Analyzing Target Customer Behavior Using Data Mining Techniques for E- Commerce Data

Dattatray V. Bhate , M. Yaseen Pasha

Abstract— **In this scenario Customer satisfaction is no longer satisfied with a simple listing of marketing contacts, but wants detailed information about Customers past purchase as well as prediction of future purchases. From long period of ago, information of customer is most important for every business. With the help of advance information technology, firms are able to collect and store mountains of data describing infinite offering and different type of customer profile. With the help of this information we are able to find customer needs and wants. Electronic commerce or ecommerce is a term for any type of business, or commercial transaction that involves the transfer of information across the Internet. It covers a range of different types of businesses, from consumer based retail sites, through auction or music sites, to business exchanges trading goods and services between corporations. It is currently one of the most important aspects of the Internet to emerge. Ecommerce allows consumers to electronically exchange goods and services with no barriers of time or distance. Electronic commerce has expanded rapidly over the past five years and is predicted to continue at this rate, or even accelerate. Traditional forecasting methods are no longer suitable for these business situations. For that type of business we are able to use the principles of data mining concept. With the help of data mining principle huge amount of customer information into cluster customer segments by using K-Means algorithm which is used to cluster observations into groups of related observations without any prior knowledge of those relationships and data from web log of various ecommerce websites.**

Keywords: Data mining , K-Means algorithm, E-commerce

I.   INTRODUCTION

It may be observed that the revenue of any organization is customer satisfaction. Every industry or organization tries to increase the satisfaction of customer and also even tries to increase the expectation of different types of customer. In addition to increase new customers, predicting potential buyers, firms are devoting a great deal of resources to delighting and retaining existing target customers to cultivate a long-term, close relationship with them.

At present, internet technologies are widely used by customer for purchase different types of goods for those Internet technologies have seamlessly automated interface processes between customers and retailers, retailers and distributors, distributors and factories, and factories and their suppliers. E-commerce is changing the face of most business functions in competitive enterprises. In the context of e-commerce, it is not that much of easy to generating large-scale real-time data. For that reason there are numerous opportunities for gathering customer information in electronic form. By using Data Mining it generate non obvious yet useful information for decision makers from very huge databases. The various mechanisms of this generation include abstractions, aggregations, summarizations, and characterizations of data. These forms, in turn, are the result of applying sophisticated modeling techniques from the diverse fields of statistics, database management, artificial intelligence, and computer graphics.

Even after gathering consumer information, but still how to analyze these data effectively is of interest to marketers and researchers. The traditional methods for analyzing and predicting customer demands have found a wide range of applications. They are mainly used for predicting the total quantity of products that belong to the same family rather than the relationship between the different customer groups and associated product groups. By using K-Means algorithm it generate clusters customer segments and data from web log of various ecommerce websites. It is clear that, the results showed that there was a clear distinction between the segments in terms of customer behavior. It is seen that this data mining model can used as an efficient vehicle for firms not only to predict the products or services that should be provided or improved for their target customer groups, but also to identify the right customers for a specific product family or service.

II.   LITERATURE REVIEW

Transacting or facilitating business on the Internet is called ecommerce. Ecommerce is short for "electronic commerce." Popular examples of ecommerce revolve around buying and selling online. But the ecommerce universe contains other types of activities as well. Any form of business transaction conducted electronically is ecommerce.
Examples of E commerce:

A. *Online shopping*

Buying and selling goods on the Internet is one of the most popular examples of ecommerce. Sellers create storefronts that are the online equivalents of retail outlets. Buyers browse and purchase products with mouse clicks. Though Amazon.com is not the pioneer of online shopping, it is arguably the most famous online shopping destination.

B. *Electronic Payments*

When you are buying goods online, there needs to be a mechanism to pay online too. That is where payment processors and payment gateways come into the picture. Electronic payments reduce the inefficiency associated with writing and mailing checks. It also does away with many of the safety issues that arise due to payment made in currency notes.

C. *Online Auctions*

When you think online auction, you think eBay. Physical auctions predate online auctions, but the Internet made auctions accessible to a large number of buyers and sellers. Online auctions are an efficient mechanism for price discovery. Many buyers find the auction shopping mechanism much interesting than regular storefront shopping.

D. *Internet Banking*

Today it is possible for you to perform the entire gamut of banking operations without visiting a physical bank branch. Interfacing of websites with bank accounts, and by extension credit cards, was the biggest driver of ecommerce.

E. *Online Ticketing*

Air tickets, movie tickets, train tickets, play tickets, tickets to sporting events, and just about any kind of tickets can be booked online. Online ticketing does away with the need to queue up at ticket counters.

