Cluster Comput
DOI 10.1007/s10586-017-1154-3

Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling

Adnan Idris[1] · **Aksam Iftikhar**[2] · **Zia ur Rehman**[3]

Received: 1 April 2017 / Revised: 3 August 2017 / Accepted: 28 August 2017

Abstract Nowadays, telecom industry faces fierce competition in satisfying its customers. This competition thus requires an efficient churn prediction system to identify customers who are ready to quit. Such churn customers are then retained through addressing relevant reasons identified by the churn prediction system. Therefore, now the role of churn prediction system is not only restricted to accurately predict churners but also to interpret customer churn behavior. In this paper, searching capabilities of genetic programming (GP) and classification capabilities of AdaBoost are integrated in order to evolve a high-performance churn prediction system having better churn identification abilities. For this, frequently selected features in various GP expressions evaluated through AdaBoost based learning, are marked and analyzed. Moreover, the issue of imbalance present in telecom datasets is also addressed through particle swarm optimization (PSO) based undersampling method, which provides unbiased distribution of training set to GP-AdaBoost based prediction system. Particle swarm optimization based undersampling method in combination with GP-AdaBoost results a churn prediction system (ChP-GPAB), which offers better learning of churners and also identifies underlying factors responsible for churn behavior of customers. Two standard telecom data sets are used for evaluation and comparison of the proposed ChP-GPAB system. The results show that the proposed ChP-GPAB system yields 0.91 AUC and 0.86 AUC on Cell2Cell and Orange datasets, in addition to identifying the reasons of churning.

Keywords Telecom · Churn prediction · Genetic programming · AdaBoost · Ensemble classification · Feature identification

✉ Adnan Idris
adnanidris@upr.edu.pk

Aksam Iftikhar
aksamiftikhar@ciitlahore.edu.pk

Zia ur Rehman
zia1986@gmail.com

[1] Department of Computer Sciences & IT, The University of Poonch, Rawalakot, Pakistan

[2] Department of Computer Sciences, COMSATS Institute of I.T, Lahore, Pakistan

[3] Department of Computer Sciences, COMSATS Institute of I.T, Abbottabad, Pakistan

1 Introduction

Telecom industry has worldwide witnessed rapid development amongst subscription based businesses over the last decade. The number of mobile phone users has reached nearly 7 billion around the world by the start of year 2015, which is almost equivalent to world population [1]. Some of the advanced countries already have more telecom subscribers than inhabitants [2]. Therefore, telecom markets are facing customer saturation and retaining customers is becoming top priority for telecom service providers in order to maintain stable profit. Moreover, it is widely accepted in relevant literature that retaining a customer is more beneficial practice than acquiring a new client [3]. Acquiring a new customer not only costs more (5–6 times than retaining a customer) but also requires a time period for developing customer loyalty with the service provider subject to the satisfaction of demanded services. Whereas, retaining a customer does not involve any additional marketing or other expense, rather an attention to the resolution of customer's concerns is enough in most of the cases. Further, long term customers generate more profit because they are not easily attracted by other competitors and in addition, can refer new customers and eventually become less costly to attend. Thus a fractional improvement in customer retention can considerably impact the growth and sustainability of telecom business [4]. Consequently, an accurate churn prediction system having better interpreting capabilities is essentially required to identify the churn customers and their reasons of churning.

Generally, classification approaches utilize customer characteristics expressed in personal demographics, account and billing information, and call details for predicting future behavior of customers. Initially, data mining approaches for churn prediction were primarily used to accurately predict the telecom churners. For example, decision trees and neural networks have been used to develop accurate churn prediction systems [2]. Various variants of Rotation forest and RotBoost have also been explored for predicting telecom churners with higher accuracy [3]. De Bock et al. have reported a hybrid learning system that combines K-means clustering and a rule technique with an objective of attaining higher prediction accuracy [4]. Similarly, a genetic algorithm based neural network approach has been presented in [5] to maximize the prediction accuracy of telecom churners. In short, numerous data mining approaches based on individual or ensemble classification methods have been used to attain improved prediction performance for customer churn in telecom industry as given in Table 1. However, most of churn prediction approaches utilize feature extraction, sampling or both before employing a classification algorithm. Burez J. et al have presented a comprehensive study analyzing the impact of sampling methods in improving churn prediction performance [6]. In their study, Random undersampling and advanced undersampling methods were explored with gradient boosting and weighted random forest for attaining improved prediction performance. Moreover, it is also highlighted in the same study through using CUBE sampling technique that applying sophisticated sampling methods do not result significant improvement in prediction results, supporting the fact quoted in [7]. Similarly, Verbeke W. et al have reported that mere duplicating instances through oversampling does not offer considerable advantage in improving prediction results [8]. This also supports the notion that only duplicating minority class through random oversampling or discarding majority class using random undersampling may not improve prediction results [6]. Whereas in [9], a PSO based intelligent undersampling method helps in the improved learning of KNN and random forest and thus resulting in better churn prediction performance.

Similarly, a number of studies have focused in using only meaningful features for inducing classification algorithms for telecom churn prediction. A set of features including personal demographics, bill and payment records, call details, account information, services orders, line information and Henley segmentation are provided to decision trees, multilayer perceptron neural network and support vector machines for predicting telecom churn [10]. Huang B. et al proposed a multi-objective feature selection method using NSGA-II for telecom churn prediction [11]. In another approach, Bayesian Belief Network extracts most important features to be used for customer churn prediction [12]. Likewise, in [13] a hybrid two phase method based on feature extraction, has been proposed for predicting telecom churners. In another approach presented in [14], certain number of features from the call detail records of telecom customers is derived and then logistic regression is employed for churn prediction.

However, in a few approaches only, besides focusing on predicting churners with higher accuracy, comprehensibility and intuitiveness of a churn prediction system are also studied to identify the reasons of customer churn behavior [12,15,16]. Thus, in this work, we propose an efficient churn prediction approach, based on exploiting powerful searching capabilities of genetic programming (GP) supported through AdaBoost based iterative approach. The features frequently selected in evolving GP programs over various AdaBoost based iterations can be considered to identify factors representing churn behavior of telecom customers. The aim of this study is to exploit the learning and searching capabilities of proposed integrated approach (GP-AdaBoost) for developing an efficient and intuitive churn prediction system. In addition, PSO based undersampling method is also employed for coping with data imbalance in the proposed approach prior to training of GP-AdaBoost ensemble.

