Telecommun Syst (2017) 66:603–614
DOI 10.1007/s11235-017-0310-7

A churn prediction model for prepaid customers in telecom using fuzzy classifiers

Muhammad Azeem[1] · Muhammad Usman[1] · A. C. M. Fong[2]

Published online: 29 March 2017

Abstract The incredible growth of telecom data and fierce competition among telecommunication operators for customer retention demand continues improvements, both strategically and analytically, in the current customer relationship management (CRM) systems. One of the key objectives of a typical CRM system is to classify and predict a group of potential churners form a large set of customers to devise profitable and targeted retention campaigns for keeping a long-term relationship with valued customers. For achieving the aforementioned objective, several churn prediction models have been proposed in the past for the accurate identification of the customers who are prone to churn. However, these previously proposed models suffer from a number of limitations which place strong barriers towards the direct applicability of such models for accurate prediction. Firstly, the feature selection methods adopted in majority of the past work neglected the information rich variables present in call details record for model development. Secondly, selection of important features was done through statistical methods only. Although statistical methods have been applied successfully in diverse domains, however, these methods alone without the augmentation of domain knowledge have the tendency to yield erroneous results. Thirdly, the previous models have been validated mainly with benchmark datasets which do not provide a true representation of real world telecom data consisting of noise and large number of missing values. Fourthly, the evaluation measures used in the past neglected the True Positive (TP) rate, which actually highlights the ability of a model to correctly classify the percentage of churners as compared to non-churners. Finally, the classifiers used in the previous models completely neglected the use of fuzzy classification methods which perform reasonably well for data sets with noise. In this paper, a fuzzy based churn prediction model has been proposed and validated using a real data from a telecom company in South Asia. A number of predominant classifiers namely, Neural Network, Linear regression, C4.5, SVM, AdaBoost, Gradient Boosting and Random Forest have been compared with fuzzy classifiers to highlight the superiority of fuzzy classifiers in predicting the accurate set of churners.

Keywords Churn prediction · Fuzzy classification · Feature selection · Telecommunication · K-nearest neighbor

1 Introduction

In telecommunication sector, customer retention has become a major area of concern for the CRM departments. Customer retention refers to the ability of a company to hold its existing customers [1]. It has been an agreed fact that acquisition cost is far higher as compared to the cost of retaining the customers [2]. Based on an earlier research in [3], the average churn rate for telecom operator is 2.2% per month. Furthermore, Asian telecom service providers deal with a much higher customer churn rate of 4.8% per month. Customer retention relies heavily on the accurate and timely prediction of churners.

Churn prediction, as an analytical part of CRM, deals with predictions of churners in order to devise the strate-

✉ Muhammad Usman
dr.usman@szabist-isb.edu.pk

Muhammad Azeem
mazeem10@gmail.com

A. C. M. Fong
alvis.fong@wmich.edu

1 Department of Computing, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan

2 Western Michigan University, Kalamazoo, USA

gic approaches towards customer retention. It is therefore; very critical to develop an accurate churn prediction model.

Churn prediction is generally referred to as a problem of binary classification where organizations typically apply data mining techniques for customer churn analysis [4]. In binary classification, companies thrive to predict the answer to the following important question from each of its customer. "*Is this customer going to leave us within the next X months*?" There are only two possible answers, yes or no, and it is what we call a binary classification task. Here, the input of the task is a customer and the output is the answer to the question (yes or no). In telecom domain, churner to non-churners ratio is very low, so class imbalance has to be handled by enhancing the instances of churner class [5]. Under sampling and Oversampling are the most promising techniques that have been utilized to balance both classes. The objective is to develop a most suitable churn prediction model which helps to identify those customers who are most likely to churn.

For achieving the aforementioned objective, several churn prediction models [4,6–15] have been proposed in the past for the accurate identification of the customers who are prone to churn. A number of machine learning methods, such as Neural Networks [13], Linear regression [4], Support Vector Machines [7], Decision Tree [14] and ensemble of hybrid methods [15] have been used in the previously proposed models to provide accurate prediction of churners. However, these prior models suffer from a number of limitations which make it difficult to apply such models for accurate prediction in a real world environment.

Firstly, the feature selection methods adopted in majority of the previous models neglected, with the exception of [14], the information rich variables present in call details record (CDRs) for model development. Instead, features from service logs [16], customer demographics [17], complaints data [4], bill and payment data [12], contractual data [14] and operation support system (OSS) data [18] has been used.

Moreover, the prior work focused on post-paid customers mainly. We argue that, modeling postpaid customers is far less challenging as compared to the prepaid clients. Prepaid customers do not pay the monthly bill and their service usage is also irregular which makes it really hard for the prediction models to accurately identify the potential churners.

Secondly, in the prior work, the selection of important features was done through statistical methods only [4]. Although statistical methods have been applied successfully in diverse domains, however, these methods alone without the augmentation of domain knowledge have the tendency to yield erroneous results and poor performance of predictive model.

Thirdly, the previous models have been validated mainly with benchmark datasets which don't provide a true representation of real world telecom data consisting of noise and large number of missing values. The presence of noise and missing values in the variables degrades the performance of the prediction models. There is a requirement of developing the models which effectively deal with the noisy data and could work well for a typical real world telecom data containing large number of missing values and noise.

Fourthly, the evaluation measures used in the past neglected the True Positive (TP) rate or *Sensitivity* measure, which actually is the proportion of positive examples (churners). It highlights the ability of a model to correctly classify the percentage of churners as compared to non-churners. Moreover, Area Under Curve (AUC) measure which has been widely utilized in the previous literature failed to give a much larger value than 0.5 which is considered to be a baseline for classifier's performance evaluation. A good classifiers should yield much larger value as compared to 0.5 [17] and it was not observed in the previously reported work.

Finally, the classifiers used in the previous models completely neglected the use of fuzzy classification methods which perform reasonably well for data sets with noise. To the best of our knowledge, fuzzy classifiers have never been applied to customer churn prediction in telecom domain. Moreover, their performance has never been evaluated in terms of TP rate and AUC.