E-commerce can change the face of business. It provides better customer management, an expanded range of product, new strategies for marketing, and more efficient operations. A key enabler of this change is the widespread use of increasingly sophisticated data mining tools. Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, costs, cuts or both.

Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Primarily Data mining is technique used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data. The term 'data mining' is used to describe the process of analyzing a company's internal data for customer profiling and targeting. In e commerce application, the end goal of data mining is to improve processes that contribute to delivering value to the end customer.

At the basic level, the information available in web log files can illuminate what prospective customers are seeking from a site. Are they purposefully shopping or just browsing? Buying something they're familiar with or something they know little about? Are they shopping from home, from work, or from a hotel dialup? The information available in log files is often used to determine what profiling can be dynamically processed in the background and indexed into the dynamic generation of HTML, and what performance can be expected from the servers and network to support customer service and make e-business interaction productive.

E-commerce data are classified as usage data, content data, structure data, and user data. Usage data provide details information about user sessions and page views. The content data in a site are the collection of objects and relationships that are conveyed to the user. Structure data represent the designer's view of the content organization within the site. Structure data also include the intra-page structure of the content represented in the arrangement of HTML or XML tags within a page. The user data may include demographic or other identifying information on registered users, user ratings on various objects such as pages, products, or movies, past purchase or visit histories of users, as well as other explicit or implicit representations of users' interests. Once the data types are clear, data preparation is easily achieved. The author then proposes association rules, sequential and navigational patterns, and clustering approaches for personalization of transactions as well as web pages. Liuying Shen and Jana Hawley describe an approach to predict user behavior in e-commerce sites. The core of their approach involves extracting knowledge from integrated data of purchase and path traversal patterns of past users (obtainable from web server logs) to predict the purchase and traversal behavior of future users.

Most of the algorithms presented in the literature to deal with clustering web sessions treat sessions as sets of visited pages within a time period and do not consider the sequence of the click-stream visitation. This has a significant consequence when comparing similarities between web sessions. Quinlan, J.R proposes an algorithm based on sequence alignment to measure similarities between web sessions where sessions are chronologically ordered sequences of page accesses. Nonetheless, reviews and research in this area are handicapped by the proprietary nature of the data and algorithms. A great deal of effort is being expended in this area, but most of it is secret. Certainly Amazon, Google, and Microsoft are deeply

engaged in statistical research, and in time the broader research community may learn more about their findings. But for now, all this paper can really be done is to lay out the main strategies in the relevant areas.

III. DATA MINING MODEL

A systematic method was used to collect e-commerce transactions from web log file of e-commerce site. There were 2518 transactions collected from 1st to 31st of December 2013. Each session gave detail information about web accessed from each user such as the web sites, requested web pages and their order, and the period of time pages were viewed. This data is very important to analyze the customer.

A. Preprocessing of Data

Data preprocessing techniques are used to increase the quality of collected data, this can help to increase the accuracy and efficiency of the subsequent mining processes. It is clear to see that data preprocessing is an important step in the knowledge discovery process, as high-quality decisions must be based on high-quality data. Detecting data anomalies, rectifying them early, and reducing the data amount to be analyzed can result in substantial benefits and advantages for the decision makers. For usage profiles, a session file from data preparation stage was used as input in data mining. While K-means algorithm was applied for the purpose of clustering some incomplete data were deleted. The usable data were 2363 transactions.

B. Data Analysis

Data mining software, Mineset was used in data analysis. The usable customer data of 4263 ecommerce transactions were divided into two groups. Group 1 was about 70% of the total transactions and was used as training data. Group 2 was about 30% of the total transactions and was used as testing data. Five factors used in data segmentation included: age, gender, online in time, address, language, and target customer behavior type.

1. Age was divided in 6 kinds
Age 1 - 11 years old to 15 years
Age 2 - 16 years old to 20 years
Age 3 - 21 years old to 25 years
Age 4 - 26 years old to 30 years
Age 5 - 31 years old to 35 years
Age 6 - 36 years old to 40 years

2. Gender was divided in 2 kinds
Gender 1 - man.
Gender 2 - woman.

3. How long online was divided in 4 groups
Group 1 -0.00 hours to 05.59 hours
Group 2 -06.00 hours to 11.59 hours
Group 3 -12.00 hours to 17.59 hours
Group 4 -18.00 hours to 23.59 hours

4. Online Address
Address 1 - At work
Address 2 - At home

5. Language
Language1 - China
Language2 – English

6. Target customer behavior
Behavior 1 -Buying computer products
Behavior 2 -Buying cloth
Behavior 3 -Buying gifts and flower
Behavior 4 -Buying books
Behavior 5 -Buying CDs and DVDs
Behavior 6 -Buying toys and children product
Behavior 7 -Buying airline tickets and other tickets
Behavior 8 -Online trading
Behavior 9 - Online banking

C. Result analysis

K-Means algorithm used to transfer user data into cluster. Results were segmented in five clusters. Based on the results of customer usage, e banking transactions can be classified into 5 clusters.

