ISSN : 2350-0743

www.ijramr.com

International Journal of Recent Advances in Multidisciplinary Research
Vol. 03, Issue 05, pp.1444-1447, May, 2016

REVIEW ARTICLE

SEGMENTING CUSTOMER FOR MOBILE PLANS

Tanjila Tamboli, Priyanshi Shah, Lokesh Saboo and *Saurabh Singh

Department of Computer, Maharashtra Institute of Technology, Pune-411038, India

ARTICLE INFO

Article History:
Received 08th February, 2016
Received in revised form
13th March, 2016
Accepted 25th April, 2016
Published online 30th May, 2016

Keywords:

Segmenting, Data-plans,
Analyze market, Data mining.

ABSTRACT

The telecom industry is a highly competitive market, with frequent innovations required to sustain the business. The current business situation indicates a mismatch between user expectations and tariff plans available in the market. Furthermore, most service providers suffer from a high churn rate. The client (service provider) would therefore like to analyze the market and offer more customized plans that help to increase the retention rate. So the business objective is to develop a model to pro le the customers based on their usage patterns, the activities they indulge in order to design a service plans that will cater to their areas of interest. To realize the business objective, there is a need to apply unsupervised data mining techniques on given data to bring out clusters of people with similar data-usage attributes and mobile preferences. The clusters, soidenti ed can then be targeted by service providers with customized data plans and contracts. Thus this help in choosing the best recommendation plans to increase the pro t of the service provider and even the customers.

INTRODUCTION

Competition in the mobile telecommunications industry is becoming ercer. Mobile operator's pro ts and Average Revenue per User are facing tremendous challenges. Customers demand becomes diversi ed, differentiation and requirements of service quality become more rational and strict. In order to improve mobile operator's competitiveness and customer value, several data mining technologies can be used. One of most important data mining technologies is customer clustering analysis. The aim of clustering is to categorize prospective customers into distinct groups for distinctive contact strategies and proximal offerings. This targeting practice has been proven manageable and effective for mobile telecommunications industry. Most telecommunications carriers cluster their mobile customers by billing system data, (Anshul Arora and Dr. Rajan Vohra, 2014; Qining Lin and Yan Wan, 2009). Billing system data describe customer subscribe, spend and payment behavior. Call detail records describe customer utilization behavior. They have more information to describe customer behavior than billing system data. Therefore, this project discusses how to cluster mobile customers based on their called tail records and analyze their consumer behaviors. Competition in the mobile telecommunications industry is becoming more intensi ed. By using Call Data Record of every customer our project is segmenting customers into various clusters and recommends the best plan for that customer.

CLUSTERING

The objective of cluster analysis is the organization of objects into groups, according to similarities among them.

Corresponding author: Saurabh Singh
Department of Computer, Maharashtra Institute of Technology, Pune-411038, India.

Clustering can be considered the most important unsupervised learning method. As every other unsupervised method, it does not useprior class identi ers to detect the underlying structure in a collection of data. A cluster can be de ned as a collection of objects which are similar between them and dissimilar to the objects belonging too ther clusters. Figure showsthiswith asimple graphical example. In this case the clusters into which the data can be divided were easily identi ed. The similarity criterion that was used inthiscaseisdistance: two or more objects belong to the same clusterif they are close according to a given distance (in this case geometrical distance). This is called distance-basedclustering. Another way of clustering is conceptual clustering. Within this method, two or more objects belong to the same cluster if this one de nes a concept common to all that objects. In other words, objects are grouped according to their ttodescriptiveconcepts, not according to simple similarity measures. Some clustering algorithms are as follows.

K-means Algorithm

Use K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assumekclusters) xedapriori. The main idea is to de ne k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible farawayfromeachother. Then extstepis to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the rst step is completed and a nearly group ageis done. Atthispointweneedto recalculateknewcentroidsasa centre of the clusters resulting from the previous step. After we have these k new centroids, a

new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally,this lgorithmaimsatminimizinganobjectivefunctionknowassquareder ror function given by(Anshul Arora and Dr. Rajan Vohra, 2014; Fadly Humka *et al*., 2013; Qining Lin and Yan Wan, 2009; Qining Lin, 2007; Jansen, 2007; Jiawei Han *et al*., 2012)

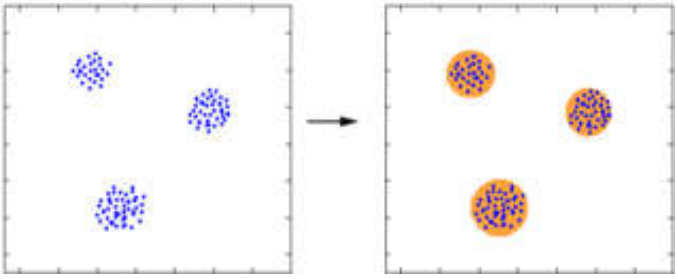

Fig. 1. Example of Cluster Formation

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\|x_i - v_j\| \right)^2$$

where,

'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j
'c_i' is the number of data points in i^{th}cluster.
'c' is the number of cluster centers.