The rest of paper is organized such that Sect. 2 discusses proposed methodology including details of the GP-AdaBoost ensemble. The parameters of GP-AdaBoost ensemble is also discussed in the same section. Section 3 presents the telecom datasets used in this work. Results, discussions and comparative analysis are made in Sect. 4. Finally conclusion is drawn in Sect. 6.

2 Proposed methodology

GP has successfully handled various problems of clustering, regression, and discovering association rules. But, the flexible and distinctive capabilities of GP make it an appropriate technique for classification [17–19] as well and so thus for churn prediction. Likewise, AdaBoost is another classification technique based on boosting, which works iteratively to ensemble multiple weak classifiers. In each iteration, a classifier is evolved by learning from the hard instances that are not correctly classified in previous iteration [20]. This work employs an ensemble classification approach (GP-AdaBoost) based on incorporating the concept of boosting in evolving various GP programs. A fixed number of GP programs is are evolved per class in a single iteration. Further, AdaBoosting extends the weight updation

Table 1 Recent research work on telecom churn prediction

Title	Methods	Source
Bagging and boosting classification trees to predict churn	Bagging, stochastic gradient boosting	Lemmens and Croux [29]
Applying data mining to churn management	Decision trees, NN	Shin et al. [2]
Hybrid models using unsupervised clustering for prediction of customer churn	C5.0 boosting	Bose and Chen [30]
Handling class imbalance in customer churn prediction	Random forests, weighted random forests gradient boosting machine	Burez and Poel [6]
Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services	GA, NN	Pendharkar [5]
Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications	NSGA-II	Huang et al. [11]
Building comprehensible customer churn prediction models with advanced rule induction techniques	AntMiner+, ALBA, C4.5, RIPPER	Verbeke et al. [15]
Monitoring and backtesting churn models	Logistic regression, decision trees	Lima et al. [31]
Empirical evaluations of rotation based ensemble classifier	Rotation forest, rotBoost	Bock and Poel [3]
Applying Bayesian Belief Network approach to customer churn analysis: a case study on the telecom industry of Turkey	Bayesian Belief Network	Kisioglu and Topcu [12]
A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data	H-MK-SVM	Chen et al. [32]
Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models	Bagging, random subspace method, generalized additive models	De Bock et al. [16]
New insights into churn prediction in telecommunication sector: a profit driven data mining approach	SVM, LSSVM	Verbeke et al. [8]
An effective hybrid learning system for telecommunication churn Prediction	K means, weighted K means, inductive rule learning	Huang and Kechadi [4]
Improved churn prediction in telecommunication industry using data mining techniques	ANN,KNN, SVM, decision trees	Keramati et al. [33]
Improved churn prediction in telecommunication industry by analyzing a large network	Logistic regression, MLP	Kyoungok et al. [34]
A customer churn prediction model in Telecom industry using boosting	Boosting	Ning et al. [35]
Customer churn prediction in the telecommunication sector using a rough set approach	Rough set approach	Amin et al. [36]
Customer churn prediction in telecommunication industry: with and without counter example	Rule generation algorithms	Amin et al. [37]
Aprudent based approach for customer churn prediction	Ripple down rules and prudence analysis	Amin et al. [38]

over instance space. In this way, the optimization process of GP is supported by AdaBoost to enhance the prediction performance. Thus, capabilities of GP-AdaBoost ensemble are used to predict churners with higher accuracy. In addition, GP-AdaBoost also identifies the features which represent the reasons for churn behavior of customers. The block diagram given in Fig. 1 shows the proposed churn prediction system (Ch-GPAB).

2.1 Pre-processing of dataset

The telecom dataset having anomalies like missing values or empty features, are tackled using filters provided by freely available WEKA tool. Moreover, the nominal features in the dataset are converted to numerical format by grouping the instances in small, medium and large categories; subject to the number of occurrences belonging to each category [21]. The preprocessing of dataset establishes a uniform numerical format in the training set.

2.2 PSO based undersampling

Telecom datasets generally have fewer instances of churner class and most of the dataset comprises of non-churner instances. The imbalance class distribution present in telecom datasets mostly results in low prediction performance of classification algorithms. The majority of non-churner