To overcome the aforementioned limitations, a fuzzy based churn prediction model has been proposed in this paper. The proposed model utilizes domain knowledge and statistical measures for feature selection. Moreover, it takes into account the information rich variables present in CDRs of prepaid customers and highlights the effectiveness of predicting the prepaid customer churn using fuzzy classifiers [19,20].

In terms of validation, the model has been validated using a real data from a telecom company in South Asia. For reasons of confidentiality, country and company names are not disclosed. To evaluate the performance of the fuzzy classifiers experiments are conducted and the results in terms of TP rate and AUC have been compared with the predominant classifiers reported in the telecom churn prediction domain. Experimental results show that the proposed model with fuzzy classifiers is more effective and accurate as compared to the existing models for customer churn prediction in telecommunication sector.

The remainder of this paper is organized as follows. Section 2 presents an overview of the related work to highlight the rationale and significance of this research. Section 3 describes our methodology followed by Sect. 4 in which experimental results and findings are discussed. Section 5 concludes this paper and specifies directions for future research in this area.

2 Related work

Churn prediction is one of the most effective strategies used in telecom sector to retain existing customers. It consists of detecting which customers are likely to cancel a subscription to a service based on how they use the service. Focus of telecom sector has been shifted from acquiring new customer to retaining existing customers due to high cost associated with acquiring new customers [9].

Literature review of churn prediction in telecom domain revealed that a number of classification techniques have been used by different authors to enhance the accuracy of prediction model. The major limitation is that dataset used for prediction contains small number of instances and attributes however in actual large size datasets in telecom sector break down the performance of these classifiers.

Various data mining techniques have been compared in [11] using data set of a telecom service provider in Taiwan considering the variables such as customer demographics, contact/service status, billing information, call detail records and service change logs. To simplify the modeling framework, each subscriber has been assigned a 'propensity to churn' score regardless of profitability. Empirical results show that Neural Networks (NN) and decision trees (DT) techniques give accurate measures to churn prediction. Data preprocessing has been performed however class imbalance problem was not addressed while building churn prediction model.

Similarly churn prediction using Neural Networks (NN) was presented in [13] where a mixed process neural network has been proposed for time variant and traditional statistical data simultaneously. Further to simplify the structure, an optimized cMPNN has been proposed which uses orthogonal furrier based transformation technique for preprocessing.

Machine learning technique (support vector machine) for churn prediction has been presented in [7] where a comparison is made between two parameter selection techniques that are based on grid search and cross-validation. Support vector machine's predictive performance is benchmarked for random forest and logistic regression techniques. Data set contains variables such as socio-demographics information, subscription, un-subscription, auto-renewal details and client company information etc. Class imbalance problem is not handled and data preprocessing was also not performed.

To build a binary classifier for churn prediction, a hybrid approach was presented in [8] which is combination of logistic regression (LR) and K nearest neighbor (KNN). The classifier building process has been divided into two phases using hybrid KNN-LR. In phase 1, KNN is applied on each predictor variable and in phase 2 LR is trained with new predictor variable after phase 2. Methodology has been implemented and trained on 4 different data sets taken from UCI machine learning repository and all of these data sets have 2 classes. Data preprocessing and class imbalanced is not handled. Furthermore, overall accuracy is higher 97.67%, however precision and recall measures were not calculated in order to get prediction accuracy of churner class.

A churn prediction method has been proposed in [9] which compared with artificial neural network, decision tree, logistic regression, and naive Bayesian classifier. Strength of their work is that the method has the best accuracy rate, hit rate, covering rate, and lift coefficient, and therefore, provides an effective measurement for customer churn prediction. Limitation is that data set chosen for experiments contains small number of instances and variables however in telecom sector, data sets are of huge size and proposed method is not applicable in this domain.

Class imbalance problem has been addressed in [15], author has investigated the increase in performance of sampling, Usefulness of regression and decision trees approach to the problem of modelling churn in the prepaid sector of the cellular telecommunication company has been discussed in [5]. Linear models are more stable than decision trees that get old quickly and their performance weakens in time. Previous studies used small number of variables whereas large number of variables (all extracted from CDRs) has been used in this study however smaller number of instances has been used for training and testing.

Furthermore, preprocessing and class imbalance not handled in their work. Data set consists of the train sample – 85,274 observations, the calibration sample – 36,824 observations and the test sample 45,497 observations. Data in the train sample and the calibration sample come from the dataset collected at the same time, which was then split randomly into the train and validation part. The test sample was collected six months after the train and calibration sample.

Two most common modeling techniques (decision trees and multilayer perceptron neural networks) have been presented in [18] where one of the most promising approaches (support vector machines) are selected as predictors. Data preprocessing has been done however class imbalanced was not handled, also data set contains small number of instances and attributes.

According to [16] Improved Balance Random Forest (IBRF) produces better prediction results than other random forests algorithms such as balanced random forests. Author improved balanced random forests by combining balanced random forests and weighted random forests.

For churn prediction, ensemble classifier is another method which is used in [21,22] as modeling techniques. Ensemble classifier utilizes the benefits of multiple classifiers and enhances the overall accuracy of churn prediction model.

Data preprocessing has been done in [15] for optimal sub sampling, i.e. nominal data has been converted into numerical. Different feature reduction techniques such as Principle Component Analysis (PCA), Fisher's ratio, F-score and Minimum Redundancy and Maximum Relevance (mRMR) have been implemented to handle the imbalance data distribution, whereas Random Forest (RF) and K Nearest Neighbor

(KNN) classifiers are used to evaluate the performance on optimally sampled and reduced features dataset.

Predictive performance of Naïve Bayesian classification techniques has been studied in [4] where feature selection has been performed using Markov Blanket method. Results have been discussed in different perspectives, i.e. using different methods of feature selection and without feature selection. The reduction of features using Markov Blanket did not reduced performance. Study indicates that Augmented Naïve Bayes technique does not lead to a good network whereas General Bayesian Network results in simple and interpretable networks.

