

Customer Segmentation by Various Clustering Approaches and Building an Effective Hybrid Learning System on Churn Prediction Dataset

E. Sivasankar and J. Vijaya

Abstract Success of every organization or firm depends on Customer Preservation (CP) and Customer Correlation Management (CCM). These are the two parameters determining the rate at which the customers decide to subscribe with the same organization. Thus higher service quality reduces the chance of customer churn. It involves various attributes to be analyzed and predicted in industries like telecommunication, banking, and financial institutions. Customer churn forecast helps the organization to retain the valuable customers and it avoids failure of the particular organization in a competitive market. Single classifier does not result in higher churn forecast accuracy. Nowadays, both unsupervised and supervised techniques are being combined to get better classification accuracy. Also unsupervised classification plays a major role in hybrid learning techniques. Hence, this work focuses on various unsupervised learning techniques which are comparatively studied using algorithms like Fuzzy C-Means (FCM), Possibilistic Fuzzy C-Means (PFCM), K-Means clustering (K-Means), where similar type of customers is grouped within a cluster and better customer segmentation is predicted. The clusters are divided for training and testing by Holdout method, in which training is carried out by decision tree and testing is done by the model generated. The results of the churn prediction data set experiment show that; K-Means clustering algorithm along with the decision tree helps improving the result of churn prediction problem present in the telecommunication industry.

Keywords CRM • Churn • K-Means • FCM • PFCM • Decision tree

E. Sivasankar (✉) • J. Vijaya
Department of CSE, NIT, Trichy 620015, Tamilnadu, India
e-mail: sivasankar@nitt.edu

J. Vijaya
e-mail: vijayacsdept@gmail.com

1 Introduction

Emergence of information and communication systems has made remarkable growth in every industry. One such industry is telecommunication where numerous competitors have been evolved. In such a competitive market, customer retention plays a key role for the successful running of the organization, as retaining an obtainable customer incurs more expenditure than adding a fresh customer to the organization. Hence, customer correlation management (CCM) in mobile communication services focus on reducing customer churn. Analysis in international telecommunication market says that for certain number of years the customers churn has reached more than 70% [1]. Hence there are various factors like customer's residence, type of service, cost of service, need, and utilization of the services involved in determining customers churn. As a result, forecast of customers churn is very difficult for any organization. Also such an intelligent forecast helps to retain the precious customers which will increase the profit of the organization. Hence various Intelligent Mining Techniques for instance Logistic Regression, k-Nearest Neighbor, Decision Tree, Naive Bayes, Artificial Neural Networks, Support Vector Machine, Inductive Rule Learning, etc., were used to forecast customers churn behavior [2–7]. Even though classification involving single classifiers act as a good model, their forecast accuracy is not appreciable. Hence, currently there are hybrid techniques involving both supervised and unsupervised techniques. But there is only few contribution that exists towards unsupervised learning. In hybrid models, unsupervised learning techniques play a vital role for predicting better results, because similar type of customers are grouped with in cluster and analyzed for better customer segmentation. Hence, this work focuses on various unsupervised learning techniques which are comparatively studied using algorithms like Fuzzy C-Means (FCM), Possibilistic Fuzzy C-Means (PFCM), K-Means clustering (K-Means), where similar type of customers are grouped within a cluster and better customer segmentation is predicted. The clusters are divided for training and testing by holdout method, in which training is carried out by decision tree and testing is done by the model generated. The results of the churn prediction data set experiment shows that; K-Means clustering algorithm along with the decision tree helps improving the result of churn prediction problem present in the telecommunication industry.

