

Customer Segmentation in Online Electronic Retail Store based on Customer's Expectations as Values using Social Media Analytics

Swetha S

Dept. of ISE, R. V. College of Engineering, Bangalore, India

Abstract

Customer segmentation for online electronic retail store is very important because the services are provided based on the customer segments generated. The improper customer segments force the online electronic retail store to service wrong customers. The traditional ways of customer segmentation like behavioral segmentation, demographic segmentation, and attitudinal segmentation have disadvantages because of which the customer segments are not much effective. Only with the availability of customers' personal information and transactional database it is not possible to understand the expectations. Here we proposed a value based segmentation focuses on the value of customers. Value of customer segmentation is based on the customers' Expectations, Needs, and wants and how the online electronic retail store can serve these which make use of social media, a rich source of about customers' expectations. Since the social media data is dynamic, the information is collected regularly. The value based customer segmentation is built by merging the social media information with the available customer details of a online electronic retail store into a data warehouse using dimensional modeling. Then the analytics is done on the aggregated data using neural networks and k-means algorithm. The resultant customer segments will be dynamic and more valuable for decision making process compared to traditional methods.

Keywords

Customer segmentation; Social Media Analytics(SMA); Data Warehouse; Neural Networks; and K-means

I. Introduction

Online stores have become a necessary market for customers with an intention to purchase what they want. Initially, most customers were suspicious about whether it was good for them to purchase products through online stores. However, as the Internet is becoming more commonly used in various areas, online shopping is taking root as a part of our lives nowadays. A great number of online stores appear each day in cyber due to the entrance barriers, which offer both an opportunity and a threat to these stores. A gradual decrease in their shares of the pie forces online stores to look for ways to survive in the fiercely competitive online market. Online stores have recognized that they need to present personalized, differentiated, and customized services and products to their customers for competing against their rivals in the online market. In general, most customers make a decision to purchase in online stores by relying on information through a computer screen. The purchasing behaviors of customers are strongly influenced by the information and the services that are provided by the online store. Online stores need to understand these customers' behaviors to develop and maintain relationships with customers. Thus, they need to evaluate the value of customers and select customer segmentation methods in order to design and establish different strategies to maximize the value of customers[2]. Customers have varying preferences; so customer segmentation can help online

stores to serve their customers according to their preferences.

Customer segmentation is a research area that intrigues researchers in data mining because vast data are suitable for data mining and patterns or rules are identified as a result of data mining that could play a vital role in marketing. Data mining has been applied for segmenting customers by demographic and transaction data that are available just by searching databases with regard to sales between stores and customers[3].

However, data that are selected from databases do not include anything related to customers' beliefs that also may strongly affect customers' intention to purchase. Thus, researchers have interested in the factors that affect the behavior of customers in the decision to purchase. Most studies related to consumer behavior have presented their research model using the factors and tested it statistically with data that are gathered from survey [6]. Although a lot of studies have contributed to explaining consumer behavior, they have not sought to utilize customers' psychographics for segmenting customers using data mining. The reason for this omission is that the psychographic data that are needed for data mining are stored in customers' minds, and not in a well-formed database. This study has proposed an value based approach that enables online electronic retail store to segment their customers using data mining techniques based on their expectation collected by social media.

Advantages of the proposed method is it Explores the social media information to know the customer behavior and it integrate social media information along with available customer information helps to know what exactly the customers want and how online electronic retail store can serve the customers.

II. Traditional Customer Segmentation Methods

The traditional approach of segmentation attempts to segment the market using the behavior patterns of customers, the location and environment of the customers, based on how profitable the customers are, and how the customers think of the products. Each of these four approaches to segmentation contributes an important piece in understanding the market; yet, each of these approaches displays significant limitations[1].

Behavioral Segmentation is the effort to cross-tabulate or associate behaviors with known demographics. It is the oldest method of segmentation. In essence, it classifies a customer into a "bucket" based upon whether customers are similar to other customers who have performed that behavior in the past. Historically, behavioral segmentation has been generated by cross-tabulation analysis, or more recently by tree-based classification tools such as CHAID or CART. The problem with these methods is, they have been superceded by data mining (modeling-based) approaches to segmentation. the newer data mining methods produce a probability score for the customer or prospect. This allows the segmentation analyst to prioritize within a "bucket" who is most likely to perform the behavior (e.g., purchase a given product)[4].