Finally classifiers performance has been evaluated using ROC and maximum profit criterion. Maximum AUC achieved was 0.88% on benchmark dataset. Other than AUC, maximum profit criterion has also been used for performance measure however; AUC was more discriminative than maximum profit criterion. AUC measures performance of complete dataset whereas maximum criterion deals with optimal fraction of customer therefore it is used in telecom for retention campaigns.

A very important problem of class imbalance has been addressed in [20] using various sampling techniques, 6 different samples have been taken and implemented the one-class and two-class classifications and results have been evaluated using AUC and F-measures. Random over sampling is not an effective way to handle class imbalance however oversampling using Synthetic Minority Over-Sampling Technique (SMOTE) in case of imbalance dataset has proven to be more effective.

Using two-class classification (Gaussian SVDD and Parzen-density estimation), AUC 96 was reported. It also has been observed that ensemble do not give significant improvement over individual classifiers. In ensemble classifiers, Stacking (SVM) gave better performance in terms of AUC.

From literature review, we have identified that several churn prediction models have been proposed in the past for the accurate identification of the customers. However, these proposed models suffer from a number of limitations which place barriers towards the implementation of such models for accurate prediction. Moreover, the fuzzy classifiers were never utilized in telecom churn prediction domain. Keeping in view the ability of fuzzy classifiers dealing with noisy data, we believe there is a strong need of building model based on fuzzy classifiers.

3 Methodology

From literature review, we have identified that none of the existing techniques have been targeted to enhance TP rate and AUC along with handling class imbalance on real world large dataset. We have proposed a framework to enhance TP rate and AUC of large size dataset. As shown in Fig. 1, in first step, data preprocessing has been performed which covers dealing with missing values and removing the noisy data, feature extraction and normalization.

Important variables have been extracted on the basis of domain knowledge. Churner ratio is too low i.e. 9% as compared to non-churners, so class imbalance has been handled by enhancing the instances of churner class in final dataset. As dataset size is huge, in next step data set has been divided into 2 parts, one for training and other for testing, ratio

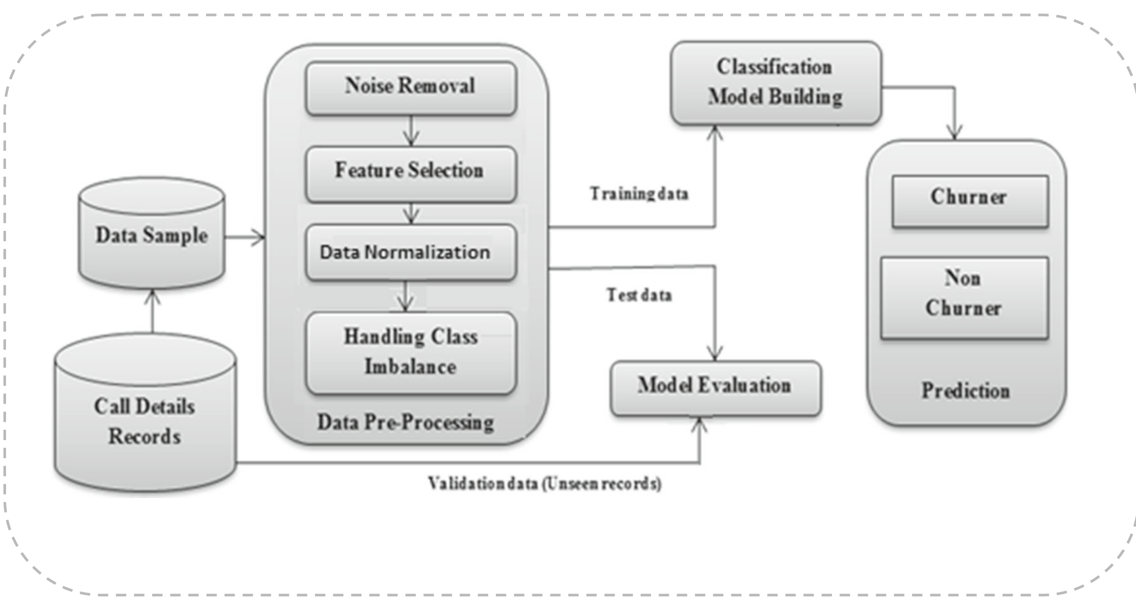

Fig. 1 Proposed model for accurate customer churn prediction

of training and testing has been kept consistent throughout implementation i.e. 80:20. Various classification techniques have been applied both on balanced and imbalanced class, performance of precision and recall has been evaluated through AUC curves in performance measure phase.

3.1 Data set

Dataset has been taken from a telecom company operating in South Asia. Dataset contains 600,000 instances with 722 attributes all extracted from customer service usage pattern like usage details of voice SMS and data.

Dataset contains 3 month usage information of Dec 2014, Jan 2015 and Feb 2015 with class value either "Yes" or "No" for churner and non-churners respectively. Attributes have numeric values which have been extracted from customer call detail records based on usage detail of Voice calls, SMS count and data session length. Furthermore, data set contains loan, promotion messages Opt in and revenue related information.

3.2 Data preprocessing

3.2.1 Noise removal

It is very important to remove noisy data as it can lead to erroneous results. In telecom data set, there are a lot of missing, garbage and inappropriate values like "Null", spaces or special characters. In this step noisy values have been removed to some extent, however, as we intend to use fuzzy classifiers for model building therefore thorough cleaning of data was not required at the pre-processing stage.

3.2.2 Feature selection

Feature selection is one of the most important and critical step of implementation where only important and relevant features have been selected from large size dataset based on the domain knowledge. A range of techniques have been used in literature [12,16] to get the relevant attributes for churn prediction. In telecom domain, different researchers have selected various attributes most of them are related to call details, customer demographics, customer bill information and complaints history. Table 1 represents the churn indicators along with their respective features.

In general following are the main indicators which help in identifying the potential churners before they actually decide to churn. In our case, important features have been selected based on few churn indicators extracted from domain knowledge which have been explained in following sections:

Decline in weekly spend rate: A decrease in a subscriber's spend rate not only results in a decrease in revenue but is also an indication of churn. Timely identification of such an event enables a Telco to offer & cross-sell bundles according to a subscriber's need. This helps in retaining a customer, increasing his loyalty, and enhancing revenue.