2 Literature Survey

Shin-Yung, David C, and Hsiu-Yu (2006) designed two hybrid classification models which are based on three existing methods such as K-Means, decision tree, and back propagation neural network. In the first classification model, K-Means clustering in addition to decision tree was combined as well as in the second model, decision tree and back propagation neural networks were combined. Finally, they evaluated the above models using Hit ratio and LIFT [8]. Lee and Lee (2006)

proposed SePI (Segmentation by Performance Information), which comprises of three models core protocol, bias protocol, and support protocol. decision tree which is out performs in single level classification is deployed in core protocol. The bias protocol uses the result of the main protocol. Support protocol uses data which are incorrectly predicted by main protocol in which artificial neural network used. Hence the key idea of their work is that the data which are incorrectly predicted by main protocol and can correctly be predicted by the support protocol [9]. Indranil Bose and Xi Chen (2009) selected two set of seven attributes in which the first set is based on minute of usage and second is based on revenue contribution. The following five clustering techniques support vector machine, K-Means, birch, self-organizing map, K-Medoid are used and the result of the above clustering techniques are given as an input to the decision tree and performance is evaluated using top decile lift for test data [10]. Huang, Kechadi, and Buckley (2012) extracted new set of features which is evaluated using seven traditional classification methods. There accuracy is increased with the newly extracted features than existing ones. Performance was measured in terms of true positive and false positive rate [2]. Mehdi, Ali, Bijari (2012) implemented a mixture model by combining Artificial Neural Network and Multiple Linear Regression model. Their proposed Artificial Neural Network among Multiple Linear Regression gives better result compared to single Artificial Neural Network and traditional classification techniques like support vector machine, linear discriminant analysis, tree induction k-Nearest neighbor, and quadratic discriminant analysis [11]. Ying Huang, Tahar Kechadi (2013) proposed weighted K-Means algorithm founded on path analysis. It is used to cluster the data, then resemblance between test instance and k number of clusters are measured which is named as close instance. Customer churn was predicted based on close cluster which has maximum close instances. In case of close cluster having churn label and non-churn label, the test value is given to first-order inductive learning classification algorithm for prediction [12]. Yangming Zhang et al. (2006) proposed hybrid learning system which combines k-nearest neighbor and logistic regression and they evaluated using two metrics—one is based on accuracy and second on receiver operating characteristic curve. Yeshwanth et al. (2011) solved the problem of churn prediction based on decision tree combined with genetic algorithm [13]. Wouter Verbeke et al. (2010), they took churn prediction data set and created various set of sub-samples based on over-sampling, ALBA, and ALBA oversampled. Then each sub-sample was tested based on Ant Miner+, SVM, Majority rule, Logistic Regression, Ripper, and some hybridization algorithms like Ant Miner with Decision Tree [14]. Tsai et al. (2012) solved the problem of churn prediction based on Back propagation artificial neural network with itself (ANN-ANN) and Self-organizing map (SOM-ANN), they took three type of test data patterns. The Accuracy is increased with the new hybrid artificial neural network with existing single artificial neural network model [15].

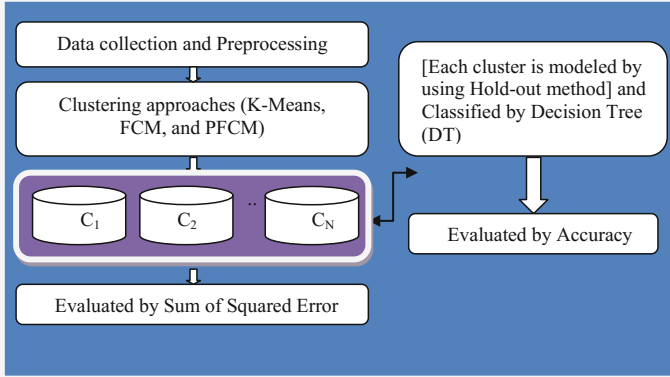


Fig. 1 Block diagram of hybrid churn prediction mode

3 Churn Prediction Models and Methods

The block diagram shown in Fig. 1 denotes the steps involved in the proposed work. The dataset is initially preprocessed and the missing values are eliminated based on mean value of attributes. After that string value features are converted into numerical format as clustering algorithm only supports features in numerical form. In order to evaluate their effectiveness to segment the customers, various clustering methods are employed. The clusters are divided for training and testing by holdout method, in which training is carried out by decision tree and testing is done by the model generated.