Disadvantage of k-means is that it accept number of clusters from user. But many times at start we don't know the number of cluster for that some modification his done in k-means, so that it takes the number of clusters dynamically. Flow diagram for k-means is shown in Fig 2. In this algorithm we find inter-cluster and intra-cluster distance as-

$$Inter = \min(m_k \quad m_{kk}) \qquad \forall \quad k = (1,2 \ldots k \quad 1)$$

$$and \;\; kk = k + 1, \ldots k$$

$$Intra = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (x_i \quad x_m)^2}$$

K-medoids Algorithm

K-medoid is a classical partitioning technique of clustering the data set of n objects intokclustersfromaknownpriori. A useful tool for determining kisthesilhouette. It is more robust to noise and outliers as compared to k-means because it minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances. The most common realization of k-medoid clustering is the Partitioning around Medoids (PAM) algorithm and is as follows: (Jansen, 2007; Jiawei Han *et al*., 2012)

1. Initialize: randomly select(citation needed) (without replacement) k of the n data points as the medoids.
2. Associate each data point to the closest medoid.
3. While the cost of the con guration decreases: For each medoid m, for each non-medoiddatapoint o: Swapmando, recomputed thecost(sumofdistances of points to their medoid)
4. If the total cost of the con guration increased in the previous step, undo the swap
5. Else STOP

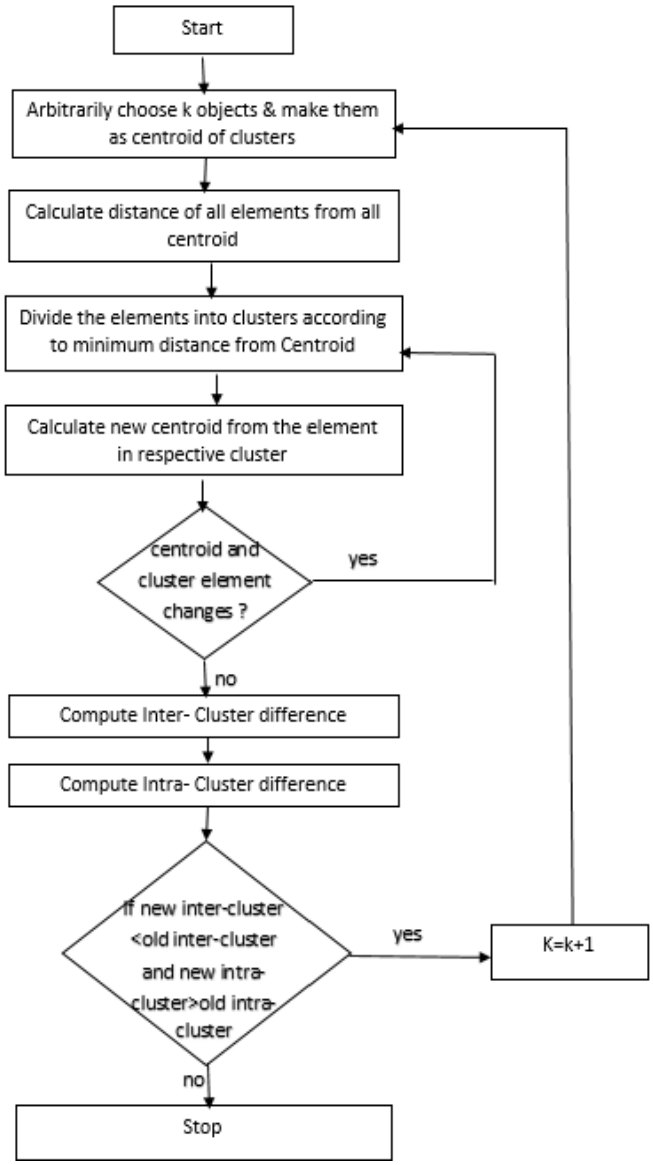

Fig. 2. K-means Flow Diagram

HierarchicalClustering

In data mining and statistics, hierarchical clustering (also called hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types: (Jiawei Han *et al*., 2012)