Decrease in outbound call count: A number of reasons can lead to a decrease in outbound call count, including lost/broken handset or charger, customer travelling abroad, network congestion, customer unhappy or customer has started using data services.

This is also one of the indications of churn. For a service provider this provides an opportunity to contact the subscriber and addresses his needs, before customer is contacted and exposed to a competitor's offer.

Increase in days on zero balance: An increase in the days on zero balance is an indicator of a sleeper situation as a subscriber is unable to perform any revenue generating activity unless he recharges. This can be one of the indicators of customer churn and addressing such subscribers timely can increase customer loyalty and retention.

Balance burn out: The burn rate of a subscriber is an indication of the days left before a subscriber uses up all of his balance, provided the subscriber does not recharge. This indication is important in order to remind a customer to recharge his account.

Increase in calls to one competitor: A subscriber who frequently calls to off-net numbers is more likely to churn. If the number of calls by a subscriber to a particular competitor increases, then he is vulnerable to accept offers from the competitor due to influence from his friends. Therefore, it is important to timely contact such subscribers before they are exposed to competitors' offer to not only prevent them from defecting to the competitors but also to use their influence to convert their off-net friends.

Decrease in residual balance: Residual balance is the amount left in the account before recharge. A decrease in residual balance is an indication that a customer is late in recharging his account, which can also an indication to churn. This is an indication important to remind a customer to recharge his account.

Shift of calling circle to off-net: If the number of off-net friends for one competitor increases significantly, then the subscribers is vulnerable to accept offers from the competitor due to an increase in influence from his friends. Therefore, it is important to timely contact such subscribers before they are exposed to competitors' offer to not only prevent them from defecting to the competitors but also to use their influence to convert their off-net friends. One of the important reason is that customer gets charged extra due because of interconnect charges, to avoid these charges there is a high chance that customer will decide to switch network.

Decrease in outbound international calls: A decrease in the number of outbound international calls is an indication of churn, especially if a competitor has an active campaign/package for international calls. Timely identification of such subscribers provides an opportunity to cross-sell international calling bundles to take the activity to the previous level.

Decrease in spend rate by day of the week: This event is different from Point 1 as this looks at daily spend rate that may vary by day of the week.

Increase in outbound voice inactive days: Increase in the inactivity of a subscriber, is an indication that a subscriber is entering into the sleeper situation and churn. This event is more important for high revenue generating subscribers.

Increase in calls to one competitor: This event is an indication that a subscriber is late in recharging his account. This is an indication for a reminding a customer to recharge his account.

3.2.3 Data normalization

As CDR dataset contains only numeric fields, so data normalization has been performed on complete dataset where all numeric values have been transformed into a range from 0–1 using min-max data normalization technique. Its a simple normalization technique in which we fit the data, in a pre-defined boundary, or to be more specific, a pre-defined interval i.e. [C, D].
Formula to calculate the normalized value is given in Eq 1

$$B = \left(\frac{(A - Min(A))}{Max(A) - Min(A)}\right) * (D - C) + C \tag{1}$$

where B contains the min–max normalized data if pre-defined boundary is C, D and A is the original data which gets normalized using above formula.

3.2.4 Handling class imbalance

As churner ratio is too low i.e. 9% as compared to non-churners, so in our case class imbalance has been handled by enhancing the instances of churner class in final dataset. Oversampling has been performed to balance both classes. In final dataset churner to non-churner ratio is 60, 40 respectively.

3.3 Classification model building

In order to build model, we have used different classifiers in our methodology. In recent years, fuzzy classifiers are getting attention for classification in telecom churn prediction due to nature of data. In churn prediction domain, data is very noisy and fuzzy classifiers especially Vaguely Quantified Nearest Neighbors (VQNN) effectively handles such data. In next sub section, short description of only fuzzy classifiers with promising results is given.

3.3.1 Fuzzy NN

In order to classify an object based on likeness and similarity with another one, the Fuzzy K nearest neighbor (FNN) algorithm is introduced. Let's suppose an object y belongs to class C which can be written as:

$$C'(y) = \sum_{x \in N} R(x, y)\, C(x) \tag{2}$$

where N is the y's K nearest neighbors, and R(x, y) is the [0, 1]—similarity of x and y. It can be written as following.

$$R(x, y) = \frac{||y - x|| - 2/(m - 1)}{\sum_{j \in N} ||y - j|| - 2/(m - 1)} \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm, and 'm' is a parameter which controls the weighting of the similarity, in our case 'm' is set to the default value 2. Assuming crisp classes, Algorithm 1 shows an application of the FNN algorithm that classifies a test object 'y' to the class with the highest resulting membership.

The idea behind this algorithm is that the degree of closeness of neighbors should influence the impact that their class membership has on deriving the class membership for the test object.
Algorithm 1: The fuzzy KNN algorithm

FNN(U,C,y,K).
U, the training data; C, the set of decision classes;
y, the object to be classified; K, the number of nearest neighbors.

(1) N ← getNearestNeighbours (y, K);
(2) ∀C ∈ C
(3) C(y) = P $\sum_{x \in N}$ R(x, y) C(x)
(4) output arg max C ∈ C (C(y))

3.3.2 Fuzzy-rough nearest neighbors (FRNN)

This algorithm combines the approach of FNN algorithm and fuzzy rough approximations. From the FNN approach