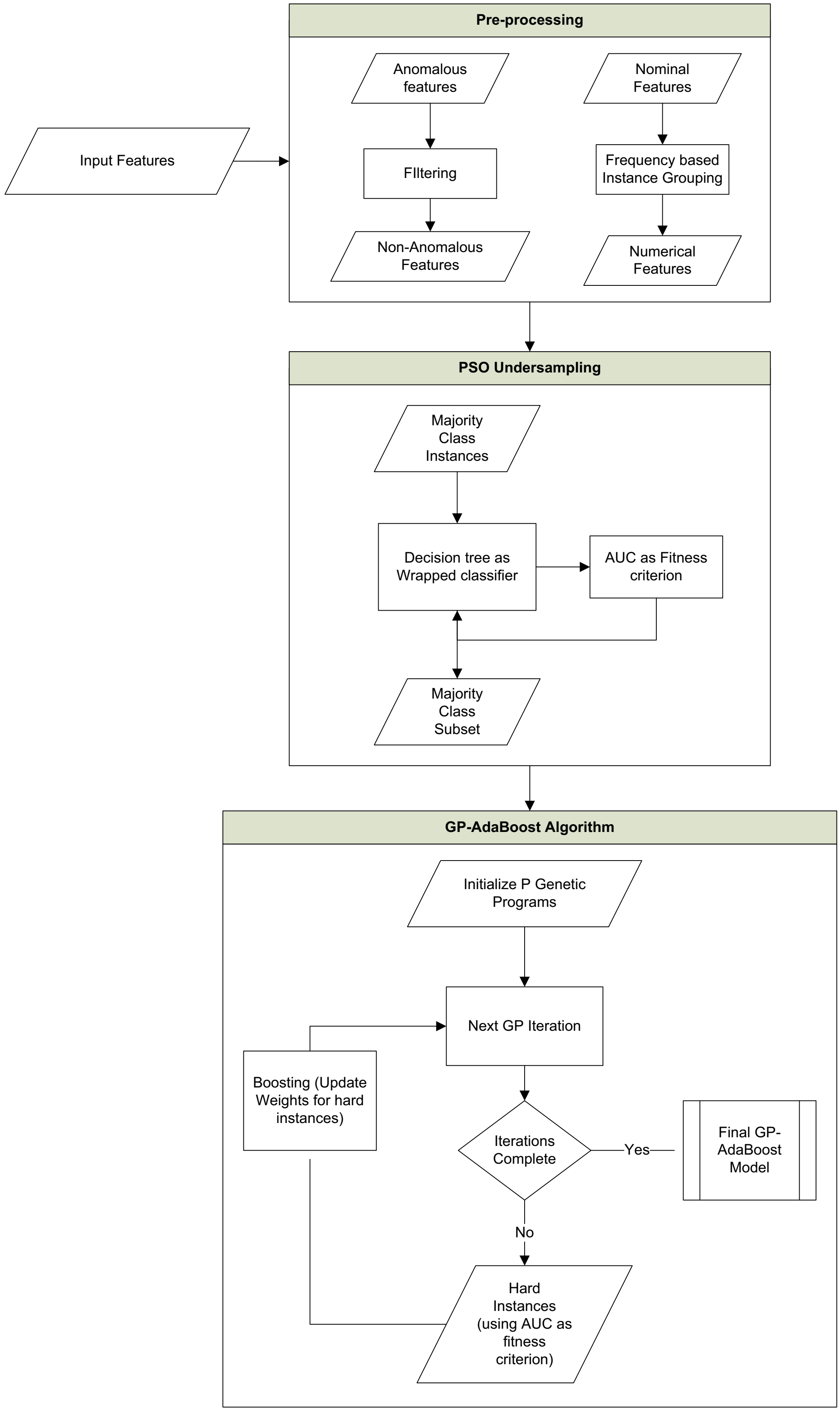


Fig. 1 Block diagram of the proposed Ch-GPAB churn prediction system

instances in training sets tend to overwhelm the classifier through biased learning, which ultimately results in lower prediction performance. Oversampling is a resort for resolving the problem of imbalance in class distribution. However, due to inherently large volume of telecom datasets, oversampling is not a feasible option as it may result in even larger sized dataset. Undersampling, on the other hand, downsamples majority class instances by selecting discriminative samples for yielding a balanced class distribution [9,22,23]. Therefore, an intelligent PSO based undersampling method is employed along with a wrapped classifier, which yields a balanced and discriminative training set.

Particle swarm optimization based undersampling method [24] can be used to tackle the imbalanced class distribution of the telecom dataset. A classifier is involved in undersampling, which recognizes less informative instances of the majority class. Such instances are removed and only good instances of majority class are retained which are equal in number to minority class instances. In this manner, a number of discriminative subsets are produced comprising minority and majority class instances. Each candidate subset is evaluated by using a wrapped classifier. The classifier evaluates the fitness of each subset using area under the curve (AUC) criteria. A frequency list is maintained that represents the number of times an instance from majority class is selected. Finally, a balanced training set is developed by combining the frequently selected instances of majority class with minority class instances. Particle swarm optimization based under sampling steers the particles in search space, while considering subsets of majority class instances as particles. Pericles' movement is steered by their own best and swarm's best positions.

We consider n particles in a population, where index i (i = 1, 2, 3,…, n) represent a particle in the swarm, index j (j = 1, 2, 3,…, m) indicates dimensions and t counts the number of iterations. A particle having good AUC is selected as a result of an iterative process. Following equations are used for updating the velocity $v_{i,j}(t)$ and position$x_{i,j}(t)$ of the ith particle:

$$v_{i,j}(t+1) = w.v_{i,j}(t) + c_1 R_1.\left(pbest_{i,j} - x_{i,j}(t)\right) + c_2 R_2.\left(gbest_{i,j} - x_{i,j}(t)\right) \tag{1}$$

$$x_{i,j}(t+1) = \left\{ \begin{array}{l} 0 : if random\,() \geq s\left(v_{i,j}(t+1)\right) \\ 1 : if random\,() < s\left(v_{i,j}(t+1)\right) \end{array} \right\} \tag{2}$$

$$S\left(v_{i,j}(t+1)\right) = \frac{1}{1+e^{-v_{i,j}(t+1)}}\ (3) \tag{3}$$

where $pbest_{i,j}$ show the previous best position and $gbest_{i,j}$ represents the best position obtained by informers. c_1 and c_2 are cognitive and social accelerators, respectively. R_1 and R_2 are random numbers between 0 and 1, and w is the inertia weight. The cognitive and social accelerators are indicated as c_1 and c_2 respectively, whereas R_1 and R_2 are random numbers between 0 and 1. Similarly, w is used for inertia weight. Lastly, the ranking of majority class instances are accomplished on the basis of their selection frequency in model building as represented in Fig. 2. The most frequent an instance is selected with regards evaluating its prediction capability, the highest the rank is assigned to the instance. The balanced training set is developed by combining instances having higher ranks in the frequency list with the minority class instances. m dimensional particle space is considered, which corresponds in size, equal to the number of majority class instances.

Empirical evaluation has been performed for selecting suitable parameter values in GP-AdaBoot and PSO based undersampling methods. After multiple experiments, best parameter values have been selected as listed in Table 2 and used in subsequent experiments.

2.3 GP-AdaBoost based churn predictor

Customer churn prediction is a binary classification problem [25,26]. Classification algorithms establish criteria of deciding a particular class for a test instance, subject to the values of certain features. In churn prediction, a binary classifier is induced with a set of instances labeled with correct classes and then learned classifier is used to differentiate test instances of churners and non-churners. GP is searching method capable of attaining optimal solution [27,28] and thus can be used to induce a classifier to be a churn predictor.

