ISSN : 2231-0711

Enhanced Customer Relationship Management Using Fuzzy Clustering

Gayathri . A, Mohanavalli . S

Department of Information Technology,Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, Chennai, India, 603 110

***Abstract*— Data mining is the process of extracting patterns from data. Data mining is becoming an increasingly important tool to transform this data into knowledgeable information. Clustering is a popular mining strategy that separates those data into subsets called clusters. The two main approaches of clustering is soft clustering and hard clustering. This project aims to implement soft clustering and hard clustering to enhance CRM (Customer Relationship Management). CRM is the process of managing a good relationship with customer and improve their profitability of their interactions with the customer. Data mining is an efficiently used tool in CRM. The telecom industry specifically handles the large data being generated by applying various data mining techniques. This project primarily focuses on applying apt clustering algorithm to identify the soft partitions of the customers namely the fuzzy c means (FCM) clustering algorithm to determine the churn ratio accurately.**

Keywords-Telecom; Data mining; Customer Relationship Management; Clustering; K-Means; Fuzzy C Means

I. INTRODUCTION

The telecom industry is rapidly growing, in terms of the market dynamicity and competition. In turn, it creates new technologies and products, which open a series of options and offers to customers in order to satisfy their needs and requirements. However, one crucial problem that telecom companies in particular suffer from is a loss of valuable customers. A customer who leaves a carrier in favour of competitor costs a carrier more than if it gained a new customer. Therefore, retaining the customers can be seen as one of the most challenging problems that the telecom companies face in general. To tackle the problem one needs to understand the behaviour of customers and classify them according to their usage levels, so that the necessary decisions will be taken to enhance in providing services. Data mining model in order to efficiently understand the system behaviour and allow to make the right decisions[8].

Data mining is the process of extracting patterns from data. It seems to be an increasingly important tool for modern business to transform data into business intelligence giving an informational advantage. It is currently used in a wide range of profiling practices, such as marketing, surveillance, fraud detection, and scientific discovery. Data mining can help telecom industries in discovering meaningful trends, patterns and correlations in their customer, product, or data, to drive improved customer relationships.

Customer Relationship Management (CRM) is very important for any organization through which the companies wanted to know the relationship between customers and their organization.CRM basically begins with the deep analysis of customer behaviour. The telecom company needs to create personal interaction with the customer that they must be gather all the customer related data and using this data they will achieve the customer satisfactions[4,6]. Data mining is the strategy through which a company can collect all the information about the customer. The systematic application of data mining techniques reinforces the knowledge management process and allows marketing personnel to know their customers using their services.

The basic data mining techniques include association rules, classification, clustering, regression analysis, sequence analysis, etc. Classification and clustering are two main techniques in the customer segmentation. Cluster analysis is a statistical technique that is used to identify both minimize within group variation and maximize between group variation based on a distance or dissimilarity function, and its aim to find an optimal set of clusters. The Clustering technique groups telecom customers based on only the information found in the customer using the services provided by the telecom company that describes the telecom customers and their relationships with the company [6].

A large number of clustering algorithms have been developed for different type of applications. To cluster massive amount of data in telecom companies, we proposed k-means clustering and fuzzy c means (fcm) clustering algorithms.

Though telecom companies customer data sets are very larger and are in terabyte (TB) levels, to handle those data in order to find results and predict customer, k-means and fuzzy clustering techniques are the efficient one. In this paper, we can analyze how far both the algorithms implemented in the large volume of real time telecom data sets. The results of both the algorithm, shows the nature of customers. Hence the services which are used in highest level, lowest level and middle level can be found out and according to the analysis of outcomes, categorization can be made among customers and then CRM can be enhanced efficiently in telecom industry.

II. BACKGROUND WORK

In this project our ultimate goal is to retaining the customers in the telecom company. Customer retention can be achieved only when the details of customers can be understood more clearly and accurately. The telecom company stores the value of customer in a large volume of data bases. So it is more challenging process for a telecom company to retain their customer by having millions and millions of customer and their details. The telecom company was one of the first to adopt data mining technology. This is most likely because telecommunication companies routinely generate and store enormous amount of customer information such as different number of customers using different number

of services. Those collections of data sets are in very extensive and huge. By using data mining technique, Telecom Company can identify customers, retain customers and maximize their profit obtained from each customer. Huge amount of data is stored in the database and ultimately we will have to execute and process this huge data. Data mining is the solution to extract knowledge from these data base and which is evaluated [1].