Demographic and life-cycle Segmentation attempts to determine customers based on different combinations of demographics. This

is historically important and intuitive approach to segmentation because customers do buy different products at different stages of their life. For instance, first mortgages are highly associated with prospects in their mid- to upper-20s, and home equity loans are highly associated with customers in their 40s with children entering college. The major difficulty with life-cycle segmentation approaches is product-specificity. Since products differ by life phase, life-cycle segmentation must be developed for each product [7].

Profitability or current value segmentation tries to determine the customers that are most profitable. This segmentation is based on the transactions previously made and the personal profile. This will identify the customers who have made huge number of transactions, or who have made transactions with huge amounts at regular intervals, or whose personal profile shows that they are rich and chances of huge transactions are more. In other way, if the personal profile is profitable based on the kind of service then it will relate with behavioral segmentation but the attitude of these cannot be predicted just by their profile which will not be fixed. Moreover, it restricts only to available customers and never expands to new customers who can be profitable [11].

Attitudinal Segmentation segments the market based on how well customers perceive the product or service to be performed. The rationale behind this type of segmentation is very strong; it recognizes that attitudes drive behavior. The difficulty lies in the fact that most approaches to attitudinal segmentation utilize only performance data. While performance data does tells us how respondents think we are doing, it tells us precious little about what respondents actually will do. Another type of data, importance or preference data is much more suited for projecting what a respondent will do. Moreover, two other limitations of attitudinal segmentation should be mentioned. Most attitudinal segmentations are developed using cluster analysis. Cluster analysis is a very powerful approach, but it is "brittle." No single cluster method always works best, but often only one particular type of cluster analysis is run on the data. Secondly, even if the cluster is well separated from the other clusters, attitudinal segments are generally more difficult to classify by simple cross-tabulation, as compared to behavioral clusters [5].

III. Proposed Method of Customer Segmentation in Online Electronic Retail Store

The proposed method of customer segmentation is Value-Based Segmentation which emphasizes on how valuable the customers are. The value is analyzed in terms of satisfaction of customers, profitability of customers, customers who give valuable feedbacks about services, or the combination of these customers. The value is increased further by knowing what value in terms of the customers is. Customers usually find the value in what they actually want. For example just because a customer whose income/family income is low is profitable, it is not appropriate to deliver the item. The focus here is on what the customer wants. So this segmentation works on the value with respect to the customer and not on the producer. First the customers who find the value in the products being produced should be found. Then they should be segmented based on different demographics [8-9].

IV. Data Warehouse and Generic Method of Analytics for Customer Segmentation

Online electronic retail store normally have a huge transaction database in which the data is normalized to increase the efficiency of updating the transaction database. But this transaction database

is not suitable for analysis and decision making [10]. To perform analysis the transaction database is modified into a de-normalized database called data warehouse. Architecture of a data warehouse is shown in fig. 1:

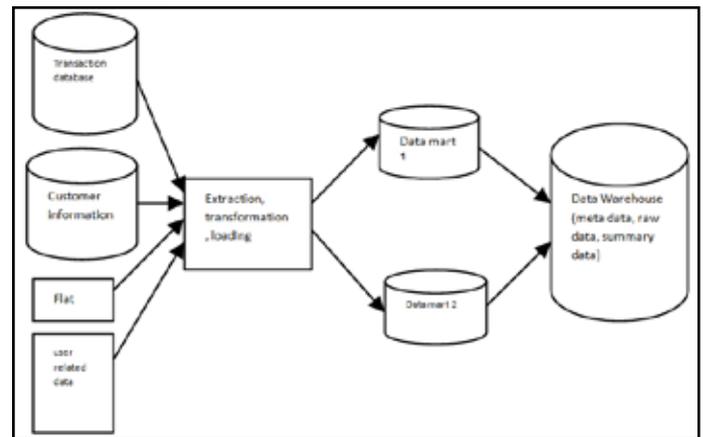


Fig. 1: Architecture of a Data Warehouse

The transaction data, customer information and user related data are collected together and ETL process is applied where the data is extracted and transformed into a form that is easy to analyze and loaded. The data marts are created for specific user requirements using star schema or snow flake schema and a fact table. All the data marts are collected together to form a data warehouse consisting of meta data, raw data and the summary data [12]. Once the data warehouse is available then the analytics is performed on the data warehouse. The Analytics is the process of obtaining an optimal or realistic decision based on existing data. The goal of analytics is to improve the business by gaining knowledge which can be used to make improvements or changes. The simple definition of analytics is the science of analysis. The general architecture of analytics is shown in fig. 2:

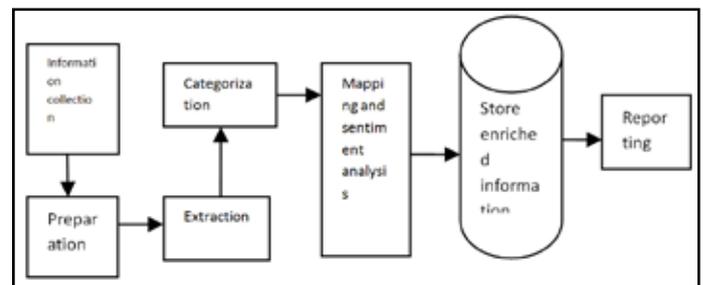


Fig. 2: Architecture of a Analytics

As shown in fig. 2, the data is collected and prepared for analysis. The data is extracted and categorized on the basis of requirements. Using the dictionary of the subject the mapping and sentiment analysis is performed using different mapping and data mining algorithms. The informative data is stored and displayed in different forms to the user.

V. Segmentation Process

Customer segmentation of online electronic retail store is done by considering two terms attributes and concepts. The customers were divided based on their characteristics as done in traditional methods based on visits, preferences, profitable, etc., and each such characteristic is called as attribute [12-13]. The set of customer attributes $A = \{A_1, A_2, A_3, \dots, A_n\}$ describes the customer characteristics and behaviors which constitute n dimensional attribute space A_n of customers. The set $G = \{G_1, G_2, \dots, G_n\}$

represents the different pre defined concepts like ‘buy an item’, based on all the customer characteristics like ‘preferences’ as P different concept dimensions. The P concept dimensions constitute P dimension concept space G^p of the customers. A can determine a set of concepts G which is shown as a relation $A^m \rightarrow G^p$.

This determination of concepts G from the attributes A is performed using Self Organizing Mapping (SOM) neural networks and K-means algorithm [7]. The K-means algorithm is supportive and very effective for large data warehouses, but its performance is very much dependent on previous data. Because we are using the $A \rightarrow G$ relation based on assumptions and not on facts, to avoid this un-wanted dependency we are using SOM which gives a stable and dependable relation to the K-means algorithm. The results obtained by the K-means algorithm are satisfactory. The only problem with this approach is that user has to specify the number of clusters (k) and the initial clustering centre (z).

VI. Customer Segmentation Based on Customer’s Expectations as Values Using SMA

The proposed value based customer segmentation is done by collecting the social media information involving customer values[10]. These value based information is integrated with available customer information and a data warehouse is created on top of this data. The analytics is done as specified in the segmentation process above to obtain the customer segments. These customer segments are used for efficient decision making by the online electronic retail store.