Table 1 Relevant features against churn indicator

Churn indicators	Associated churn variables
Decline in spend rate	DAYS SINCE LAST RECHARGE, VOICE BUCKET REVENUE, ACTIVE DAYS SINCE LAST CALL, SMS BUCKET REVENUE
Decrease in outbound call count	TOTAL REVENUE VOICE, REVENUE ON NET, REVENUE OFF NET
Increase in days on zero balance	ACTIVE DAYS SINCE RECHARGE
Balance burn out or revenue	TOTAL REVENUE, SMS CHARGED OUTGOING COUNT, GPRS BUCKET REVENUE, REVENUE SMS, CRBT REVENUE
Increase in calls to one competitor	REVENUE OFF NET, CHARGED OFF NET, MINUTE OF USE OFF-NET
Decrease in residual balance	BALANCE AVERAGE DAILY, BALANCE LAST RECHARGE
Shift of calling circle to off net	OFF NET OUTGOING MINUTE OF USE, CHARGED OFF NET MINUTE OF USE
Only free resources utilization	FREE MINUTE OF USE, FREE ON NET MINUTE OF USE, FREE SMS
Increase in outbound voice inactive days	REVENUE FIX, ACTIVE DAYS RECHARGE
Decrease in recharge frequency	RECHARGE COUNT, RECHARGE VALUE, LAST RECHARGE VALUE
Decrease in voice calls frequency	ACT DAYS MINUTE OF CALL
Number of promotion opt ins	PROMO OPT IN
Loan information	LOAN COUNT, ACTIVE DAYS SINCE LOAN
Increase in inactive days	INACTIVE DAYS CALLS, INACTIVE DAYS SMS, INACTIVE DAYS DATA

get nearest neighbors and from the fuzzy rough approximations get fuzzy upper and lower approximations of decision classes. For example, consider a set of objects called U. One of the objects is considered test object called t and the remaining are training objects. Establish a fuzzy relation between a test object and each of the training objects. Calculate the similarity value of each couple that varies from 0 to 1. Choose the training objects with highest value of similarity as the nearest neighbors. Determine the upper and lower approximation of each class by means of the nearest neighboring objects. Predict the class membership of test object by using upper and lower approximations. Output the decision class with the resulting best combined fuzzy lower and upper approximation memberships.

Algorithm 2: Fuzzy-rough nearest neighbors (FRNN).

Let $D =$ set of decision classes, $U =$ training data, and $t =$ test object to be classified output class:

(1) N $\leftarrow$ getNearestNeighbors(t,K)
(2) $\tau \leftarrow 0$, Class $\leftarrow \emptyset$
(3) $\forall \mathrm{C} \in \mathrm{D}$
(4) if $(((\mathrm{R} \downarrow \mathrm{C})(\mathrm{t}) + (\mathrm{R} \uparrow \mathrm{C})(\mathrm{t}))/2 \geq \tau)$
(5) Class $\leftarrow$ C
(6) $\tau \leftarrow ((\mathrm{R} \downarrow \mathrm{C})(\mathrm{t}) + (\mathrm{R} \uparrow \mathrm{C})(\mathrm{t}))/2$
(7) Output class

There are two instance algorithms of FRNN named FRNN-FRS and FRNN-VQRS. Both differ in their approximations. FRNN-FRS uses traditional approximations of all and at least one; on the other hand FRNN-VQRS uses VQRS approximations of most and some. Consider a class, so high value of upper approximation reflects that all or most of the neighboring objects belong to class; similarly high value of lower approximation reflects that at least one or some of neighboring objects belong to class for FRS and VQRS approximations, respectively.

3.3.3 *Vaguely quantified nearest neighbors (VQNN)*

This algorithm is a variant of FRNN (Fuzzy Rough Nearest Neighbor) algorithm. According to the algorithm a test object is classified using VQRS (vaguely quantified rough set) approximations. In contrast to traditional approach, VQRS approach uses “most” and “some” quantifiers for upper and lower approximations, instead of “all” and “at least one,” respectively. So impact of noise will be less on VQRS approximations compared to traditional approximations. For example, consider a set of objects ‘U’ containing a test object

't' to be classified and training objects. A subset of training objects 'K' is considered to be the nearest neighboring objects. Classify test object to a class on basis of upper and lower approximations of a class in these nearest neighbor objects. If we consider a class C, high value of upper approximation reflects that most of the neighboring objects belong to class C; similarly high value of lower approximation reflects that some of neighboring objects belong to class C.

3.3.4 OWANN

Fuzzy-rough ownership identifies the homogenous regions for each class in test data. It also handles the fuzzy and rough uncertainties caused by insufficient knowledge i.e. attributes in training data. For any object y, the fuzzy-rough ownership function τC of class C can be represented as.

$$\tau C(y) = \frac{\sum_{x \in X} R(x, y) C(x)}{|x|} \tag{4}$$

In this, the fuzzy relation R is determined by:

$$R(x, y) = \exp\left(-\sum_{a \in A} Ka\,(a(y) - a(x))^{2/(m-1)}\right) \tag{5}$$

where m controls the weighting of the similarity (as in FNN) and Ka is a parameter that decides the bandwidth of the membership, defined as

$$K_a = \frac{|X|}{2 \sum_{x \in X} |a(y) - a(x)|^{2/(m-1)}} \tag{6}$$

τC (y) is interpreted as the confidence with which 'y' can be classified to class 'C'. The corresponding crisp classification algorithm, called FRNN-O. Initially, the parameter K_a is calculated for each attribute and all memberships of decision classes for test object 'y' are set to 0. Next, the weighted distance of 'y' from all objects in the universe is computed and used to update the class memberships of 'y'. Finally, when all training objects have been considered, the algorithm outputs the class with highest membership.

By contrast to the FNN algorithm, the fuzzy-rough ownership function considers all training objects rather than a limited set of neighbors, and hence no decision is required as to the number of neighbors to consider. The reasoning behind this is that very distant training objects will not influence the outcome (as opposed to the case of FNN).

3.4 Model evaluation

After a classification model is available, the model will be used to predict the further behavior of customers. As one of the important step to ensure the model generalize well, the performance of the predictive churn model has to be evaluated.

Confusion matrix is a standard table used to measure the accuracy of both churner and non-churner classes. Precision and recall are the measures that actually evaluate the accuracy of churn prediction model. Precision is basically ratio of relevant records (TP) identified to the total number of irrelevant and relevant records. It is usually expressed as a percentage. In our example precision will be calculated as TP/(TP + TN).

Similarly, Recall is the ratio of the number of relevant records (TP) retrieved to the total number of relevant records in the dataset. It is also expressed in percentage as Recall i.e. TP/(TP + FN).