Data mining is gradually increasing and important process for any business. The main focus of data mining is divided into classification, estimation, clustering, association and prediction. Clustering is generally used for market segmentation. Clustering separates a heterogeneous population into number of more homogeneous subgroups or clusters so that data in each cluster share some common trait. So it is more efficient for using this clustering technique in Telecom Company to group their customers according to the services used by them. In paper [9], to analyse the customer value they use RFM (Recency, Frequency, Money) rule. By using this RFM rule they can confirm the characteristics of customer. Yi-Hui-Lung defined customer value is fundamental to CRM. It is a in starting point of CRM to understand and measure the true value of a customer.

There are two main categories of segmentation methods using data mining technologies. They are multivariate statistics and neural network model. k-means clustering algorithm comes under multivariate statistics model. In paper [2], the author finds the value of a customer using k-means clustering algorithm based on RFM analysis and assessing the value of existing customer. So by using k-means algorithm, author can able to classify the customers according to their similarity level.

The accurate estimation of a customer usage level rate for a particular service can be easily found using k-means clustering. This hard clustering algorithm takes "n" number of customers and group them into "k" service group, in which each service contains the customers who is using that service in a highest level. In "k" clusters (services), each observation belongs to the cluster with the nearest mean.

In crisp clustering, each data point in the sample space assigned to only one cluster. We can able to assign each and every customer to a particular service based, in which a particular service is well used by a set of customers rather than other services. The k in the k-means algorithm stands for the number of cluster seeds initially provided to the algorithm. The goal of this algorithm is the assignment of each customer into one and only service. k-means is based on the principle which minimizes the clustering property indicator. A commonly used clustering criteria function is the minimized variance of all the same points (data or objects) to the central point in the cluster. k-means is a prototype-based, simple partition clustering technique which attempt to find an user specified k number of clusters. These clusters represented by their centroids, which is typically the mean of the points in the cluster.

Hard clustering technique identifies each customer to a certain service category, with the nature of "either-or", so this type of clustering is obvious. But in the estimation of customers in telecom company is unpredictable. The usage level may vary day by day and their usage pattern is an intermediary, with the nature of "and also", so in these sample uncertain situations, soft clustering called fuzzy clustering can be used efficiently and easily. Fuzzy Clustering provides a robust and resilient method of classifying collections of large amount of (data points) customers by allowing the same customer to reside in multiple service groups (clusters) with different degrees of membership. In fuzzy clustering, a data point can have a membership in multiple clusters (each of different degree).Membership is a measure of how strongly a data point is a member of a particular cluster [7].

Goal of cluster analysis is the recognition and quantification of the groupings. The quantification involves 2 processes (identifying membership of the data point and locating the center of cluster). Fuzzy clustering can able to handle extreme outliers, either very small or a higher than warranted membership degree in surrounding cluster. It provides a flexible and robust method of assigning data points into clusters. Each data point will have a degree of membership for each clusters in the range [0, 1] indicating the strength of its placement in that cluster [5].

III. PROPOSED SYSTEM DESIGN

3.1 K Means Clustering:

K-means clustering is the crisp clustering technique which attempts to cluster data by grouping related attributes in uniquely defined clusters. Each data point in the dataset is assigned to only one cluster. In partitioning the data, only the centers of the clusters are moved and condition of all the data points are fixed. Clustering is an iterative process of finding better and better cluster centers. Distance metric measures calculate how far away a point is from a cluster center. Cluster centers are randomly initialized and data point (x_i) is assigned into clusters (c_j ,j=1 to k). When all the data points are assigned to clusters, new cluster centers (centroids) are calculated. The process of calculating cluster memberships and recalculating cluster centers continues until the cluster centers no longer change from one cycle to the next. Thus, the cluster centers were stabled on the final.

The procedure of K-means clustering is partition into three steps.

- First, k initial centroids are selected randomly, where k is the initial number of clusters.
- Second, assigning every point in the data to the closest centroid and each collection of points assigned to a centroid form a cluster.
- Third, update the centroid of each cluster based on the new cluster. This process is repeated until no point changes clusters. The whole procedure is repeated until no point switches cluster assignment or a number of iterations are performed.