The segmentation process is shown in fig. 3

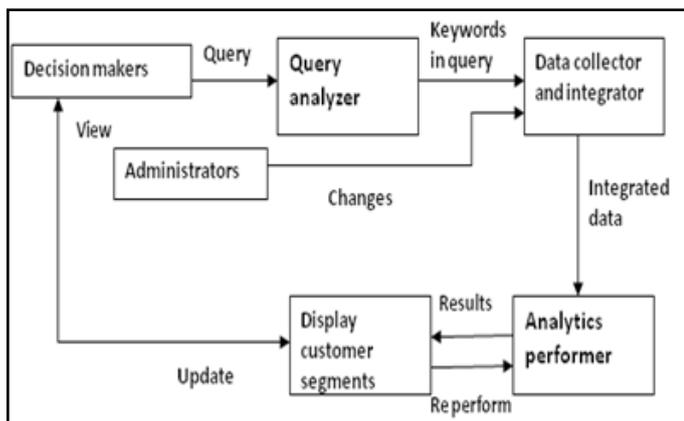


Fig. 3: Proposed Value Based Customer Segmentation

The decision makers will give the query with the initial values for k and z. after analyzing the query the query analyzer will give the keywords to data collector and integrator which will collect information from social media and integrate with available information. The analytics is performed on integrated data to get customer segments which can be again re-modified till users are satisfied.

A. Algorithm for Information Collector Module

Input

search string.

Output

The Information Collector module stores the collected data into separate tables and outputs the same.

The Information Collector module consists of 4 main functions.

- Identify Keywords
- Check Internet Connection
- Collect information from twitter

- Collect information from facebook

Start

Identify the keywords in the request query

Check for internet connection, if not connected establish connection

Set source to twitter

Collect RSS feeds of keywords in request from twitter search feed.

Parse the RSS feeds in XML format using XML parser and collect the values of each feed in separate row of the table and mention title for each column.

Change the source from twitter to facebook

Give keywords as input to graph API of facebook and save the result in JSON format.

Use JSON parser and parse

each message block into the table.

step 10. Stop.

B. Algorithm for Data Integrator Module

Input

information tables of social media sites as input.

Output

data warehouse.

The data integrator module consists of 5 main functions

- Join twitter and facebook tables
- Connect to database and collect master data
- Analyze data and form a model
- Create a data warehouse
- Integrate and load master data, transaction and media data into data warehouse

Start

Take the media information tables collected from twitter and facebook. Merge them into a single table called media data table.

Connect to database and collect the master data of online electronic retail store

Remove non-customer information from media data table

Analyze data and form a data model

Create dimension tables and fact table

Connect dimension tables and fact table according to the data model and form a data warehouse

Integrate and Load the data into the data warehouse.

Stop

C. Algorithm for Analytics Performer Module

Input

The Analytics Performer module uses the data warehouse, list of services of online electronic retail store .

Output

The Analytics Performer module displays customer segments for each service.

The Analytics Performer module consists of 5 main functions

Analyze message in the fact

Relate it to suitable category

Relate the category to service of online electronic retail store

Segment the customers for each service of online electronic retail store

Build a dashboard to view and personalize customer segments.

Set count =0

For each fact up to count = N

Analyze the message in the fact

Relate it to suitable category of messages

Check whether the categorization is satisfactory or not.
 If not satisfied go to step a.
 If satisfied, relate the category to a service of a
 online shopping mall

Count = Count +1

Segment the customers into groups, each representing customers for the service of a online electronic retail store
 Develop a dashboard to view customer segments
 Provide features to drill down the customer segments and analyze them.

VII. Results

The fig. 4 shows the analysis done on the data in the Data Warehouse against various dimensions to obtain business intelligence and it also shows the table and bar chart representing the number of messages collected from social media sites for each category. The pie chart shows the percentage of messages for each category. The fig. 4 shows the trend of variation in number of messages posted by customers for each category.