The role of AdaBoost in GP-AdaBoost algorithm is to extend boosting in evolution of multiple GP programs per class. AdaBoost iteratively evolves multiple GP programs, where each GP program recognizes those instances which are incorrectly classified in previous iteration. Boosting extends the weight updation support over instance space to handle the hard instances. The AUC is used as fitness function. Introducing boosting with in GP strengthens the optimization process and extends robustness in evolving multiple GP programs. Besides improving the results, the boosting in GP-AdaBoost also saves the time because in order to evolve new GP program, the whole new population is not created rather weights are updated over instance space for next iteration. The final prediction of a test instances is made on the basis of the higher value, from a weighted sum of the outputs of GP programs per class. The pseudo-code given in Fig. 2(a) shows the steps involved in GP-AdaBoost algorithm. Figure 2(b) represents the GP's evolution is supported by weight updation through Boosting.

Fig. 2 **a** Pseudo code of GP-AdaBoost algorithim **b** prediction of test instance using Ch-GPAB churn prediction system

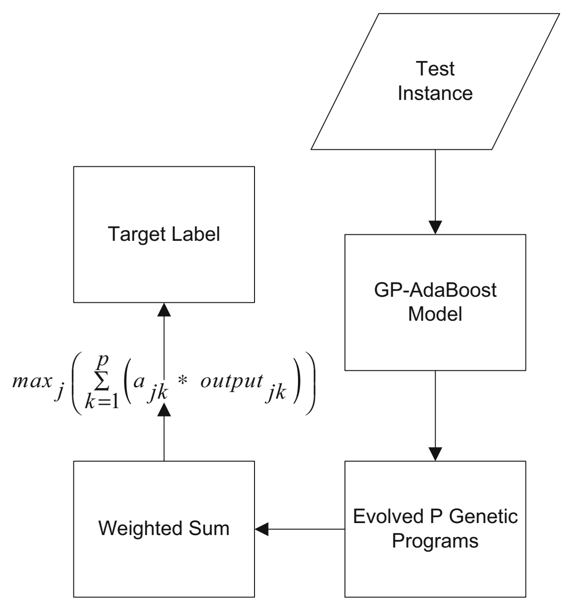
Step by step procedure of GP-AdaBoost

Inputs: *C, P, T, N*
[C = Number of target classes (in our case 2)
P = Number of GPs required in boosting
T = Training dataset
N = Number of instances in T]
For *j = 1 to C (GP strings are obtained for each class)*
GP_Init(POP) (Population Initialization)
Init_Weights(W) (Weight Initialization as $W_{i} = 1/N$ *for each instance i)*
For *k=1 to P (GP strings evolved using Ada boosting)*
Fill POP. New set of GP strings
Evolving GP string capable of recognizing Class $C_{j,}$ *keeping AUC as fitness function*
end For
end For
*(P * C) GP strings are acquired for every class Cj*
Accumulative classification result of each class Cj is calculated by weighted sum of α_{jk} *and* output_{jk} *of the GP strings.*

$$max_j\left(\sum_{k=1}^{p}\left(a_{jk} * output_{jk}\right)\right)$$

The j which, scores maximum indicates the class of a test instance

(a) Pseudo code of GP-AdaBoost Algorithim



(b) Prediction of Test Instance using Ch-GPAB Churn Prediction System

2.3.1 GP-AdaBoost parameter settings

GP-AdaBoost algorithm approaches churn prediction as a one class problem by treating churners and non-churners separately as represented in algorithm given in Fig. 2. P number of GP programs is are evolved for each class using boosting. The obtained results are summed up for each class and the higher maximum output identifies the class of a test instance. The parameters listed in Table 2 for GP-AdaBoost based algorithm are fixed empirically after performing number of experiments.

The Elite size identifies the number of GP programs per class, which are evolved through boosting. For population generation, 0.07 cross over, 0.90 mutation and 0.03 reproduction values rates are set as 0.07, 0.90, and 0.03, respectively for population generation. Higher value of mutation ensures

Table 2 Parameter values used in GP-AdaBoost algorithithm

Parameter name	Value
Number of generation	20
Number of GP strings per class	5
Fitness function	AUC
Functions	+,/,*,If,>,<,Pow,&,|,Max,Min, Exp,Log
Max depth of trees	5
Mutation	0.90
Cross over	0.07
Reproduction of new programs	0.03
Population size	100
Population initializer	Ramped half and half method

Table 3 Characteristics of used telecom datasets

	Orange telecom	Cell2Cell
Total instances	50k	40k
Total features	260	76
Numerical features	190	68
Nominal features	70	8
Data distribution	Imbalanced(7.3% minority class)	Balanced
Missing values	Yes	No

diversity for each consequent generation. Ramped half and half method is used for population initialization while AUC is used as fitness measure. GP-AdaBoost algorithm is used to develop efficient churn predictor for telecom thus AUC is the appropriate fitness criteria to be considered for evolution of a churn predictor.

2.3.2 Datasets

The data sets used in this work are readily available on internet and are mostly used in contemporary literature of telecom churn prediction. Orange telecom provides a data on its internet based resource [39]. The other dataset, provided by Cell2Cell is available at Center for customer relationship management Duke University's website [8]. Orange dataset is provided online with its original imbalanced class distribution, whereas Cell2Cell dataset is preprocessed and a balanced version is provided for research purposes. The feature names in Orange dataset are hidden to keep the customers' information private. Some of the features in Orange dataset have no value at all or have only one value. Considering the uninformative nature of such features, they are removed from the dataset. Finally, both the datasets include a few nominal features, which are transformed into their equivalent numerical representation for maintaining consistency of numerical format in whole dataset. Table 3 represents important attributes of both datasets.

2.3.3 Handling imbalance through PSO based undersampling

Generally, telecom datasets have imbalanced class distribution because there are fewer class instances of minority class. Particle swarm optimization based undersampling method is used in proposed approach to diminish the imbalanced class distribution of a telecom dataset. This undersampling method selects the majority class instances, which are suitable for inducing decision tree classifier. Such instances of majority class are combined in comparable numbers with instances of minority class and a balanced training dataset is obtained. Orange dataset comprises of 50,000 instances where only 3276 are churner instances. After applying PSO based undersampling, the training dataset has balanced class distribution with equal number of churners and non-churners. Whereas, in case of Cell2Cell dataset, a processed version of dataset is provided, that has already balanced class distribution (Table 4).

