

Market Analysis and Review Synthesis

Surabhi R. Gosavi^{1*}, Poonam R. Gohad², Priya P. Sandanshi³,
Sushmita T. Mahajan⁴ and Prof. N. R. Wankhade⁵

^{1*,2,3,4,5}Dept. of Computer, LGNSCOE, Nasik, University of Pune
surabhogosavi@gmail.com

www.ijcseonline.org

Received: 8 March 2014

Revised: 17 March 2014

Accepted: 25 March 2014

Published: 30 March 2014

Abstract- For the analysis of business, a lot of research attention in the field of computational statistics and data mining has been made. Due to recent technological advances in the field of data clustering, the researchers face ever-increasing challenges in extracting relevant information from enormous volumes of available data. The paper focus on large data sets obtained from online web visiting and categorizing this into clusters according some similarity it helpful tool for the top level management to take optimized and beneficial decisions of business expansion. Clustering is the assignment of a set of observations into subsets. Cluster analysis is widely used in market research when working with multivariate data from surveys. Market researcher partition the general population of consumers into market segments and understand the relationships between customers.

Keywords- System Architecture, Cluster, Porter Stemmer Algorithm

I. INTRODUCTION

The focus area of this system is market research and analysis. It is a web-based application and aims at determining target markets and consumer density and identifying potential customers. We have used the concept of Cluster analysis for the same. This application will help determine the user's browsing details and monitor customer population.

It becomes highly challenged to digital firms due to the exponential growth of customer data in today's electronic business environment. Hence market segmentation proves to be one of the possible solutions. The goals of market segmentation include retention of customer, allocation of advertising resource, and an increase of profit margins. The outcome of market segmentation plays important role in market development, advertising, product pricing, attracting new customers and other marketing strategies. A good analytical tool should be able to compare the characteristics of different segments (group) and identify important attributes of each segment (group) to create opportunity for targeting customer which can be accomplished by getting the complete context of the customer.

With this system interested to target the customer of my own web portal in major cities of Maharashtra. To find that from which location of the world the visitors of the website to trace the future market at that place. Maximum used services by this visitors also referral of the website and whole control of the system at administrator. The similar kind of facility available with Google Analytic Tools but for every domain we have to pay the different fees also the whole database with Google. Report formation not as per our requirement which is the major task in every system.

For the analysis of business, a lot of research attention in the field of computational statistics and data mining has been made. Due to recent technological advances in the field of data clustering, the researchers face ever-increasing challenges in extracting relevant information from enormous volumes of available data [6].

The paper focus on large data sets obtained from online web visiting and categorizing this into clusters according some similarity it helpful tool for the top level management to take optimized and beneficial decisions of business expansion. Clustering is the assignment of a set of observations into subsets. Cluster analysis is widely used in market research when working with multivariate data from surveys. Market researcher partition the general population of consumers into market segments and understand the relationships between customers. To achieve robustness and efficiency in data

clustering combine Partition and hierarchical clustering results.

II. SURVEY OF LITERATURE

For any project literature survey is considered as the backbone. Hence it is needed to be well aware of the current technology and systems in market which is similar with the system to be developed.

In recent years, data mining has gained widespread attention and increasing popularity in the commercial world. According to the professional and trade literature, more companies are using data mining as the foundation for strategies that help them outsmart competitors, identify new customers and lower costs. In particular, data mining is widely used in marketing, risk management and fraud control.

Although recent surveys found that data mining had grown in usage and effectiveness, data mining applications in the commercial world have not been widely. Literature about data mining applications in the fields about economic and management are still few. With the realization of importance of business intelligence, we need to strengthen the research on data mining applications in the commercial world.

The premise of data mining is that it is necessary to establish the enterprise-level customer-information data warehouse which can assemble the customer data and can provide a correct, complete and single customer data circumstance for instituting better customer service strategies.

It becomes highly challenged to digital firms due to the exponential growth of customer data in today's electronic business environment. Hence market segmentation proves to be one of the possible solutions. The goals of market segmentation include retention of customer, allocation of advertising resource, and an increase of profit margins. The outcome of market segmentation plays important role in market development, advertising, product pricing, attracting new customers and other marketing strategies.

A good analytical tool should be able to compare the characteristics of different segments (group) and identify important attributes of each segment (group) to create opportunity for targeting customer which can be accomplished by getting the complete context of the customer. Clustering is just as the normal saying that "things of one kind come together", that is, to classify a group of individuals into several groups. Its aim is to make the distance between the individuals of the same group as short as possible, and the distance between the individuals of different groups as far as possible. Clustering analysis can

find out the data distribution mode and proper inter relationship between data properties from the macroscopical view, search for the useful relation between the object data in certain data set and divide the records in database into a series of meaningful subclass.