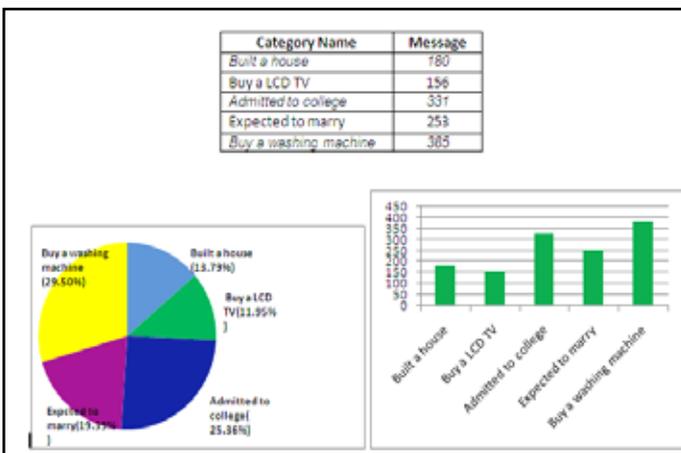


Fig. 4: Analysis of no. of Messages with Respect to Categories

The fig. 5 shows the table and bar chart representing the number of messages collected from social media sites in each month December 2010, January 2011 and February 2011 respectively. The pie chart shows the percentage of messages for each month. The figure shows the trend of variation in number of messages posted by customers across time. It also seen that increase in number of messages as time progresses. This shows the increase in use of social media by customers.



Fig. 5: Analysis of No. of Messages Against Time

The fig. 6 shows the table and bar chart representing the number of messages collected for each category from social media sites facebook and twitter respectively. It shows the trend of variation in number of messages posted by customers for different categories

in different social media sites.

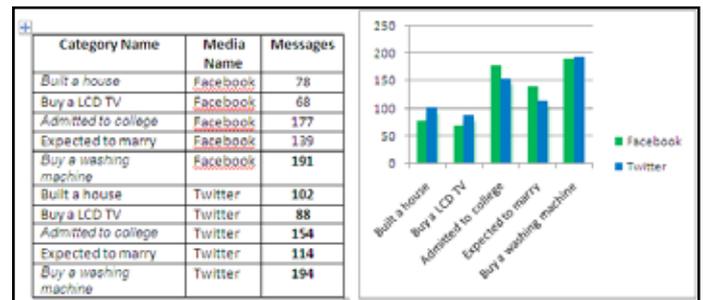


Fig. 6: Analysis of No. of Messages for each Category from different Media

The fig. 7 shows the table and line chart representing the number of messages posted from different locations in different intervals of time. The figure shows the trend of variation in number of messages posted by customers at different locations in different months. In December 2010 the messages were very few. But we can see in January 2011 and February 2011 the total number of messages from all cities are more and there is slight variation in number of messages across various locations.

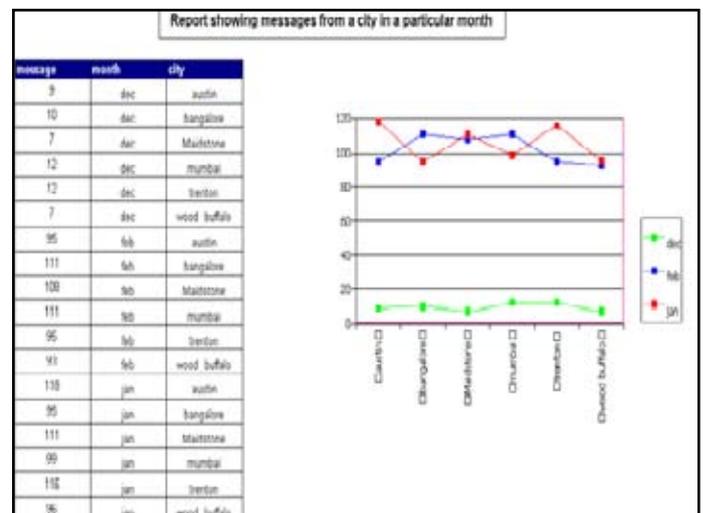


Fig. 7: Analysis of No. of Messages in various Locations across Time

VIII. Conclusion

This paper provides an approach to segment customer which collects the social media information of customers having keywords that express need of a service, online electronic retail store can provide. It thus avoids unwanted social media information and reduces information overload and integrates social media information of customers with transaction data and master data with the online electronic retail store and forms a data warehouse in star schema with dimensions and facts. Thus information is structured and de-normalized making the analysis easier. self organized clusters are formed by applying SOM neural network to the initial data. Instead of giving random centroids initially to k-means algorithm, the centroids of these clusters are given as input to the k-means algorithm. So, obtained customer segments are well separated and effective.