Compared to other random undersampling methods, PSO based undersampling method is more effective for its capabilities to remove those instances, which are less informative for inducing learning to a classifier. The instances of majority class which are higher ranked on the basis of evaluated fitness are combined with instances of minority class to develop a balanced dataset. Thus, PSO based undersampling avoids the problem of loss of information caused by random undersampling [9].

Table 4 Class distribution of datasets after applying PSO based undersampling

	Orange telecom	Cell2Cell
Minority class	3276	20,000
Majoirty class	3276	20,000
Total	6552	40,000

3 Performance measures

When the predictor is developed, it is important to evaluate its performance to assess how well it can decide about the future conduct of telecom customers. AUC, sensitivity and specificity based measures are used to evaluate the prediction performance of the proposed Ch-GPAB churn prediction system. The said performance measures are employed due to their popularity considered in contemporary literature for the evaluation of classification approaches, which are used for churn prediction [34,40].

Sensitivity and Specificity measures are defined with the help of following counts: True positives (TP), false negatives (FN), true negatives (TN) and false positives (FP), where positives (P) refer to churners and negatives refer to non-churners (N). Thus, P = TP + FN and N = TN + FP. Equations (4) and (5) represents sensitivity and specificity measures, whereas AUC is computed with the help of formula given in equation (6) [6]:

$$Sensitivity = \frac{TP}{P} \tag{4}$$

$$Specificity = \frac{TN}{N} \tag{5}$$

$$AUC = \int_0^1 \frac{TP}{P} d\frac{FP}{N} = \frac{1}{P.N}\int_0^N TPdFP \tag{6}$$

4 Results and discussion

This section discusses the telecom datasets used in current study and explores prediction performance of the proposed Ch-GPAB churn predictor on these data sets. A comparative analysis of Ch-GPAB with other existing approaches is also included in this section. The prediction results obtained in this study are evaluated using 10-fold cross validation. A telecom dataset (D) is divided into 10 subsets (D_1, D_2, $D_3 \ldots D_{10}$). The subsets [$\forall_i | D - D_i$] fold are used in training phase of Ch-GPAB, whereas D_i folds are used for testing in a single iteration ranging from 1 to 10. Final results are computed by accumulating the results over all iterations. Telecom datasets are usually of larger size and therefore, 10-fold cross validation becomes computationally expensive. However, the results obtained using 10-fold cross validation is considered to be more reliable. Two of the commonly used telecom datasets are used in this study for performance evaluation of the proposed churn prediction system.

4.1 Performance analysis of Ch-GPAB

The proposed churn prediction approach is based on the idea of combining GP and AdaBoost algorithms. In [41], an intelligent churn prediction system is developed for predicting churners, however, the reported technique does not have the capability of interpreting customer's churn behavior. Whereas, the proposed Ch-GPAB interpret the customers' churn behavior through analyzing frequently selected features in evolving multiple GP programs per class. Moreover, proposed Ch-GPAB also attains improved prediction performance by incorporating PSO based undersampling. This work uniquely attempts of using GP-AdaBoost ensemble classifier in combination with PSO based undersampling to interpret customers churn behavior in addition to achieving improved prediction performance. GP is well known independent evolutionally algorithm, which offers flexibility as well as interpretability to model many complex problems. Similarly, AdaBoost extends better classification capabilities through adjusting weights of hard instances. GP-AdaBoost adopts the iterative boosting approach for weights updation over instance space while evolving P number of GP programs. The GP's evolution process is supported through using boosting in GP-AdaBoost algorithm. Moreover, GP-AdaBoost works as a one-class classifier by evolving multiple GP programs per class. Final output prediction for a test instance is made using the high-level output obtained using weighted sum of the outputs of programs per class. It is observed that integration of the boosting in GP is efficient both in terms of prediction performance and time required in evolving P GP programs. GP-AdaBoost algorithm does not need to create a whole new population of programs for evolving each second program of a same class because of a program specializing into the hard instances may already be found in the population. This obviously results in time efficiency.

4.2 Improved learning using balanced dataset

Tables 5 and 6 show the performance attained by various classifiers for predicting churners on Orange and Cell2Cell datasets. The imbalanced original orange dataset is used here for experimentation. It can be clearly observed from the results given below in Table 5 that all of the classifiers show deteriorated performance on Orange dataset. Although, GP-AdaBoost comparatively performs better but it is not up to desired level. Whereas, in case of Cell2Cell dataset (Table 6),

Table 5 Prediction performance on orange dataset

	Orange telecom		
	Sensitivity	Specificity	AUC
Random forest	0.0049	0.9991	0.571
Rotation forest	0.0026	0.9998	0.583
RotBoost	0.0291	0.7212	0.601
GP-AdaBoost	0.3120	0.7501	0.631

Table 6 Prediction performance on Cell2Cell dataset

	Orange telecom		
	Sensitivity	Specificity	AUC
Random forest	0.690	0.601	0.592
Rotation forest	0.646	0.666	0.610
RotBoost	0.664	0.632	0.699
GP-AdaBoost	0.87	0.891	0.910

used classifiers are performing well, where higher prediction performance is attained by GP-AdaBoost. Results shown in Table 5 clearly hint that imbalanced class distribution in training set induce the classifiers with biased learning, resulting in suffered performance. The used classifiers perform well on Cell2Cell dataset for its balanced class distribution compared to orange dataset where data is highly imbalanced. Therefore, a sampling methodology is required to be used for settling the imbalanced class distribution of a training set.

4.3 Performance improvement using PSO based undersampling

This section compares the results of prediction performance of Ch-GPAB with and without applying PSO undersampling. Such a comparison facilitates realizing the actual performance boost-up caused by employing PSO undersampling. Figure 3 shows that GP-AdaBoost algorithm enhances the prediction performance when PSO balanced training set are provided for learning. Improved sensitivity and AUC bars ascertain the fact that GP-AdaBoost is able to efficiently predict churners when PSO balanced training set is used.

