International Journal of Management, IT & Engineering
Vol. 7 Issue 10, October 2017,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

CUSTOMER SEGMENTATION USING SOCIAL MEDIA DATA

Dr. Neetu Narwal*

Abstract— Customer Segmentation was introduced by Smith(1956), thereafter it has become prominent part of marketing strategies and practices followed by different organization. The organizations are using different techniques to gather data required for segmentation of customer, which is costly in terms of data gathering and analysis. This paper proposes a methodology that helps the organization to use the data available in social media to segment the customer base belonging to specified demographic area.

Keywords—Social Media Mining, Customer Segmentation, Classification Techniques.

* Asst. Prof. , Maharaja Surajmal Institute

I. INTRODUCTION

Social Media is considered as fastest and largest source of data in the present scenario. On an average around 6,000 tweets are tweeted every second on Twitter, which corresponds to over 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year [1]. With the availability of such huge source of data, it has become challenging to store the data and perform mining on the data and extract useful information out of it. Various organizations have been using these data to know about the reviews of customer about their product and services. They are using it as information base to formulate their business strategies and marketing plans to target such group of customers. And also this information is used to improve their product and services based on customer remarks.

Social media is being used by retail and marketing department of large enterprise companies and it has become a promising field of research. Its applications include customizing advertising campaigns, localizing unexplored market segments, and projecting sales trends.

In this paper we propose a methodology that uses the social media data to extra information about the customer. It makes use of some user related features to segment the customer base. Using some well known Clustering techniques these user groups are partitioned into segments. These Customer segments are further analyzed to provide certain demographics and psychographic details. The use of social media data will eliminate the need to collect customer related data which is generally done in the form of questionnaire, online or offline survey or customer personal transaction logs. This methodology will in turn large organization to cut their cost in terms of time, money and efforts spend to collect customer dataset.

Customer Segmentation was introduced by Smith (1956), it has become an essential concept in marketing theory and practices. It has changed the focus of marketing strategies from manufacturing oriented to customer oriented.

Smith[2] stated “Market Segmentation involves viewing a heterogeneous market as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desires of consumers for more precise satisfaction of their varying wants”.

Smith recognizes that segments are directly derived from heterogeneity of customer needs.

For performing segmentation of the heterogeneous group of customers, there are bases/features on which the customers can be classified into homogeneous groups.

These features are categorized as:

General bases – These features are independent of product or services.

Product Specific bases – These features are related to both, customer and product, services and particular circumstances.

Frank, Massy and Wind(1972) first proposed the classification of segmentation bases into four categories:

Observable and General Bases – Cultural, geographic, demographic and socio economic variable

Observable and Product Specific Bases- User status, usage frequency, store loyalty

Unobservable and General Bases - Values, Personality, life style.

Unobservable and Product Specific Bases – Psychographic, benefits, perceptions.

Figure 1 shows the four broad categories of segmentation base used for performing customer segmentation.

These bases are used for segmenting customer into homogeneous groups using one of the various methods available for performing segmentation.

These segmentation approaches are classified into a-priori and post-hoc method. *A-priori* approach is used when the type and number of segments are determined in advance by the researcher. *Post-hoc* method is used when the numbers of segments are determined on the basis of the result of data analysis.

Another way of classifying segmentation approaches is descriptive or predictive statistical approach. *Descriptive method* analyzes the association among the set of segmentation bases/ variable, with no difference between dependent or independent variable. *Predictive methods* analyses the association between dependent and independent variable.

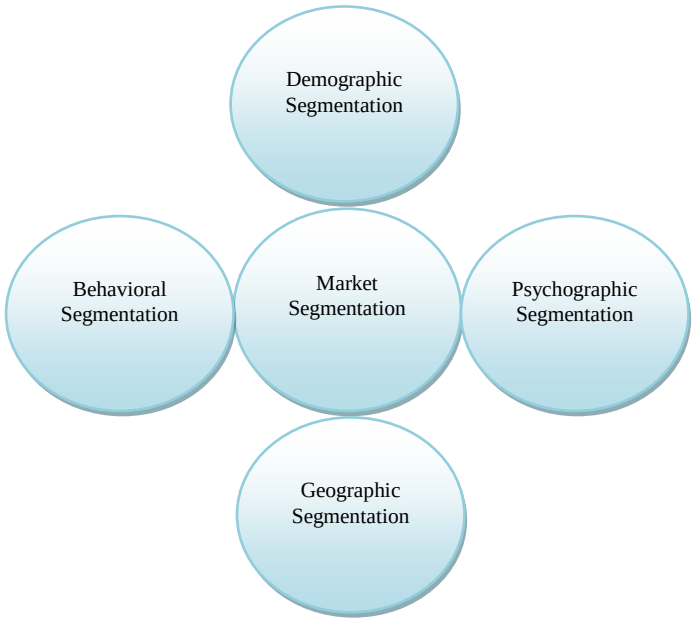

Figure 1: Types of Segmentation

Classification of segmentation methods are as shown in Table 1:

Apriori and Descriptive approach – Contingency table, log-linear models.