Agglomerative: This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

- **Divisive:** This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

In general, the merges and splits are determined in a greedy manner. The results of hierarchical clustering are usually presented in a dendrogram. Inthegeneralcase, the complexity of agglomerative clustering is $O(n^3)$,which makes them too slow for large data sets. Divisive clustering with an exhaustive search is $O(2^n)$ (Jiawei Han *et al*., 2012)

USER SEGMENTATION

Segmentation is a way to have more targeted communication with the customers. The process of segmentation describes the characteristics of the customer groups (called segments or clusters) within the data. Segmenting means putting the population in to segments according to their af nity or similar characteristics. Customer segmentation is apreparations tep for classifying each customer according to the customer groups that have been de ned. The segmentation is based on the users usage patterns and their various attributes like call duration, data usage, data plans, night calls, group SIM, SMS, MMS etc. Segmentation is essential to cope with today's dynamically fragmenting consumer marketplace. By using segmentation, marketers are more effective in channeling resources and discovering opportunities. The construction of user segments is not an easy task. (Anshul Arora and Dr. Rajan Vohra, 2014; Qining Lin and Yan Wan, 2009)

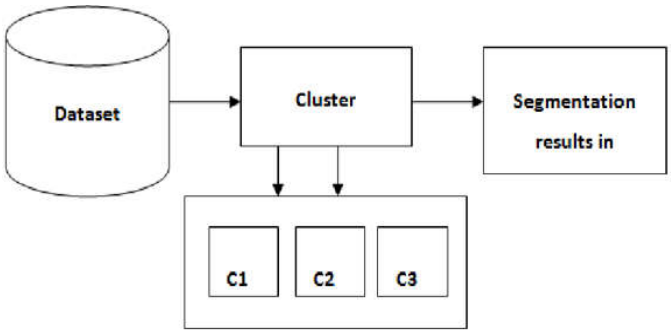

Fig. 3. Clustering and Segmentation

The term data mining was used. Data mining is the process of searching and analyzing data in order to nd implicit, but potentially useful, information. It involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and ultimately comprehensible information, from large databases.Data mining uses a broad family of computational methods that include statistical analysis, decision trees, neural networks, rule induction and re nement, and graphic visualization. Although, data mining tools have been available for a long time, the advances in computer hardware and software, particularly exploratory tools like data visualization and neural networks, have made data mining more attractive and practical. The typical data mining process consist of the following steps:

- Problem formulation
- Data preparation
- Model building
- Interpretation and evaluation of the results

Pattern extraction is an important component of any data mining activity and it deals with relationships between subsets of data. Formally, a pattern is de ned as: A statements in L that describes relationships among a subsets of facts F_s of a given set of facts F, with some certainty C, such that S is simpler than the enumeration of all facts in Fs. Data mining tasks are used to extract patterns from large data sets. The various data mining tasks can be broadly divided into six categories as summarized. The taxonomy re ects the emerging role of data visualization. Ataxonomyof dataminingtasksaseparatedatamining task, even as it is used to support other data mining tasks. Validation of the results is also a data mining task. By the fact that the validation supports the other data mining tasksand is always necessary within a project, this task was not mentioned as a separate one. Different data mining tasks are grouped intocategories depending on the type of knowledge extracted by the tasks. The identi cation of patterns in a large data set is the rst step to gaining useful marketing insights and marking critical marketing decisions. The data mining tasks generate an assortment of customer and market knowledge which form the core of knowledge management process. The speci c tasks to be used in this project are Clustering (for the customer segmentation), Classi cation (for estimating the segment) and Data visualization. Clustering algorithms produce classes that maximize similarity within clusters but minimize similarity betweenclasses. A draw back of this method Is that the number of clusters has to be given in advance. The advantage of clustering is that expert knowledge is not required. For example, based on user behavior data, clustering algorithms can classify the customers intocallonly users, International callers, SMS only users etc. Classi cation algorithms groups customers in prede ned classes. For example, the service provider can classify its customers based on their age, gender and type of subscription and then target its user behavior. Data visualization allow data miners to view complex patterns in their customer data as visual objects complete in three or two dimensions and colors. In some cases it is needed to reduce high dimensional data into three or two dimensions. To provide varying levels of details of observed patterns, data miners use applications that provide advanced manipulation capabilities to slice, rotate or zoom the objects.