Cluster 1: The male customers who are 21 to 30 years old used personal computers to purchase computer products and books. They accessed ecommerce site via personal computers between 6.00 hours to 17.59 hours at home. The language was Chinese.

Cluster 2: This was the second smallest cluster. The male Customers were online between 6.00 hours to 17.59 hours. They accessed e-commerce site to buy books, airline tickets and other tickets at work. The language used was Chinese and their age was from 26 to 35.

Cluster 3: This cluster gained a majority of the ecommerce application since it had the largest population (26%). By using personal computers in Chinese language at home or at work, the female and male Customers accessed e commerce site to purchase books, CDs and DVDs, airline tickets and other tickets. The online time was between 12.00 hours to 23.59 hours and their age was from 21 to 30.

Cluster 4: The male customers who are 11 to 20 years old used personal computers at home. They accessed e-commerce to purchase gifts (including Jewelry) and flowers, toys and children products between 0.00 hours to 05.59 hours. The language was Chinese.

Cluster 5: This was the second largest cluster, where female customers used personal computers for ecommerce between 6.00 hours to 17.59 hours. The language used was Chinese and transactions were bill payments. Their age was from 21 to 30 and they purchase cloth, toys and children products.

Cluster 6: Male customers who are 31 to 40 years old used personal computers for e commerce site between 12.00 hours to 23.59 hours at work. This was the only cluster where customers used English. The major behaviors were online trade.

IV. CONCLUSION

Data mining is technique primarily used today by companies with a strong Customer focus - retail, communication, financial and marketing organizations. Data mining is having lot of importance because of its huge applicability. It is being used increasingly in business applications for understanding and then predicting valuable data, like Customer buying actions and buying tendency, profiles of Customers, industry analysis, etc. Data Mining is used in several applications like market research, Customer behavior.

E-commerce companies are shifting from the old world of mass production where standardized products, homogeneous markets, and long product life and development cycles were the rule to the new world where variety and customization supplant standardized products. Instead of tens of thousands of products in a superstore, consumers may choose among millions of ones in an online store to satisfy the personalization demands. It is clear that target customers marketing can be effective when a e-commerce company is able to collect rich information about buyers behavior on ecommerce site. According to this study, the majority of the customers in e-commerce were male and online period was between 6 hours to 17.59 hours. The age is also an important factor that affects customer behavior. This can cause a market segment in the e-commerce. Personal computers at home are more popular for ecommerce than at work that can be attributed to the convenience of online purchasing at home. Chinese is more popular than English because Cluster 6 which used English was about 16%. It shows that some customers begin to access English e-commerce site to engage in international trade.

V. REFERENCES

[1] Liuying Shen, Jana Hawley, etc. "E-commerce Adoption For Aupply Chain Management in U.S. Apparel manufacturers". Journal of Textile and apparel, technology and management, vol.4, no.1, 2004, pp.1-10.

[2] N.R. Srinivasa Raghavan. "Data Mining in E-commerce: A Servey". Sadhana, vol,30, no.2, 2005, pp.275-289.

[3] Mobasher. B. Web Usage Mining and Personalization. Practical Handbook of Internet Computing (ed.) M P Singh (CRC Press), 2004.

[4] Randall S. Sexton, Richard A. and Michael A. "Predicting Internet/e-commerce use", Internet Research, vol.5,2002, pp. 402-410.

[5] Strader, T.J., Shaw, M.J. Electronic Markets: Impact and implications, in: Shaw, M., Blanning, R., Strader, T., Whinston, A. Handbook on Electronic Commerce, Springer, Berlin, 2000, pp.77-98.

[6] Quinlan, J.R., Induction of Decision Trees, Machine Learning, Morgan Kaufmann Publishers, Inc, CA, 1990.

[7] John A. Rodgers, David C. Yen and David C. Chou. Developing E-business: a Strategic Approach, Information Management & Computer Security, no.4, 2002, pp.184-192.

Dattatray V. Bhate received his B.E. in Computer Engineering from university of pune in 2009, pursuing M. Tech in CSE from JNTU, Hyderabad. Presently he is working in ZES's Dnyanganga College of Engineering and Research, Pune from 2 years. Areas of research interest in Data Mining, Text Mining.