A churn prediction system should be measured by its ability to identify maximum churners for retention campaigns, and we therefore use the True Positive rate and Receiver Operating Characteristic (ROC) curve to give a comprehensive evaluation of our prediction model where value of area under curve (AUC) measures the effective of churn prediction model.

AUC is obtained by plotting the precision and recall on a graph. So, we use AUC instead of accuracy because percentage of correctly identifying the customer is not sensitive to a single cut-off value, hence recall rate and AUC are more adequate measures to evaluate churn prediction model.

The lift curve is also very popular technique in order to retain the most likely to churn customers by assigning each customer a "propensity to churn" score. The lift curve helps us determine how effectively we can skip customers with lowest churn score by selecting a relatively small number of customers and retaining a large portion of the customers.

The input required to construct a lift curve is a validation dataset that has been "scored" by appending to each case the estimated probability that it will belong to a given class. The lift will vary with the number of cases we choose to act on. A good classifier will give us a high lift when we act on only a few cases.

4 Experimental setup and results

We have used machine Core i5 with 8GB RAM and all experiments have been conducted using WEKA software. Dataset contains 3 months of call details records used for training and 4th month data has been used for testing and predicting potential churners. We have selected the same attributes for each month and our proposed model has been validated using different classifiers. Eleven classifiers have been implemented on sample training dataset. Classification techniques include Naive Bayes, Multilayer Perceptron, Linear Regression, C4.5, SVM, Decision Trees, Bagging, AdaBoost, Gradient Boosting, Random Forest and Fuzzy classifiers.

Training and testing ratio for each dataset has been consistent through implementation and that is 80:20 respectively. Furthermore, ratio of non-churner to churner is consistent in testing datasets, however class imbalance was handled only for training dataset. Testing dataset has been kept separately and was not provided during training of model. Below are the details of implementation and results, dataset contains same number of attributes i.e. 84 and 600,000 instances.

4.1 Result 1: Balanced versus imbalanced class

Table 2 represents the distribution of training and testing for complete dataset with imbalance and balanced class accuracy is 0.95 whereas AUC is very low i.e. 0.55. Table 3 represents balanced dataset.

Experiment results in Fig. 2, show that overall accuracy was higher when class was imbalanced as model got biased for non-churner class however AUC for churner class was increased from 0.40 to 0.62 when class was balanced as shown in Fig. 3.

Table 2 Dataset with imbalanced class

Imbalanced class	Non-churner	Churner	Total
TRAIN	436,734 (91%)	43,266 (9%)	480,000

Table 3 Dataset with balanced class

Imbalanced class	Non-churner	Churner	Total
TRAIN	288,000 (60%)	192,000 (40%)	480,000

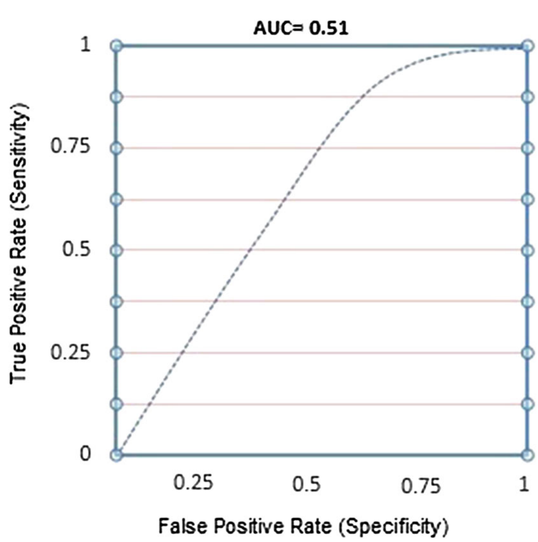

Fig. 2 AUC with imbalanced class

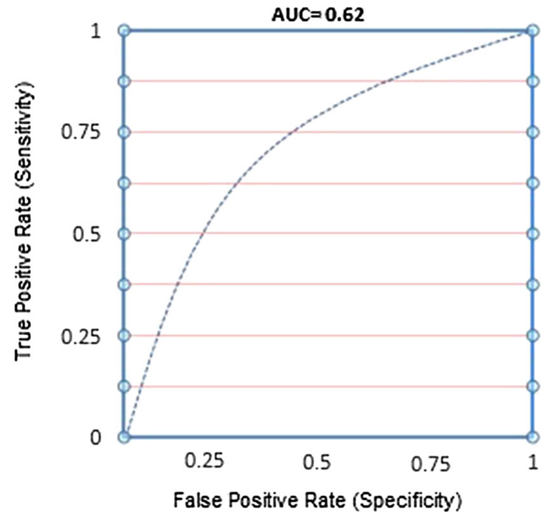

Fig. 3 AUC with balanced class

We have performed the same steps for M2 and M3 datasets and results revealed that significant improvement was observed in Recall rate when churner and non-churner ratio was increased. Three iterations have been performed on complete dataset of 3 months in order to evaluate the effect of class imbalance on recall and accuracy.

4.2 Result 2: Classifiers performance

Different classifiers have been implemented on same data set to evaluate the performance. Neural Netwrok gave better performance in terms of AUC as compared to SVM decision trees and C4.5 and covered more area under curve as TP rate is higher i.e. 0.56. On top of this, from Fig. 4 we can see that TP rate and AUC of fuzzy classifiers was higher because fuzzy