The K-means algorithm is among the most popular clustering methods that group observations with similar characteristics or features together [5]. It is widely used in many marketing applications, especially in cluster-based market segmentation.[1]

III. PROPOSED SYSTEM

The proposed system consists of following modules:

1. Module for Admin: Whenever user visits a particular website his details (e.g. login details or account details) are sent to server of the website. Admin module uses this information stored in web server to generate reports. Admin is the owner of the website. He has all the access control. Ultimately he is responsible for the change in the website. Admin performs functions as follows:

1. Creation of account
2. Deletion of account
3. Updating of account
4. Restricted access control to other intruders or users.

2. Module for Web user analyzer: In this module the users (visitors) information is used for classification of the data or to form clusters of the similar data by using the various data mining techniques. This module is the heart of the variance rover system.

Main formation of the clusters is largely dependent on the data formation by hits of the visitors. In our system we are forming clusters by considering IP address of the visitor.

We can form clusters on the basis of the other attribute like product on the website etc. After forming clusters the classified data is then stored in the web server.

3. Module for Geo Location: Geo-Location Module helps to locate visitor's geographical location using inputs like longitude and latitude of the user from user details for locating him so as to form clusters. User details consist of IP, longitude and latitude. Then we apply these characteristics on Google maps for getting users geographic al location. Clusters are formed by tracking visitors region wise, so in reports it will be easy for organization to take business related decisions or to expand business.

4. Module for Report Generator: All the clustered data generated by web user analyzer is stored in the web server. That data is processed by this module and it can be used for market research and for deciding advertising strategies to enhance the business and to maximize profit as well as to satisfy customer needs.

IV. ALGORITHMIC STRATEGIES

NEAREST NEIGHBOR ALGORITHM

The Nearest neighbor algorithm is one popular partitionial algorithm for data clustering. The *Nearest neighbor* algorithm assigns each point to the cluster whose centre (also called *cancroids*) is nearest [4]. The centre is the average of all the points in the cluster — that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster.[3]

The algorithm steps are:

- 1). Choose the number of clusters, k .
- 2). Randomly generate k clusters and determine the cluster centres, or directly generate k random points as cluster centres.
- 3). Assign each point to the nearest cluster centre, where "nearest" is defined with respect to one of the distance measures discussed above.
- 4). Recompute the new cluster centres.

- 5). Repeat the two previous steps until some convergence criterion is met (usually that the assignment hasn't changed).

The main advantages of this algorithm are its simplicity and speed which allows it to run on large datasets. Its disadvantage is that it does not yield the same result with each run, since the resulting clusters depend on the initial random assignments. From the algorithmic framework, we can see that the algorithm need to adjust the sample classification continuously, and calculate the new cluster centres constantly. Therefore, the time consumption is fairly considerable when the dataset is large. It is necessary to improve the efficiency of the algorithm's application. If a data point is far away a centre, it is not necessary to calculate the exact distance between the point and the centre in order to know that the point should not be assigned to this centre. Most distance calculations in standard Nearest neighbor are redundant.

PORTER STEMMER ALGORITHM:

In 1980, Porter presented a simple algorithm for stemming English language words.

The algorithm has been widely adopted and extended so that it has become the standard approach to word conflation for information retrieval in a wide range of languages.

Natural language texts typically contain many different variants of a basic word. Morphological variants (e.g., COMPUTATIONAL, COMPUTER, COMPUTERS, COMPUTING etc.) are generally the most common, with other sources including valid alternative spellings, misspellings, and variants arising from transliteration and abbreviation. The effectiveness of searching, most obviously but not exclusively in terms of recall, would be expected to increase if it were possible to conflate (i.e., to bring together) the variants of a given word so that they could all be retrieved in response to a query that specified just a single variant.[2]

The steps of the algorithm are:

1) Step 1

This step is designed to deal with past participles and plurals. This is the most complex step and is separated into three parts in the original definition, 1a,1b and 1c. Step 1 also removes inflectional suffixes (i-suffixes).

a) Step 1a:

SSES → SS

caresses → caress

b) Step 1b:

(*v*) ING →

Opening → agree

c) Step 1c:

(*v*) Y → I

History → histori

2) Step 2

This step is much more straight forward. It deals with pattern matching on some common suffixes. It removes derivational suffixes (d-suffixes) and follows some rules such as follows:

(m>0) ATIONAL → ATE

Relational → relate

3) Step 3

It deals with special word endings. It also removes derivational suffixes (d-suffixes). Composite d-suffixes are reduced to single d- suffixes one at a time. Therefore, if a word ends with -icational, Step 2 reduces it to -icate and Step 3 reduces it to -ic. Below is the example of rules applied in Step 3.