3.1 Dataset Preprocessing

In the preprocessing phase useless features (State, Phone number) are eliminated from dataset and reduced to 18. After that string value features are converted into numerical format as clustering algorithm only supports features in numerical form. Preprocessing is one of the most important phases as it reduces noise from data, thus making the data more consistent

3.2 Clustering Algorithm

Clustering technique has wide range of applications in data mining like image segmentation, market analysis, etc. The technique groups data objects based on their similarities. Minimum distance data points are present with in a cluster. Distances between two data points are calculated by various formulas like Euclidean distance, Manhattan distance, Minkowski distance, and so on [16]. Generally, clustering can be classified into fuzzy clustering, model-based clustering, partitional

clustering, density-based clustering, hierarchical clustering, and grid-based clustering. In this document, we are using partitional clustering and fuzzy clustering techniques.

3.2.1 K-Means

K-Means algorithm was proposed by Stuart Lloyd in 1957, based on partitional-based clustering techniques. It constructs K partition of N input tuples, where K less than or equal to N. Partitional-based clustering must satisfy the following requirement (i) Every cluster to have a minimum one sample (ii) Each sample necessity to present on precisely one cluster. Algorithm 1 illustrates the main step involved inside K-Means clustering.

Algorithm 1. Working Steps of K-Means Clustering

Input: Entire data D containing N samples.

Output: K number of cluster groups.

Step 1: Initialize K value, where K less than or equal to N

Step 2: Randomly choose K object as initial cluster center.

Step 3: Find distance measure from each object to chosen K cluster center.

Step 4: Similar object are grouped based on their minimum distance

Step 5: Update the cluster means and take as new cluster centers (C_i), $\{i = 1 \dots k\}$

Step 6: No change in the cluster center then the algorithm stop, or go to step 3.

3.2.2 Fuzzy C-Means

Fuzzy C-Mean's algorithm was projected with Dunn in 1973, based on Fuzzy techniques [17]. The main idea of Fuzzy C-Means algorithm includes membership function μ_{ij} . We have to find the K clusters membership matrix for each data point. For all data objects, the summation of K cluster membership value is equal to 1. The data object is assigned to highest membership value cluster. In fuzzy cluster method, single object will be present in two or more cluster and in some cases the cluster does not have any data points. Algorithm 2 illustrates the main steps involved in Fuzzy C-Means clustering.

Algorithm 2. Working Principles of Fuzzy C-Means Clustering

Input: Entire data D containing N samples.

Output: K number of cluster groups.

Step 1: Initialize K value, where K less than or equal to N

Step 2: Randomly choose k cluster center

Step 3: Calculate fuzzy membership μ_{ij} using Eq. 1 in Table 1

Step 4: Similar object are grouped based on their maximum membership value μ_{ij}

Step 5: Update new cluster center V_j using Eq. 2 in Table 1

Step 6: The algorithm stops when minimum objective function $J(U, V)$ is reached using Eq. 3 in Table 1

3.2.3 Possibilistic Fuzzy C-Means (PFCM)

Possibilistic Fuzzy C-Means algorithm was anticipated by Nikhil R et al. [18], based on (FCM) Fuzzy C-Means and (PCM) Possibilistic C-Means algorithm. The main idea of Possibilistic Fuzzy C-Means algorithm includes Typicality value t_{ij} and Membership function μ_{ij} . Algorithm 3 illustrates the main steps involved in Possibilistic Fuzzy C-Means clustering. Fuzzy C-Mean's algorithm only depends on Membership utility, so the output of the Fuzzy C-Means algorithm has lot of outliers. But Possibilistic Fuzzy C-Means depend on Typicality value and Membership function; it eliminates various problems present in FCM and PCM.

Algorithm 3. Possibilistic Fuzzy C-Means Clustering Steps

Input: Entire data D containing N samples.

Output: K number of cluster groups.

Step 1: Initialize K value, where K less than or equal to N

Step 2: Randomly choose k cluster center

Step 3: Calculate Fuzzy membership μ_{ij} using Eq. 1 in Table 1

Step 4: Calculate Typicality value t_{ij} using Eq. 4 in Table 1

Step 5: Similar object are grouped based on their maximum membership value μ_{ij}

Step 6: Update new cluster center V_j using Eq. 5 in Table 1

Step 7: The algorithm stops minimum objective function $J(U, T, V)$ is reached Eq. 6 in Table 1

3.3 Classification Algorithm

Classification technique is a process of assemblage related information which is already present in the form of class labels. Initially, we have to build two types of data grouping called training and testing. The classification protocol is created using training data part and tested using testing part.

