining for lifestyle segmentation", Elsevier, 2012.
- [10] David Stern, Ralf Herbrich, Thore Graepel, "Matchbox: Large Scale Online Bayesian Recommendations", International World Wide Web Conference Committee (IW3C2), April 20–24, 2009.



[11] Alípio Jorge, Mário Amado Alves, Paulo Azevedo, “*RECOMMENDATION WITH ASSOCIATION RULES: A WEB MINING APPLICATION*”, Department of de Informatics, University do Minho, Portugal.

[12] Narimel Bendakir and Esma Aimeur, “*Using Association Rules for Course Recommendation*”, American Association for Artificial Intelligence (www.aaai.org), 2006.

[13] Hasan Ziafat *, Majid Shakeri, “*Using Data Mining Techniques in Customer Segmentation*”, Hasan Ziafat Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 9(Version 3), September 2014.

[14] Rakesh Agrawal, Ramakrishnan Srikant, “*Fast Algorithms for Mining Association Rules*”, Proceedings of the 20th VLDB Conference Santiago, Chile, 1994.