Apriori and Predictive approach – Cross Tabulation, Regression, logit and Discriminant analysis

Post-hoc and Descriptive approach - Clustering methods, Non overlapping, Fuzzy techniques, ANN.

Post-hoc and Predictive approach - AID, CRT, Cluster wise regression, ANN, Mixture model

Table 1: Classification of Segmentation Method

Method	**Apriori**	**Predictive**
Descriptive	Contingency tables, log liner models	Hierarchical Clustering, optiomal clautering, latent class cluster models
Predictive	Cross-tabulation, regression, discriminant analysis, decision trees, neural network	Auto associative neural network, latent class regression models

There are numerous applications of using customer segmentation in various industry scenarios and proposing appropriate marketing strategies based on the segmentation.

The rest of this paper is organized as follows: The next Section outlines related work done in the area of Customer segmentation and the use of social media for providing customer base; Section III, proposed methodology adopted in the system. Section IV concludes with some final remarks and directions for future work.

II. RELATED WORK

Customer segmentation is performed in different industries using different features extracted either using surveys, online questionnaires or using transactional data related to the customers. Some of the commonly used methodology prominently used by different industries has been studied and reviewed.

H Hwang et. al. [3] considered the domain of Customer Relationship Management (CRM) and focused on customer cultivation and retention. For the purpose they suggested a new lifetime value (LTV) model and customer segmentation by considering defection and cross-selling opportunity.

For study they collected six-month service data of a Wireless Company in Korea. The data was categorized as: socio-demographic data and usage information of wireless service data.

LTV models evaluate the long-term value of customer focused on the entire lifetime of customer. Since wireless industry is very sensitive to external environment and customer defection. The study focused on short-term value of customer.

For conducting customer segmentation they considered on three dimensions: current value, potential value and customer loyalty.

Current value is the profit contributed by customer during certain period.

Current Value = $\frac{\text{Avg. amount asked to pay – Cumulative amount in arrears}}{\text{Total Service Period}}$

Potential value = $\sum_{j=1}^{n} Prob_{ij} \; x \; Profit_{ij}$

Probability that customer_i would use the service j among n optional services.

Profit_{ij} that company can get from customer_i who uses optional services j.

Customer loyalty is calculated as

CL= 1- Churn Rate

Three perspectives on customer value are used to segment the customer base and they suggested marketing strategies based on the segments.

S. Peker et. al. [4] proposed a new RFM model, called length, recency, frequency, monetary and periodicity (LRFMP) for classifying customer in grocery retail industry.
RFM model was proposed by Hughes(1996) to analyze and predict customer behavior. RFM has been applied in customer segmentation and in a variety of industry health, textile, banking and tourism.
They proposed new features periodicity and modified recency.
Recency is calculated time interval (in days) between customer last visit data and last data of observation period.
Modified recency is calculated as average number of days between the dates of customer N recent visit and last date of observation period.
Periodicity reflects whether customer visits the store regularly. They define periodicity as standard deviation of the customer inter-visit times.
They used k-means algorithm to segment customer, the optimal value of clusters (k) is found using cluster validation indices to find the concept of compactness.
They used three cluster validity indices :

- Silhouette Index (SIL)
- Calinski Harabasz (CH) Index
- Davier Bouldin (DB) Index

They used the dataset of local grocery chain that operated in more than ten stores in Antalya, Turkey. The original dataset was extracted from company loyalty card system that contains 2 million purchase transactions of 16024 customers between the period of Oct 2012 to Aug 2014.
J. T. Wei et. al.[5] used LRFM model (length, recency, frequency, and monetary) model, by adopting self-organizing maps (SOM) technique for a children dental clinic in Taiwan to

segment its dental patients. SOM determines the best number of clusters is twelve among 2258 patients based on the characteristics of length, recency, and frequency. The average values of LRF are computed for each cluster and the overall patients, excluding monetary covered by NHI program. The values of LRF variables for each cluster greater than those of the overall average are identified. The results show that three clusters having the above average LRF values (454 patients) can be viewed as core patients.

They also used Customer Value Matrix 2x2 by combing length (L) and recency (R) to depict different clusters of customers to provide unique marketing strategies for each cluster of customers.