3.3.1 Decision Tree

1. Decision tree is a classification algorithm based on information gain
2. A binary tree is constructed using the calculated gain of the attribute
3. After that decision rules are formed out using different paths from root to leaf
4. Based on decision rules, it can predict whether the customer is churn or not.

4 Experimental Setup and Result Analysis

4.1 Dataset

Experiments are performed on a benchmark churn prediction dataset. The complete dataset be made of 5000 customers, out of 5000 customers 707 customers are churning customers and 4293 customers are non-churning customers. Every customer is represented by 20 attributes and predicted churn variable. The attributes primarily consist of the subsequent information: Demographic profiles which explain basic information of customer, account information, call details includes historical call value based on time and plan, complaint information and predicted class label churn or not.

4.2 Evaluation Criteria

Clustering: Sum of Squared Error (SSE): Sum of Squared Error measure is used to find the quality of the clustering techniques. The main intention of clustering is to reduce the sum of squared error value. If SSE value is low, then the clusters points are grouped well, i.e., object inside a particular cluster are similar and items that are belonging to different cluster are dissimilar. SSE value is calculated by the following Eq. 7.

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} \|c_i - o_{ij}\|^2 \quad (7)$$

Here, $\|C_i - O_{ij}\|^2$ refers Euclidean distance between each data point (O_{ij}) within a cluster to cluster center (C_i), it will be calculated for all K cluster and finally summarized.

4.3 Experiment Setup

This work comprises three set of experiments. The clustering result vastly decides the classification accuracy of any hybrid-based classification technique. Therefore, the first set of experiment is conducted to estimate the performance of the clustering algorithm (K-Means, FCM, and PFCM) based on SSE Measure. The second sets of experiment are conducted to calculate the efficiency of the hybrid form and in third set of experiment well known existing models are compared.

4.4 Result and Conversation

4.4.1 Experiment 1

The first set of experiment focuses on various unsupervised techniques which are comparatively studied using algorithms (K-Means) K-Means clustering, (FCM) Fuzzy C-Means, (PFCM) Possibilistic Fuzzy C-Means where similar type of customers are grouped with in a cluster and better customer segmentation is predicted. Figure 2 shows that sum of squared error result produced by K-Means (K-Means), Fuzzy C-Means (FCM), Possibilistic Fuzzy C-Means (PFCM). In this figure, horizontal axis represent number of cluster and vertical axis represent sum of squared error.

From Fig. 2, we can infer that the result generated by K-Means algorithm has lower SSE compared to Fuzzy C-Means and Possibilistic Fuzzy C-Means algorithm. Figure 3a–c shows that 3D clustering result generated Fuzzy C-Means (FCM), Possibilistic Fuzzy C-Means (PFCM), K-Means clustering (K-Means) with the number of cluster $K = 2$.

After clustering, we have to find the number of customer in each cluster which is produced by K-Means, FCM, PFCM. Table 2 shows that sample result produced by various clustering algorithm with the number of cluster = 5 where the total number of customer is 5000.