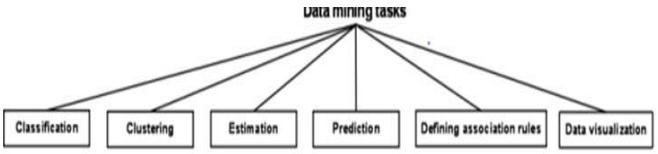

Fig. 3. Data Mining tasks

BESTPLANRECOMMENDATION

With the help of clustering the system recommends the best plan to customer on the basis of his current requirement. The recommender system compares the generated plan with the customers current plan and if the cost of current plan is more than that of generated plan then the message is send to the customer else it is not recommended. The attributes of the customer matches with the attributes of static plans stored in database. The plan with which these attributes match is generated else a new plan is generated dynamically. (Anshul Arora and Dr. Rajan Vohra, 2014; Fadly Humka *et al*., 2013)

OUTCOMES

The mobile telecommunication marketplace is highly competitive. In order to serve more customers and improve performance and marketing result, telecom operator needs different marketing strategies on the bases of their customer behavior. To support this, our paper gives a system for segmenting customer for mobile data plans, which improves performance of recommending plans to customers. (Anshul Arora and Dr. Rajan Vohra, 2014) To know the in uence of the

feature values on the outcome, a complete data analysis is required. Segmentation of mobile customers of different groups will be done based on some rules. Customers are segregated into groups under the categorization of network providers. We cancalculate the score with the help of different attributes like Local Calls, STD Calls, Local SMS, Roaming, STD SMS, Tariff Plan, Data Plan, etc. Revenue maximization pro ling, identifying maximal cluster using clustering algorithm like k-means is also done. (Fadly Humka *et al*., 2013) Mobile Network Users are also bene ted from this project as the number of promotional calls and messages from the service providers are reduced and also the plans they are genuinely interested in are suggested and also these plans are better than their current plans. Also it help mobile operator to know more about their customers and their requirements and accordingly suggests a proper plan to their customers.

APPLICATIONS

In this project the market segmentation can help companies to focus their business and marketing strategy by selecting a potential target market in offering their plans and services. Furthermore, combination between two or more dimensions may provide market researchers a ne tuned market segment that represents the preferences of using the plans. (Fadly Humka *et al*., 2013) Furthermore, interms of data collection methods, the automatic data collection in this project gives alternatives for service provider in order to segment their users based on their actual usage and behavior without having the hassle on dealing with huge data of CDR or contacting the user through interview or questionnaire. There is a win-win situation for both the customer and the service provider. The customer gets the best and the cheapest plan, moreover the service provider can keep a hold towards many clients and can earn pro t. Recommender system recommends best plan for user according to its need and current cost of the plan.

Conclusion

Clustering analysis based on call detail records can give more information than other clustering analysis for marketing managements. It helps to achieve better distinguishable groups, so that marketing managements can design more suitable marketing strategy. It performs automatic customer segmentation based on usage behaviour, without the direct intervention of a human specialist. It focused on profiling customers and ending a relation between the profile and the segments. A short characterization of each cluster was made. It recommends the best plan for the customer and is bene ted to both the service provider and the customer. It is pro table and generates a plan that user is interested in.

FUTURE WORK

To know the in uence of the feature values on the outcome of the clustering, a complete data analysis project is required. To validate the resulting clusters human expert, and his feed-back offer improving the clustering criteria.To improve the classi cation is to come to a more accurate and precise de nition of the customer pro les.An enhanced and more precise analysis of the data ware house will lead to improved features.It is challenging to classify the pro le of the customer based on the corresponding segment alone.However, this is a complex course and it essentially requires the availability of high-quality features.

Acknowledgment

It gives us great pleasure in presenting the preliminary project report on 'Segmenting Customers for Mobile Data Plans'. We would like to take this opportunity to thank my internal guide Prof. L.B. Bhagwat for giving me all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to Prof. V.Y. Kulkarni, Head of Computer Engineering Department, M.I.T. and Pune for her indispensable support, suggestions. In the end our special thanks to Mr. Vivek Singh for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for Our Project.

REFERENCES

Anshul Arora and Dr. Rajan Vohra, 2014. Segmentation of Mobile Customers for Improving Profitability Using Data Mining Techniques, *International Journal of Computer Science and Information Technologies.*

Fadly Humka, Harry Bouwman, Mark de Reuver and Maarten Kroesen, 2013. Mobile Customer Segmentation based on smartphone measurement, Elsevier, Telematics and Informatics.

Jansen, S.M.H. 2007. Customer Segmentation and Customer Pro ling for a Mobile Telecommunications Company Based on Usage Behavior.

Jiawei Han, Micheline Kamber and Jian Pei, 2012. Data Mining concepts and technique, Elsevier, Book.

Qining Lin and Yan Wan, 2009. Mobile Customer Clustering Based On Call Detail Records For Marketing Campaigns, IEEE Management and Service Science, MASS '09. International Conference on, 2009

Qining Lin, Mobile Customer Clustering Analysis Based on Call Detail Records, Communications of the IIMA, 2007

****