Social media data is being used in the past to find the meaningful information about the person behind the account. Various researchers have picked the data from social media sites and performed some useful analysis and were able to predict some information related to the user.

Drezde, et al. used Twitter feeds to predict some hidden features related to the user such as gender and ethnicity by performing clustering on observed attributes such as first name, last name, and friends list [6].

A research conducted in IBM Haifa Research Lab shown that ”using the same tags, bookmarking the same web pages, [and] connecting with the same people” are some features that clusters some like-minded people [7].

A Research Project from Pennsylvania State University partitioned users based on their levels of connectedness and engagement on social media, they found that there was significant difference amongst the clusters regarding willingness to interact with a company online [8].

III. PROPOSED METHODOLOGY

The methodology adopted in this work comprises of the following steps

Step 1 Data Gathering: This step gathers twitter data based on the latitude and longitude as required by the organization. The customer segmentation is done on the basis of area where the company needs to focus and device certain strategies for promoting their product or services. Social Media is used as data source to perform customer segmentation.

Step 2 Data Cleaning: After getting personal tweets of individual users the next step is data cleaning. The social media is a medium where people can express their views regarding any topic

and to gather information from the raw data is challenging. This step involves cleaning of the data so that some valuable information can be extracted from it.

Step 3 Data Classification: The individual tweets are present in clean form and it indicates the actual content discussed by the individual. We used this data to predict the personality of individual. The data is classified using some well known classification algorithm like Artificial Neural Network, Support Vector Machine, Radial Basis Neural Network. The personality traits of the customers are identified using classification algorithm.

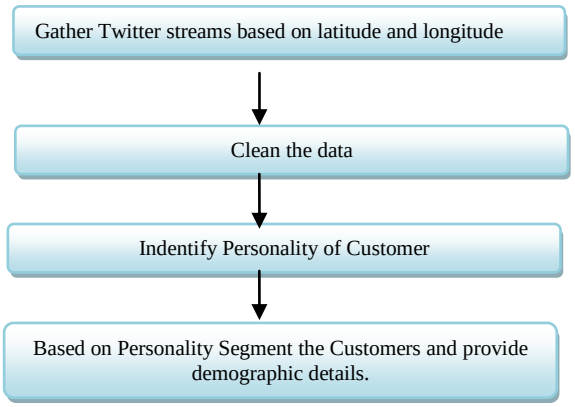

Figure 2: Methodology used in the research

Step 4: Customer Segmentation: The personality traits are used to segregate the customers into different clusters pertaining to different customers segments. Each customers belonging to specific cluster are related to each other in terms of their personality traits and behavior. Hence these formulate homogeneous customer segments which are of interest to an organization. Finally the segmented clusters are combined with their demographic details are provided in terms of location, gender, family size, annual income etc. which help the organization to gain insight into their customer and strategies their marketing plan to cater the needs for specific segment of customer.

CONCLUSION

Our system proposes a methodology that makes use of social media data to provide customer segmentation, as a result reducing the cost of conducting survey, online questionnaire and transaction log data. In future we will implement the system using R programming and using real time data from twitter to perform customer segmentation. Nowadays Customer segmentation is widely used in Industry to focus marketing strategies and plan committed towards individual

customers rather on focusing on whole community which are heterogeneous in terms of behavioral, psychographic details etc.

REFERENCES

[1] http://www.internetlivestats.com/twitter-statistics/ as seen on 28 Aug 2017.

[2] "Market Segmentation – Conceptual and methodological foundation", Michel Wedel, Wagner A Kamakura, 1999.

[3] Hyunseok Hwang, Taesoo Jung, Euiho Suh , "An LTV Model and customer segmentation based on customer value : a case study on the wireless telecommunication industry", , Experts System with Applications., 200

[4] LRFMP model for customer segmentation in the grocery retail industry: a case study", Serhat Peker, Altan Kocyigit and P. Erhan Eren, Marketing Intelligence and Planning, 2017.

[5] A case study of applying LRFM model in market segmentation of a children's dental clinic", Jo-Ting Wei a, Shih-Yen Lin b, Chih-Chien Wengc, Hsin-Hung Wuc, Expert System with Applications/

[6] S. Bergsma, M. Dredze, et al. Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter. 2013. http://www.clsp.jhu.edu/ sbergsma/TwitterClusters/

[7] I. Guy, M. Jacovi, et al. Same Places, Same Things, Same People? IBM Haifa Research Lab. 2010. http://dl.acm.org/citation.cfm?id=1718928

[8] B. Jansen, K. Sobel, et al. Classifying ecommerce information sharing behaviour by youths on social networking sites. Journal of Information Science. 2011.