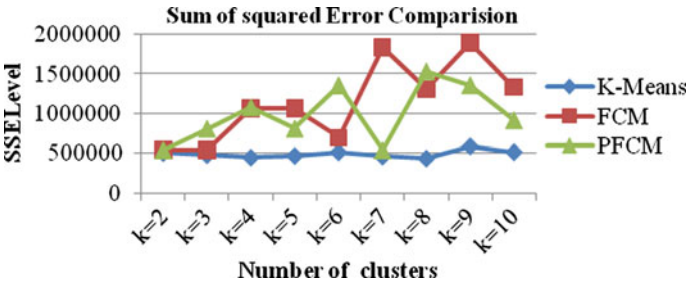


Fig. 2 SSE of different clustering techniques

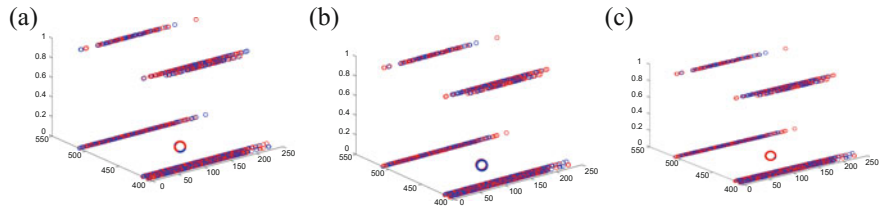


Fig. 3 a–c Clustering results of K-Means, FCM, PFCM where $K = 2$

Table 1 Formula’s used for FCM, PFCM algorithms

Equation no.	Formula name	Formula
(1)	Membership function formula both FCM, PFCM	$\mu_{ij} = \frac{1}{\sum_{k=1}^c \frac{d_{ij}^2}{d_{ik}^{\frac{2}{m-1}}}}$
(2)	Cluster center formula for FCM	$V_j = \frac{(\sum_{i=1}^n (\mu_{ij})^m x_i)}{(\sum_{i=1}^n (\mu_{ij})^m)}$
(3)	Objective function formula for FCM	$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m x_i - v_j ^2$
(4)	Typicality matrix value for PFCM	$t_{ik} = \frac{1}{1 + \frac{(b(D_{ik}))^2}{\delta_i} \frac{1}{(q-1)}}$
(5)	Cluster center formula for FCM	$v_i = \frac{\sum_{k=1}^n (a\mu_{ik}^m + b t_{ik}^n) x_k}{\sum_{k=1}^n (a\mu_{ik}^m + b t_{ik}^n)}$
(6)	Objective function formula for PFCM	$X = \sum_{i=1}^c \delta_i \sum_{k=1}^n (1 - t_{ik})^\eta$ $J(U, T, V, Z) = \sum_{i=1}^c \sum_{k=1}^n a\mu_{ik}^m + b t_{ik}^{m*} z_k - v_i ^2 + X$

Table 2 Number of customer in 5 clusters

Algorithm	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
K-Means	885	1152	850	1070	1043
FCM	97	156	2456	0	2291
PFCM	28	2478	2487	0	7

4.4.2 Experiment II

In second set of experiment, the accuracy of each hybrid model such as K-Means along with decision tree, fuzzy C-Means together by decision tree, and Possibilistic fuzzy C-Means together by decision tree are compared. The output of each model is plotted in Fig. 4. The parallel axis and the upright axis correspond to the rates of number of clusters and accuracy value predicted by each model. In this figure, we can observe that the accuracy value of K-Means along with decision tree is better than fuzzy clustering. Since K-means clustering has lowest SSE value, it can partition the customers well hence, it will produce better classification result.

4.4.3 Experiment III

In order to compare different hybrid models, the third set of experiments are evaluated. The result of the proposed hybrid model is compared against different

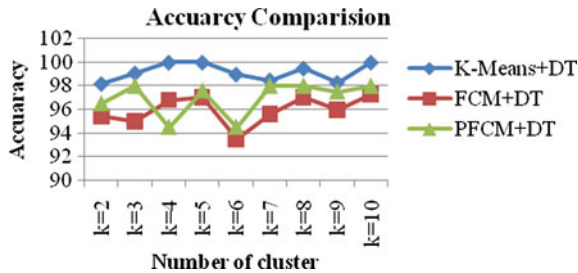


Fig. 4 Accuracy for different hybrid models

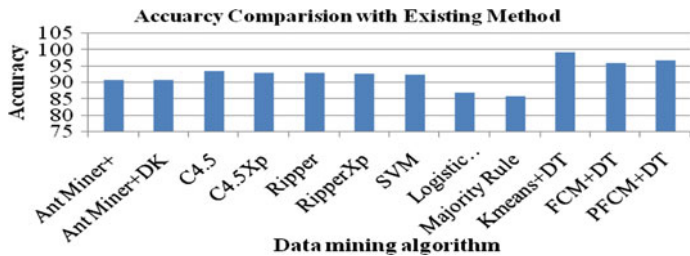


Fig. 5 Accuracy of existing model with proposed hybrid model

existing models produced by Wouter Verbeke et al. (2010) without sampling. Figure 5 illustrates that the proposed hybrid model outperform the classification techniques like SVM, Ripper, Logistic regression, Ant Miner combined with decision tree.