Table 3.1.1: Basic Parameters in k-means algorithm:

x_i → A vector of training data k → The number of fuzzy clusters c_j→ the centre of a crisp cluster .This value is repeatedly calculated by the algorithm. s_i→ A cluster in the sample space $\delta_{sj}()$ →characteristic function of set s_j

Table 3.1.2: Formulae in k-means:

(i)Cluster in the sample space:

$$s_j^{(t)} = \{X_i \mid \|X_i - C_j^{(t)}\| \le \|X_i - C_h^{(t)}\|\}$$

(ii)To Find Centroid:

$$c_j^{(t+1)} = \frac{\sum_{i=1}^{n} X_i \, \delta_{s_j}^{(t)}(x_i)}{\sum_{i=1}^{n} \delta_{s_j}^{(t)}(x_i)}$$

Table 3.1.3: Algorithm for k-means:

- Initialize k = Number of clusters
- Initialize c_j (cluster centre)
- Set cycle variable t=1
- Repeat

For i=1 to n : Distribute sample points (x_i) into the K clusters

For j=1 to k : Resolve $s_j^{(t)}$ for x_i

For j=1 to k : compare new cluster centres $c_j^{(t+1)}$

t=t+1

- Until c_i estimates stabilize.

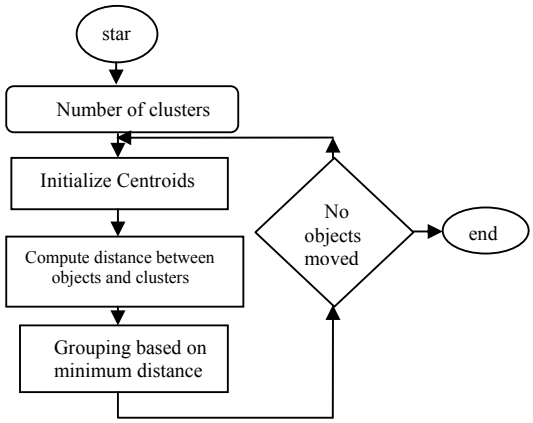


Figure 3.1.1: Flow diagram of k-means algorithm

3.2 Fuzzy C Means Clustering:

Introduced by Zadeh in 1969 to overcome the idea that all things can be absolutely True or False. According to fuzzy algebra every element of the universe can belong to any fuzzy set with a degree of membership that varies from 0 to 1 taking real values. Fuzzy clustering is an approach operating towards fuzzy logic and it provides the flexible method of assigns the data points to the clusters. Data points are given partial degree of membership in multiple nearby clusters. Central point in the fuzzy clustering is always no unique partitioning of the data in a collection of clusters. In this membership value is assign to the each cluster. Sometimes this membership has been used to decide whether the data points belong to the cluster or not.

The most well known fuzzy clustering algorithm is fuzzy-c-means, a modification of an original crisp cluster methodology. The idea of fuzzification parameter (m) in the range [1,n] , which determines the degree of fuzziness on the cluster. When m > 1 the degree of fuzziness among points in the decision space increases, when m=1, the clouds do not overlap, but as you increase the value (m=1.25) the clouds begin to overlap and share many of the same points.

The procedure of fuzzy c means algorithm undergoes the following steps:

Step 1: the membership value for each sample point using the cluster centers.

Step 2: the cluster centers using all membership values. When the cluster centers stabilize the clustering algorithm is finished.

Table 3.2.1: Basic Parameters in FCM Algorithm

X → vector of training data

d_{ij} → distance of the i-th data point form the j-th cluster center.

P → no of fuzzy clusters

M → A fuzzification parameter in the range [>0,<1]

C_i → the center of a fuzzy cluster.

$\mu_i(x_i)$ → A fuzzy membership qualification indicating the membership of sample x_i to the j-th cluster.

Table 3.2.2: Formulae in fuzzy c means

(i)Membership function:

$$\mu_j(x_i) = \frac{\left(\frac{1}{d_{ij}}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{p}\left(\frac{1}{d_{ik}}\right)^{\frac{1}{m-1}}}$$

(ii) To Find Centroid:

$$C_j = \frac{\sum_i (\mu_j(x_i))^m x_i}{\sum_i (\mu_j(x_i))^m}$$

Table 2.6: Algorithm for fuzzy c means:

Initialize p=number of clusters

m=fuzzification parameters

Initialize c_j

Repeat:

For i=1 to n: update $\mu_j(x_i)$

For j=1 to p : update c_j

Until c_j estimates stabilize.