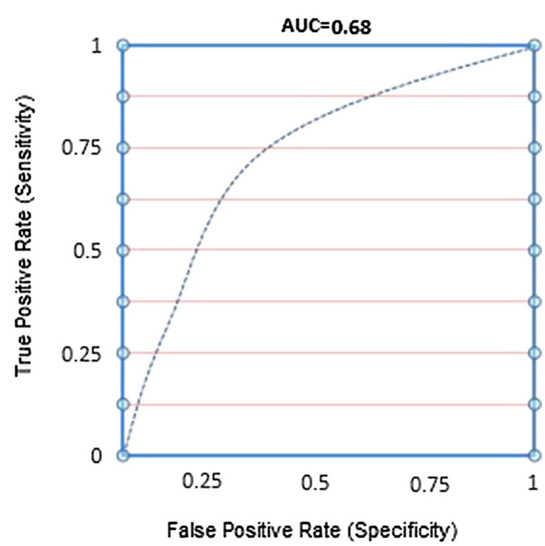

Fig. 4 AUC for OWANN fuzzy classifier

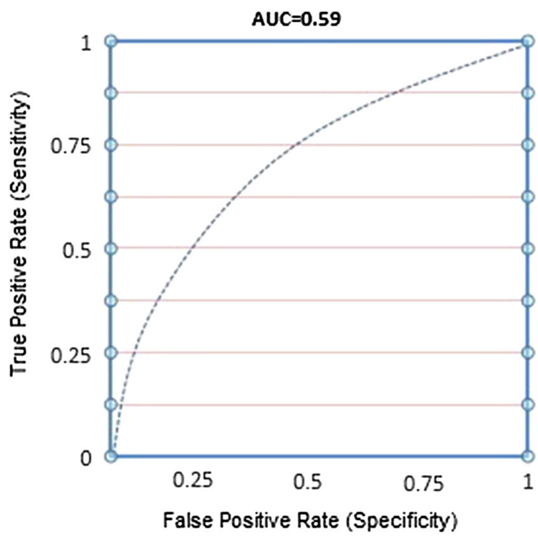

Fig. 5 AUC for OWANN with small number of attributes

Table 4 Classifiers performance

Classifier	TP rate	FP rate	AUC
C 4.5	0.27	0.04	0.57
Decision Tree	0.33	0.15	0.63
SVM	0.33	0.15	0.59
Linear Regression	0.37	0.14	0.6
Neural Network	0.56	0.32	0.64
AdaBoost	0.71	0.32	0.56
Bagging	0.8	0.75	0.58
Random Forest	0.53	0.36	0.63
Gradient Boosting	0.69	0.40	0.52
FuzzyNN	0.61	0.42	0.57
VQNN	0.94	0.94	0.62
FuzzyRoughNN	0.95	0.87	0.66
OWANN	0.98	0.92	0.68

classifiers efficiently handle the missing values as compared to state of the art classifiers.

4.3 Result 3: Classifiers performance with limited features

Set of experiments were conducted after reducing the number of attributes from original dataset. Fuzzy classifiers have been implemented and observed the significant improvement with higher number of attributes on same dataset. Figure 5 represents the AUC curve of fuzzy classifiers with smaller number of attributes (42 attributes reduced from original 84).

4.4 Result 4: Comparative analysis

Under this result set, comparative analysis of different classifiers has been performed. Table 4 gives measures of accuracy, True positive rate, false positive rate and area under curve where TP and FP rate of Fuzzy classifiers is higher as compared to other classifiers. AUC is the best performance measure which has been adopted by many researchers in churn prediction domain as a replacement for accuracy. Figure 6 shows that the best performance in terms of TP rate and AUC is given by Fuzzy classifiers as compared to non-fuzzy classifiers i.e. Multi-layer perceptron, C4.5 and SVM classifier. So, accuracy of a classifier is not the only performance measure criteria for any classifier algorithm. Figure 6 also shows that maximum churners have been captured by Fuzzy Classifiers as TP rate is very high.

Lift curve of fuzzy OWANN has been shown in Fig. 7 and it can be clearly seen that curve starts with a straight line which show that top 10% customers identified as churners are giving the maximum gain. Fuzzy OWANN also gives the best area under curve and true positive rate.

5 Conclusion and future work

In this paper, a fuzzy classifier based customer churn prediction model for telecommunication section has been proposed. The proposed model utilizes the fuzzy classifiers, namely, FuzzyNN, VQNN, OWANN and FuzzyRoughNN to accurately predict the churners from a large set of customer records. Experiments performed on the CDR details of a real telecom company for South Asia, showed that the proposed model is able to achieve a TP rate up to 98% and AUC score of 0.68. A number of predominant classifiers in the literature, namely, Multilayer Perceptron, Linear regression, C4.5, SVM and Decision Tree have been compared with fuzzy based Nearest Neighbor classifiers to highlight the supremacy of fuzzy classifiers in predicting the accurate set of churners.

Future work is intended toward using the fuzzy based feature selection methods i.e. fuzzy rough subset, to see the effect of fuzzy feature selection on classifier's prediction accuracy. Moreover, we plan to do a comparative analysis of prediction model building time with respect to different classifiers in order to assist telecom analysts to pick a classifier which not only gives accurate results in terms of TP rate, AUC and lift curve but also scales well with high dimension and large volume of call records data.