5 Conclusion and Future Work

Nowadays, both unsupervised and supervised techniques are combined to get better classification accuracy, and unsupervised classification plays a major role in hybrid learning techniques. Hence, this work focuses on various unsupervised techniques which are comparatively studied using algorithms like fuzzy C-Means (FCM), possibilistic fuzzy C-Means (PFCM), K-Means clustering (K-Means). The experimental results show that K-Means algorithm produces better cluster quality compared to fuzzy grouping so it produced better classification accuracy along with decision tree. In future, we concentrate Fuzzy C-Means clustering for hybrid model to increase the prediction of customer churn.

References

1. Rob Mattison: Telecom churns management: The golden opportunity. Fuquay-Varina, N.C: APDG Publishing.
2. Ying Huang, Bingquan Huang and M. T. Kechadi: A rule-based method for customer churns prediction in telecommunication services, *Lecture Notes in Computer Science*, 6634, 411–422, 2011.
3. Hyunseok Hwang, et al., An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry, *Journal of Expert Systems with Applications*, 26(2), 181–188, 2004.
4. Bart Lariviere, Dirk van den Poel: Predicting customer retention and profitability by using random forests and regression forests techniques, *Journal of Expert Systems with Applications*, 29(2), 472–484, 2005.
5. Chih-Ping Wei, I-Tang Chiu: Turning telecommunications call details to churn prediction: A data mining approach, *Journal of Expert Systems with Applications*, 23 (2), 103–112, 2002.
6. Guo en XIA, Wei dong Jin: Model of customer churn prediction on support vector machine, *Journal of Systems Engineering - Theory and Practice*, 28 (1), 71–77, 2008.
7. Bingquan Huang, et al., Customer churns prediction in telecommunications, *Journal of Expert Systems with Applications*, 39 (1), 1414–1425, 2012.
8. Shin-Yung Hung, et al., Applying data mining to telecom churn management, *Journal of Expert Systems with Applications*, 31 (3), 515–524, 2006.
9. Jae Sik Lee, Jin Chun Lee: Customer churns prediction by hybrid model, *Proceedings of the second international conference on advanced data mining and applications*, 4091, 959–966, 2006.
10. Indranil Bose, Xi Chen: Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn, *Journal of Organizational Computing and Electronic Commerce*, 19:2, 133–151, 2009.
11. Mehdi Khashei, Ali Zeinal Hamadani, Mehdi Bijari: Novel hybrid classification model of artificial neural networks and multiple linear regression models, *Journal of Expert Systems with Applications*, 39 (3), 2606–2620, 2012.
12. Ying Huang, Tahar Kechadi: An effective hybrid learning system for telecommunication churn prediction, *Journal of Expert System with Applications*, 40(14), 5635–5647, 2013.
13. V. Yeshwanth Raj, V. Vimal raj, M.Saravanan: Evolutionary churn prediction in mobile networks using hybrid learning, In *Proceedings of the twenty-fourth international Florida artificial intelligence research society conference, (FLAIRS)*, Palm Beach, Florida, USA, May 18–20. AAAI Press.
14. Wouter Verbeke, et al.: Building comprehensible customer churn prediction models with advanced rule induction techniques, *Journal of Expert System with Applications*, 38 (3) 2354–2364, 2011.
15. Chih-Fong Tsai, Yu-Hsin Lu: Customer churns prediction by hybrid neural networks, *Journal of Expert System with Applications*, 36 (10), 12547–12553, 2009.
16. Velmurugan. T: Performance Comparison between k-Means and Fuzzy C-Means Algorithms using Arbitrary Data Points, *Wulfenia Journal*, 19(8), 2012.
17. J. C. Dunn: A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well Separated Clusters, *Journal of Cybernetics*, 3(3), 32–57, 1973.
18. Nikhil R. Pal, et al., A Possibilistic Fuzzy c-Means Clustering Algorithm, *IEEE Transaction on Fuzzy Systems*, 13(4), 517–530, 2005.



本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。

图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：

[图书馆首页](#) [文献云下载](#) [图书馆入口](#) [外文数据库大全](#) [疑难文献辅助工具](#)