TWO LOOPS:

First loop : cal membership value

Second loop: recalculate the cluster centers.

When the cluster centers stabilize, they fail to change

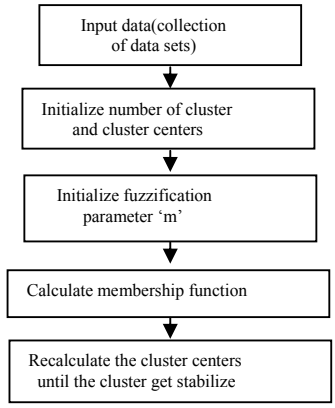


Figure 3.2.1: Flow Diagram for Fuzzy C Means Algorithm

3.3 Fuzzy clustering in Telecom Company:

Fuzzy clustering methods can be used to mine Telco customer and prospect databases to gain residential and business customer market share. The Enhancements to make fuzzy clustering effective in traditional databases follows four key steps.

They are as follows:

- First, customers often have significant membership values in more than one distinct fuzzy cluster and can be considered in a natural manner for hybrid or multiple contacts in a given marketing campaign.
- Second fuzzy clustering outcomes are shown to be dependent on the particular offer or marketing message.
- Third, there are differences in clustering outcomes over time as various offers and treatments are successively presented to consumers, and as products and tastes change. This evolution of fuzzy clusters can be used to help understand customer loyalty and to extract more optimal lifetime economic relationship value.
- Fourth, in the longer run, formal procedures can be suggested involving intuitive fuzzy-based clustering metrics for continuous process improvement, to support increasingly flexible and opportunistic campaign management.

IV. IMPLEMENTATION AND RESULTS

We propose a research framework for customer retention in telecom industry. We collected real time customer datasets from Telecom Company. Two datasets, one consisting of 99 customers with 77 (services) attributes, and other one with 1170 customers with 77 attributes were used. Among these 77 attributes, 20 attributes were filtered using a feature extraction algorithm and the attributes used are listed as follows:

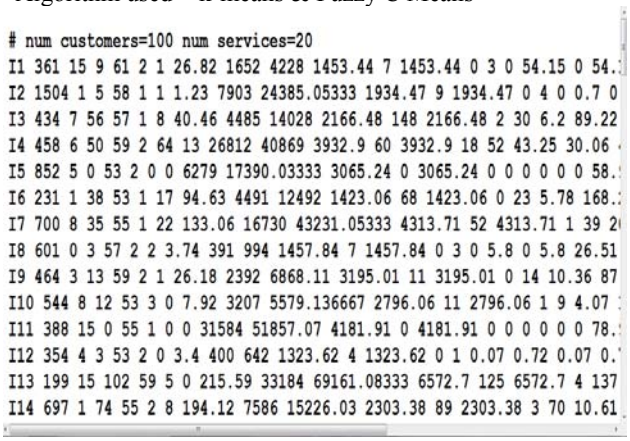
1)Eqpdays,
2)lor,
3)complete_Range,
4)months,
5)uniqsubs,
6)drop_blk_Range,
7)mou_peav_Range,
8)totcalls,
9)totmou,
10)attempt_Range,
11)adjrev,
12)iwylis_vce_Range,
13)peak_vce_Range,
14)mouowylisv_Range,
15)mou_opkv_Range,
16)mouiwylisv_Range,
17)opk_vce_Range,
18)avgrev,
19)totrev,
20)models

The selected datasets (services) of 100 as well as 1170 customers were tried with both the hard clustering (k-means) and soft clustering (Fuzzy c means) algorithms in order to find the churn rate of customers in the telecom company accurately.

The Implemented results are shown as follows:

Input given:

Number of Customers = 99 , Number of services = 20
Number of Clusters = 2,
Algorithm used = k-means & Fuzzy C Means



```
# num customers=100 num services=20
I1 361 15 9 61 2 1 26.82 1652 4228 1453.44 7 1453.44 0 3 0 54.15 0 54.
I2 1504 1 5 58 1 1 1.23 7903 24385.05333 1934.47 9 1934.47 0 4 0 0.7 0
I3 434 7 56 57 1 8 40.46 4485 14028 2166.48 148 2166.48 2 30 6.2 89.22
I4 458 6 50 59 2 64 13 26812 40869 3932.9 60 3932.9 18 52 43.25 30.06
I5 852 5 0 53 2 0 0 6279 17390.03333 3065.24 0 3065.24 0 0 0 0 0 0 58.
I6 231 1 38 53 1 17 94.63 4491 12492 1423.06 68 1423.06 0 23 5.78 168.
I7 700 8 35 55 1 22 133.06 16730 43231.05333 4313.71 52 4313.71 1 39 2
I8 601 0 3 57 2 2 3.74 391 994 1457.84 7 1457.84 0 3 0 5.8 0 5.8 26.51
I9 464 3 13 59 2 1 26.18 2392 6868.11 3195.01 11 3195.01 0 14 10.36 87
I10 544 8 12 53 3 0 7.92 3207 5579.136667 2796.06 11 2796.06 1 9 4.07
I11 388 15 0 55 1 0 0 31584 51857.07 4181.91 0 4181.91 0 0 0 0 0 0 78.
I12 354 4 3 53 2 0 3.4 400 642 1323.62 4 1323.62 0 1 0.07 0.72 0.07 0.
I13 199 15 102 59 5 0 215.59 33184 69161.08333 6572.7 125 6572.7 4 137
I14 697 1 74 55 2 8 194.12 7586 15226.03 2303.38 89 2303.38 3 70 10.61
```

Result:

From comparison of the clustering techniques, customer churn was observed to be different as per the clustering result. FCM groups the customer based on the membership value. The membership value gives the accurate churn ratio of customers moving from one group to other group. The result of soft clustering is more accurate than hard clustering because the membership value in FCM clustering groups the customer who is having largest membership value in a particular group accordingly.

So in our experimental analysis, we clustered the customers in two groups (clusters) based on Euclidean distance in kmeans, we grouped the customers in which Group A consists of 77 number and Group B having 22 number of customers. But in Fuzzy c means, the group A consists of 76 number of customers and group B consists of 23 number of customers based on their membership value. So FCM algorithm gives a better result which is more accurate than kmeans algorithm.

Similarly in this way, we tested both the algorithms for 1170 customers and selected 20 attributes. The results for this data set also give an accurate result in soft clustering algorithm than hard clustering. The membership value of FCM makes to move 15 customers from one group A to other group B gives the churn ratio of every customer accurately.

Conclusion & Future Work

It is observed that FCM gives better support to CRM rather than kmeans and other hard clustering techniques, and it is further proposed to extend this work to large real time data sets.

V. REFERENCES

[1] A.S.Al-Mudimigh, F.Saleem, Z.Ullah, F.N.Al- Aboud, "*Implementation of Data Mining Engine on CRM – Improve Customer Satisfaction"*, IEEE 2009.

[2] Binu Thomas, Raju G, "*A Novel Fuzzy Clustering Method for outlier Detection in Data Mining*,International Journal of Recent Trends in Engineering", Vol.1,No.2,May 2009.

[3] Kaiqi Zou, Zhiping Wang, Ming Hu, "*An new initialization method for fuzzy c- means algorithm",,* Springer.

[4] Levine S, *"The rise of CRM",* America's Network, Vol. 104No.6, p.34,2000.

[5] Robert L. Cannon, Jitendra V. Dave, James C. Bezdek, "*Efficient Implementation of the Fuzzy c-Means Clustering Algorithms,* IEEE Transactions on Pattern Analysis and Machine Intelligence", Vol PAMI-8, No.2.March 1986.

[6] Ruey-Shun Chen, Ruey-chyi Wu and J.Y.Chen, "*Data Mining Application in Customer Relationship Management of Credit Card Business",* proceedings of the 29th Annual International Computer Software and Applications Conference, IEEE 2005.

[7] Shokoi Z SELIM, M.A.Ismail, "*Kmean type Algorithms: A Generalized convergence Theorem and characterization of Local optimality"*, IEEE Transactions on Pattern Analysisand Machine Intelligence, Vol.6(1984), pp.81-87.

[8] Tarik Rashid, "*Classification of churn and non-churn customers for Telecom companies"* ,International Journal of Biometrics and Bioinformatics(IJBB), Volume 3, Issue 5.

[9] Yi-Hui-Liang, "*Combining the k-means and decision tree methods to promote customer value for the Automotive Maintenance Industry", IEEE 2009.*

[10] Earl cox,*"Fuzzy modeling and genetic algorithms for Data Mining and Exploration*", Chapter 7,page No 207 to 260.