4.4 Average performance of PSO based undersampling

Particle swarm optimization is a stochastic process therefore, we have performed series of simulations to validate the effectiveness of the method. Sampling process involves internal 3-fold stratified cross validation for creating balanced training set. Decision tree is wrapped in PSO based undersampling method and obtained AUC is used as a fitness

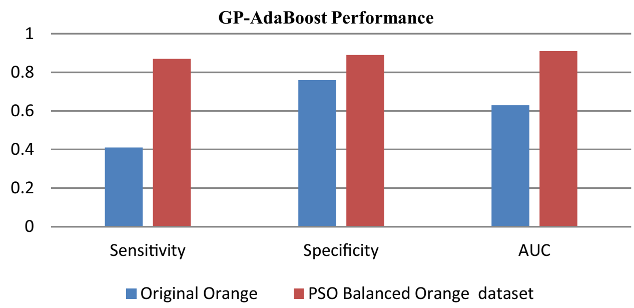


Fig. 3 Performance improvement attained by GP-AdaBoost after applying PSO based undersampling

evaluation criterion. Figure 4 shows the average performance of PSO based undersampling method. 200 iterations are executed in each PSO based simulation and the best evolved AUC is plotted in a graph (Fig. 4). AUC is considered as fitness criterion for the internal optimization. Figure 4 displays the deviation in AUC, computed over thirty PSO simulations. Error bars represents, deviations of AUC from mean value. The error bars are plotted with the condition of having standard deviation 1 on both sides of mean value.

4.5 Improved prediction performance of Ch-GPAB

Table 7 shows the prediction performance of proposed Ch-GPAB on the two datasets. The primary objective of a churn predictor is to classify the churners with higher accuracy. In this connection, the results reported in Table 7 clearly demonstrate the appreciable prediction performance of Ch-GPAB in recognizing churners. Ch-GPAB attains sensitivity scores of 0.89 and 0.93 on Orange and Cell2Cell datasets, respectively. Moreover, GP-AdaBoost, to the best of our knowledge, attains best prediction performance (in terms of AUC) reported so far on Orange and Cell2Cell datasets, respectively. The considerable difference between prediction performances on the two data sets can be explained by different demographics of both data sets. But, overall the prediction performance is a favorable indicator for Ch-GPAB system to be a reasonable candidate for industrial use. As additional evidence to its good classification performance, ROC curves are also drawn in Figs. 5 and 6 for both data sets. These figures clearly demonstrate the efficient performance of GP-AdaBoost as churn predictor.

4.6 Average performance of Ch-GPAB

Ch-GPAB churn prediction approach is based on exploiting the capabilities of GP-AdaBoost ensemble classification approach. Figure 7 shows the average performance of GP-AdaBoost algorithm. GP-AdaBoost simulation is run thirty

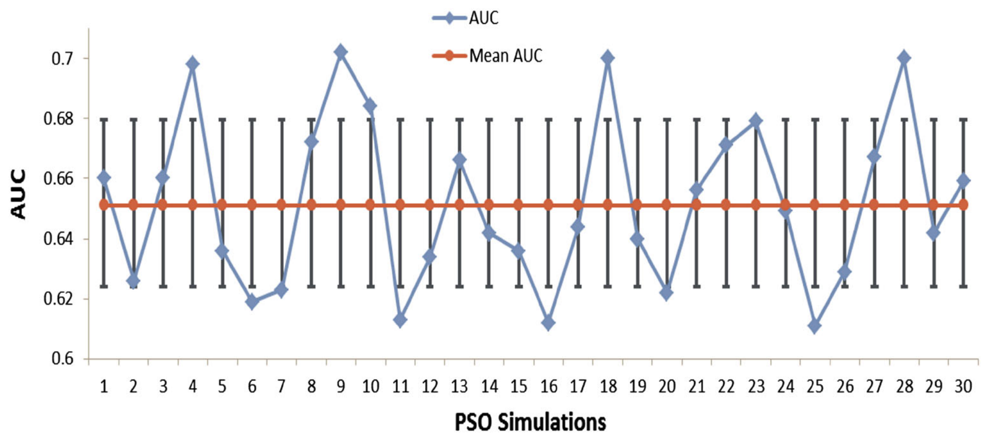


Fig. 4 Average performance of PSO based undersampling

Table 7 Prediction results on orange and Cell2Cell datasets

	Sensitivity	Specificity	AUC
Orange dataset	0.835	0.892	0.86
Cell2Cell dataset	0.904	0.933	0.91

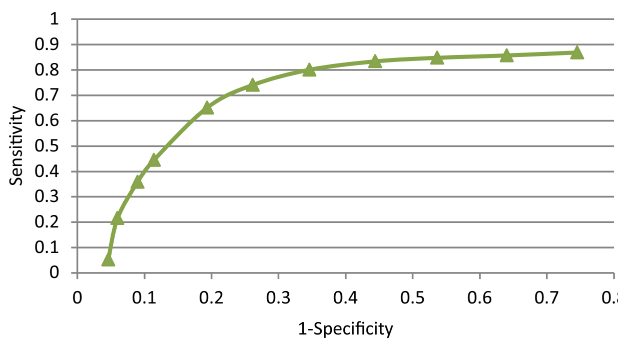


Fig. 5 Prediction performance of Ch-GPAB on orange dataset

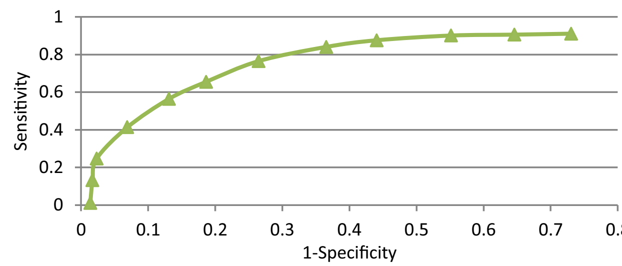


Fig. 6 Prediction performance of Ch-GPAB on Cell2Cell dataset

times using parameter values given in Table 2 to observe the average performance. The AUC attained by GP-AdaBoost ranges between 0.70 to 0.91 over thirty independent simulations as shown in graph given in Fig. 8. Error bars show the deviation of AUC from mean value. Error bars are drawn using standard deviation 1 on both sides of mean value.