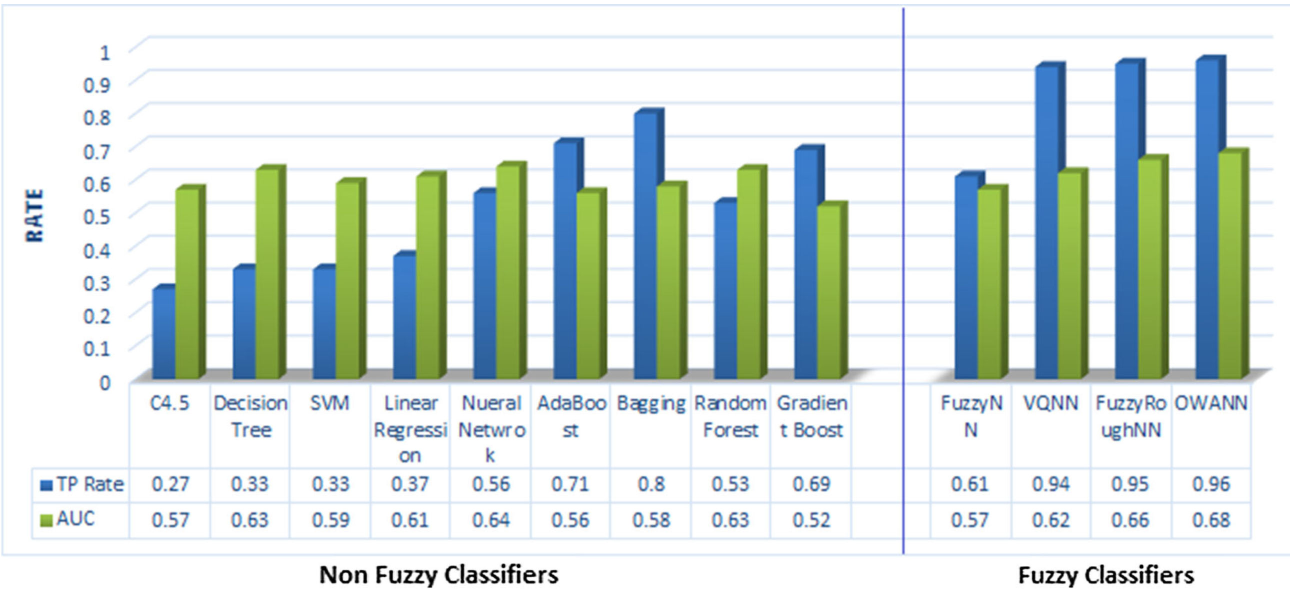

Fig. 6 Performance analysis of different classifiers

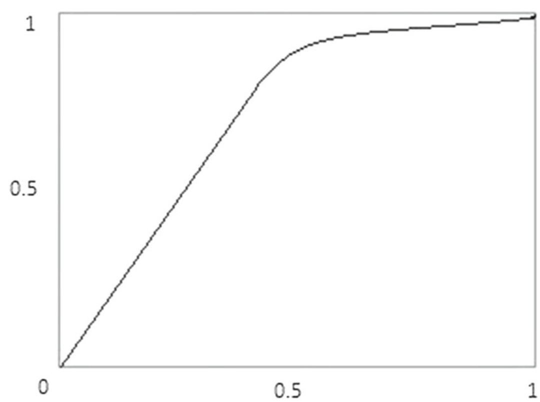

Fig. 7 Lift curve of Fuzzy OWANN

References

1. Kawale, J., Pal, A., & Srivastava, J. (2009). Churn prediction in MMORPGs:A social influence based approach. In *International conference on computational science andengineering, 2009. CSE'09* (pp.423–428).
2. Richter, Y., Yom-Tov, E., & Slonim, N. (2010). Predicting customer churn inmobile networks through analysis of social groups. In *SDM*, 2010(pp. 732–741).
3. Hung, S.-Y., Yen, D. C., & Wang, H.-Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, *31*, 515–524.
4. Tsai, C.-F., & Chen, M.-Y. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems with Applications*, *37*, 2006–2015.
5. Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, *36*, 4626–4636.
6. Song, G., Yang, D., Wu, L., Wang, T., & Tang, S. (2006). A mixed process neural network and its application to churn prediction in mobile communications. In *Sixth IEEE international conference on data mining workshops, 2006*. ICDM workshops 2006 (pp. 798–802).
7. Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, *34*, 313–327.
8. Zhang, Y., Qi, J., Shu, H., & Cao, J. (2007). A hybrid KNN-LRclassifier and its application in customer churn prediction. In *IEEE internationalconference on systems, man and cybernetics, 2007. ISIC* (pp. 3265–3269).
9. Xia, G.-E., & Jin, W.-D. (2008). Model of customer churn prediction on support vector machine. *Systems Engineering-Theory & Practice*, *28*, 71–77.
10. Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2006). Churn prediction usingcomplaints data. In *Proceedings of world academy of science,engineering and technology*.
11. Owczarczuk, M. (2010). Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications*, *37*, 4710–4712.
12. Huang, B. Q., Kechadi, T.-M., Buckley, B., Kiernan, G., Keogh, E., & Rashid, T. (2010). A new feature set with new window techniques for customer churn prediction in land-line telecommunications. *Expert Systems with Applications*, *37*, 3657–3665.
13. Tsai, C.-F., & Lu, Y.-H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, *36*, 12547–12553.
14. Wei, C.-P., & Chiu, I.-T. (2002). Turning telecommunications call details to churn prediction: A data mining approach. *Expert Systems with Applications*, *23*, 103–112.
15. De Bock, K. W., & Van den Poel, D. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, *38*, 12293–12301.
16. Huang, B., Buckley, B., & Kechadi, T.-M. (2010). Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. *Expert Systems with Applications*, *37*, 3638–3646.
17. Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, *38*, 2354–2364.

18. Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B. et al. (2015). Telco churn prediction with big data. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data* (pp. 607–618).
19. Verbraken, T., Verbeke, W., & Baesens, B. (2014). Profit optimizing customer churn prediction with Bayesian network classifiers. *Intelligent Data Analysis*, *18*, 3–24.
20. Maldonado, S., & Montecinos, C. (2014). Robust classification of imbalanced data using one-class and two-class SVM-based multiclassifiers. *Intelligent Data Analysis*, *18*, 95–112.
21. Xie, Y., Li, X., Ngai, E., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, *36*, 5445–5449.
22. Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, *218*, 211–229.

Muhammad Azeem is currently doing his MS degree in computer sciences from Shaheed Zulfikar Ali Bhutto Institute of Science & Technology of Islamabad, Pakistan. Before that he received his M.Sc. degree in computer sciences from the ARID agriculture university, Rawalpindi, Pakistan, in 2006. His main areas of interest are data mining, machine learning and predictive modeling.

Muhammad Usman has completed PhD in Computer & Information Sciences from Auckland University of Technology, New Zealand. He is currently an Associate Professor of Computer Science in the department of Computing at Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan. His research interests include Data Mining, Machine Learning, Data Warehousing, OLAP, Business Intelligence and Knowledge discovery. He is currently researching in the novel methods and techniques for the seamless integration of Data Mining and Data Warehousing technologies. He has published in international journals and conference proceedings, and he has served as reviewer for a number of premier journals and conferences.

A.C.M Fong is an associate professor in the Department of Computer Science, Western Michigan University, USA, and an accredited imaging scientist. His research interest include Machine Learning, Information retrieval, Artificial Intelligence and Multimedia. He has published in international journals and conference proceedings, and he has served as reviewer for a number of premier journals and conferences.