The customer segments can be customized and seen in the form of various graphs for better understanding and suitability. The Customer Segmentation on integrated data gives Customer Segments for each Service of the online electronic retail store. The obtained customer segments show us that the customers

belonging to a range of demographics say, in income group of 20 k to 30k wants to buy a washing machine in their social media information. The online electronic retail store can deliver the washing machine as a service for these customers. The online electronic retail store can drill down this customer segment and deliver the item only to those customers who have already done shopping atleast once .

The advantages of the proposed approach are It Exploring the social media information to know the customer behavior and The integration of social media information along with available customer information helps to know what exactly the customers want and how online electronic retail store can serve the customers. limitation of our method is Only the textual information is used from social media for analysis. In Future work can be extended to get the social media identity of customers to use their social media information which is assumed in the our approach.

References

- [1] Hung, C., Tsai, C. F., "Market segmentation based on hierarchical selforganizing map for markets of multimedia on demand". *Journal of Expert Systems with Applications*, Vol 34 Issue 1, 2008, pp 780–787.
- [2] Chen, Y., Zhang, G., Hu, D., & Fu, C., "Customer segmentation based on survival character". *Journal of Intelligent Manufacturing* , Vol 18 Issue 4, 2007, pp 513–517, 2008.
- [3] Shin, H. W., Sohn, S. Y., "Segmentation of stock trading customers according to potential value", *Journal of Expert Systems with Applications*, Vol. 27 Issue 1, 2004, pp. 27–33, 2008.
- [4] Everard, A., & Galletta, D. F., "How presentation flaws affect perceived site quality, trust, and intention to purchase from an online store", *Journal of Management information Systems*, Vol. 22, Issue 3, 2006, pp. 55–95.
- [5] DeLone, W. H., McLean, E. R., "The DeLone and McLean model of information systems success: A ten-year update", *Journal of Management Information Systems*, Vol. 19, Issue 4, 2003, pp. 9–30.
- [6] Verhagen, T., Dolen, W., "Online purchase intentions: A multi-channel store image perspective", *Information and Management*, 46(2), 2009, pp 77–82.
- [7] W.H.Lnmon, *Data Warehouse*, 3rd ed., Beijing: Mechanical Industry Press, Aug. 2005, pp. 68-75.
- [8] Fang Anru, Ye Qiang, Lu Qi, LI Yijun, "Customer segmentation framework model based on data mining", *Computer Engineering*, Vol. 35, Oct. 2009, pp. 251-253.
- [9] Rong-Shiunn Wu, Po-Hsuan Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach", *journal of Electronic Commerce Research and Applications* Vol. 10, Issue 3, 2011, pp. 331–341.
- [10] Zeng, Daniel Chen, Hsinchun Lusch, Robert Li, Shu-Hsing, "Social Media Analytics and Intelligence", *Intelligent Systems, IEEE*, Nov.-Dec. 2010, Vol. 25, Issue 6, pp.13.
- [11] Xiong Weiwen, Chen Liang, Zhang Zhiyong, Qiu Zhuqiang, "RFM Value and Grey Relation Based Customer Segmentation Model in the Logistics Market Segmentation", *Computer Science and Software Engineering*, 2008 International Conference, 12-14 Dec.2008, Vol. 5, pp. 1298 – 1301.
- [12] "The Shortcut Guide to achieving Business Intelligence in Mid Size Companies", white paper by IBM, February, 2011, Length 82 Pages.

- [13] Art Weinstein, "Market segmentation", revised ed., vision books, Nov. 93.



Swetha S. received her B.E. degree in Information Science and Engineering from Siddaganga Institute of Technology, Tumkur under Visvesvaraya Technological University, Belgaum in 2006, the M.Tech degree in Computer Science and Engineering from Dr. Ambedkar Institute of Technology, Bangalore under Visvesvaraya Technological University, Belgaum, in 2009. She worked as Lecturer in Department of CSE, SIET, Tumkur. She is presently working as Assistant Professor, Department of ISE, RVCE, Bangalore. Her research interests include Data Mining, Network Security.