4.7 Intuitive reasons of churning based on frequently selected features

Ch-GPAB efficiently predicts telecom churners and it is also effective for investigating the churn behavior of customers. In Ch-GPAB, P numbers of GP programs have been evolved per class using boosting approach. Each evolved GP program comprises of a function set, which is developed using the given terminals and available features of telecom dataset. The features which are frequently chosen in P GP programs can be used to investigate the churn behavior of the customers. The histogram in Fig. 8 (a) shows the number of times each feature is selected in evolving GP programs per class for cell2cell data set. The most frequent feature highlights the significance of a particular feature in explaining the churn behavior of a customer. The “marryno” feature identifies the marital status of a customer as single. It is most frequently selected feature from 76 available features of the Cell2Cell dataset, which shows the significance of this feature in investigating the churn behavior of customers. Similarly, both ”truck” and “refer” are the second most frequent features, which show, respectively, that customer owns a truck (probably because he is a may be a goods forwarding operator) and referrals made by a customer. “dropvce” is in the list of third most frequently selected feature from telecom dataset and it shows the status of voice quality. Considering these features, the potential areas can be marked, which may consequently be explored by marketing managers for effective decision making. Moreover, attractive tariff plans and other benefits can be offered to a niche market through investigation of the features frequently selected by Ch-GPAB. The features’ names in the orange dataset are hidden therefore frequency reported in Fig. 8 (b) cannot be discussed for investigating customers’ behavior. However, numbered features can still be identified and ranked according to their frequency of selection.

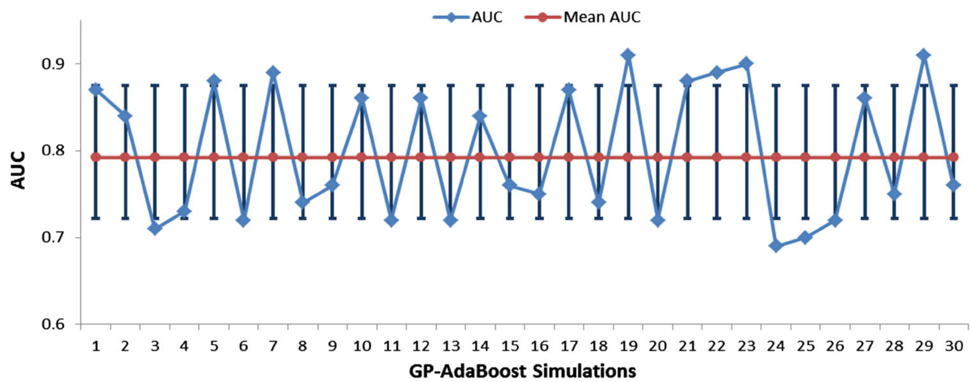


Fig. 7 Average performance of GP-AdaBoost

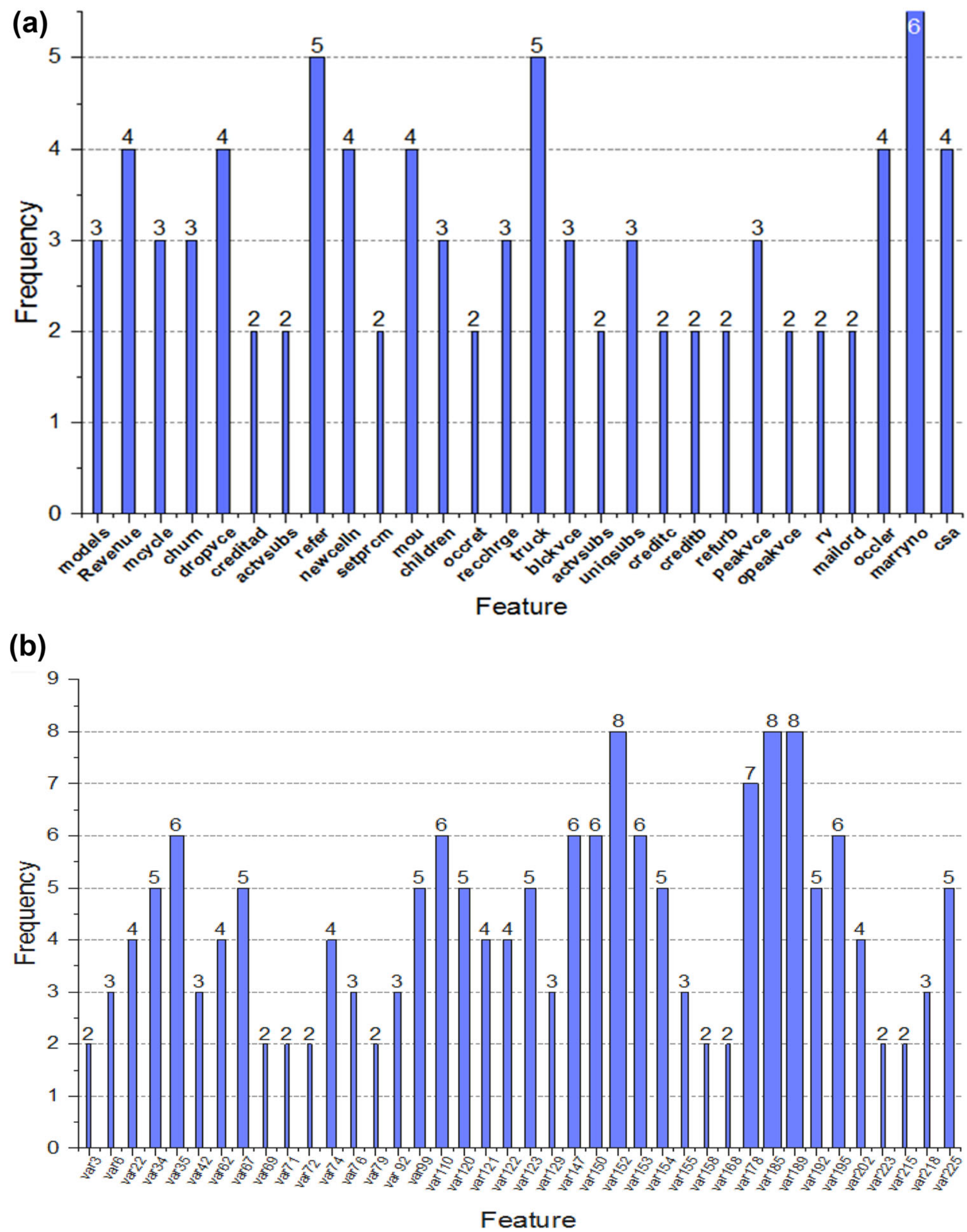


Fig. 8 The frequency of selecting different features in evolving multiple GP programs per class **a** Cell2Cell **b** Orange