(m>0) NESS →

Possibleness → possible

4) Step 4

It checked the stripped word against more suffixes in case the word is compounded. It deals with -ic,-able, -Ive and many

more which are similar strategy to step3. Example of rules involved in this step is as shown below:

(m>1) MENT →

Adjustment → adjust

5) Step 5

It tidy up the algorithm after removing suffixes in previous steps. It checks if the stripped word ends in a vowel and fixed appropriately. It consists of Step 5a and Step 5b as indicated in the example :

a) Step 5a

(m>1) E

Probate → probat

b) Step 5b

(m>1 and *d and *L) → single letter

Bill → bil

V. SYSTEM ARCHITECTURE

The focus area of this system is market research and analysis. It is a web-based application and aims at determining target markets and consumer density and identifying potential customers. We have used the concept of Cluster analysis for the same. This application will help determine the user's browsing details and monitor customer population. Web User analysis is a simple template that provides a graphical, time-phased overview of process in terms of conceptual design, mission, analysis, and definition phases.

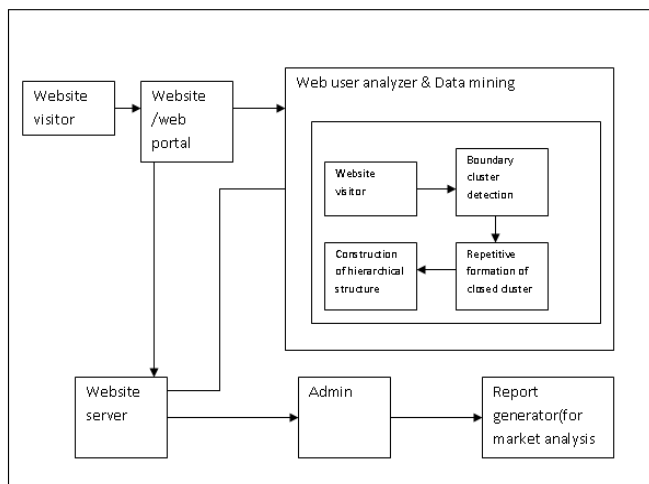


Fig. 1: System Architecture

It becomes highly challenged to digital firms due to the exponential growth of customer data in today's electronic business environment. Hence market segmentation proves to be one of the possible solutions. The goals of market segmentation include retention of customer, allocation of advertising resource, and an increase of profit margins. The outcome of market segmentation plays important role in market development, advertising, product pricing, attracting new customers and other marketing strategies. A good analytical tool should be able to compare the characteristics of different segments (group) and identify important attributes of each segment (group) to create opportunity for targeting customer which can be accomplished by getting the complete context of the customer. With this system interested to target the customer of my own web portal in major cities of Maharashtra. To find that from which location of the world the visitors of the website to trace the future market at that place. Maximum used services by this visitors also referral of the website and whole control of the system at administrator. The similar kind of facility available with Google Analytic Tools but for every domain we have to pay the different fees also the whole database with Google. Report formation is not as per our requirement which is the major task in every system.

VI. RESULTS

When the visitor visits the website, with the help of API details of the visitor such as longitude, latitude, IP, country,

city, date, time, and zip code are stored in data set. And according to the algorithmic strategies report is generated. This report consist of:

- i) Product visitors
- ii) Homepage visitors
- iii) Website visitors
- iv) Reviews of all products
- v) Charts- Bar graph, Pie chart

VII. CONCLUSION

Today, as internet has become a necessity with all the works being carried out easily through it. In addition to it, this project will surely contribute in making it easier for analyzing the products and their reviews for marketers and in near future for users as well.

REFERENCES

- [1]. Haibo Wang, Da Huo, Jun Huang, "An Approach for Improving K-Means Algorithm on Market Segmentation" IEEE Transactions on knowledge and data Engineering, Vol. 17, NO. 2, February 2012.
- [2]. Chien-Liang Liu, Wen-Hsiaio, Chia-Hoang Lee, Gen-Chi Lu and Emery Jou, "Movie Review Summarization in Mobile Environment" IEEE vol 42, No 3, May 2012.
- [3]. Bin Zhang Member, IEEE and Sargur N. Srihari, Fello, , "Fast k-Nearest Neighbor Classification using Cluster Based Trees" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No 4, April 2004.
- [4]. Shlomo Geva and Joaquin Site, "Adaptive Nearest Neighbor Pattern Classification" IEEE Transactions on Neural Networks, Vol 2, No 2, March 1991.
- [5]. Sridhar. A, Sowndarya. S, "Efficiency of K-Means Clustering Algorithm in Mining Outliers from Large Data Sets" (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 3043-3045.
- [6]. Rama.B, Jayashree.P, Salim Jiwani, "A Survey on clustering Current status and challenging issues" (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 2976-2980.