5 Performance comparison with other existing approaches

The proposed Ch-GPAB churn prediction approach shows better prediction performance compared to other existing approaches as shown in Table 8. Ch-GPAB considerably performs better in predicting telecom churners when evaluated on Orange dataset. In [9], random forest and minimum redundancy and maximum relevance based features have been exploited to predict churners from Orange dataset and reported prediction performance was 0.751 AUC. Whereas, the proposed Ch-GPAB attains 0.862 AUC. The prediction performance of Ch-GPAB is also compared with a churn prediction approach based on Gradient Boosting Machine [39], which attains 0.737 AUC on Orange dataset. Decision trees with boosting, are used as base classifier in Gradient Boosting approach, where it adopts a ranking based criteria for feature selection. The instances of the training dataset are split into 1% quantile and then half of the training dataset is used to calculate the mean response. The obtained mean response is also applied to the other half, leading to the calculation of AUC that grades the variables. This methodology is based on imputations for preforming feature selection.

The prediction performance attained using Ch-GPAB, is also compared with an AdaBoost based approach [42]. This AdaBoost based approach is enhanced through using multi armed bandit (MABs). In this methodology, AdaBoost makes a pool of simple bases classifiers and then majority voting criteria is applied for obtaining final predictions. This churn prediction methodology develops the data subsets, which are optimized through using MABs and finally AdaBoost is only employed to search these subsets instead of optimizing the base classifier over the whole space. The results given in Table 8 represent that tree and stump based learners with AdaBoost, attain 0.725 AUC and 0.715 AUC, respectively.

In another study, a Bayesian network with oversampling obtains 0.714 AUC to predict churn customers of Orange dataset [8]. Our Ch-GPAB approach attains 0.862 AUC on Orange dataset, which is best score on Orange dataset, to best of our knowledge. Moreover in Ch-GPAB approach a novel combination of AdaBoost and GP algorithm is applied for predicting telecom churners. Proposed churn prediction approach not only achieves improved prediction performance but also offers interpretation capabilities for investigating churn behavior of customers. The additional capability of interpreting customer behavior using frequently selected features in Ch-GPAB is a considerable contribution to the field of telecom churn prediction. Almost all of the methods listed in Table 8 focus only on accurately predicting telecom churners with higher accuracy. On the other hand, our proposed Ch-GPAB offers both higher prediction performance as well as capabilities for investigating the reasons behind churn behavior of customers.

Table 8 Performance comparison on orange dataset

Method	AUC
Ch-GPAB	0.862
Chr-PmRF approach[9]	0.751
Gradient boosting machine[39]	0.737
Decision stump based model[42]	0.725
Decision tree based model [42]	0.715
Bayesian net (BN) based approach[8]	0.714

Table 9 Performance comparison on orange dataset

Method	AUC
Ch-GPAB	0.910
Naïve Bayes (NB) based approach [8]	0.818

Table 10 Mecnemar's confusion matrix (Ch-GPAB versus Chr-PmRF) for Cell2Cell Dataset

	Chr-PmRF +	Chr-PmRF −	Total
Ch-GPAB +	14,590	10,200	24,790
Ch-GPAB −	450	14,760	15,210
Total	15,040	24,960	40,000

Table 11 Mecnemar's confusion matrix (Ch-GPAB versus Chr-PmRF) for Orange dataset

	Chr-PmRF +	Chr-PmRF −	Total
Ch-GPAB +	2610	1300	3910
Ch-GPAB −	22	2620	2642
Total	2632	3920	6552

Table 9 displays the performance comparison of Ch-GPAB with other existing churn prediction methodologies using Cell2Cell dataset. Ch-GPAB obtains enhanced prediction performance on Cell2Cell dataset as well, compared to Naïve Bayes's results [8]. In Verkerke's work, Naïve based approach achieves good predictions results only on Cell2Cell dataset and lacks in attaining good prediction performance on other datasets used in the study, resulting an inconsistent performance. Moreover, in our Ch-GPAB system, 10-fold cross validation is deployed to evaluate the prediction per-

Table 12 Mecnemar's confusion matrix (Ch-GPAB versus RUS-BOOST) for Orange dataset

	RUS-BOOST +	RUS-BOOST −	Total
Ch-GPAB +	787	3125	3910
Ch-GPAB −	11	2629	2642
Total	798	5754	6552

formance, whereas a single random split of the training set is performed to evaluate performance in Verkerke's work. Conclusively, Ch-GPAB shows improved performance of 0.910 AUC and 0.862 AUC for predicting churners from Cell2Cell and Orange dataset, respectively (Tables 8, 9). This is highest prediction performance so far reported on both datasets. Thus, proposed Ch-GPAB churn prediction approach shows promising results in accurately predicting telecom churners with additional intuitive and interpreting capabilities.

5.1 McNemar's statistical test

We have also performed McNemar's statistical test to evaluate the confidence level of prediction performance of proposed Ch-GPAB system. In this regard, we performed the comparison of Ch-GPAB with Chr-PmRF [9] and RUSBoost [8] approaches using both Orange and Cell2Cell datasets. RUSBoost uses random undersampling for coping imbalanced class distribution and then AdaBoost algorithms is applied for classification. Tables 10, 11 show the McNemar's confusion matrices for Orange and Cell2Cell datasets. + sign indicates the truly classified instances, whereas—sign shows misclassified instances. The number of instances correctly predicted by proposed Ch-GPAB but incorrectly classified by Chr-PmRF are 10,200 from Cell2Cell dataset as given in Table 10. Similarly, 1300 instances are correctly classified by Ch-GPAB, which are incorrectly predicted by Chr-PmRF from Orange dataset as given in Table 11. Whereas, there are 450 from Cell2Cell and 22 instances from Orange datasets, which are incorrectly predicted by proposed Ch-GPAB but correctly predicted by Chr-PmRF. In comparison with RUSBOOST, the proposed Ch-GPAB correctly classified 3,125 instances whereas only 11 instances are correctly classified by RUSBOOST which are incorrectly classified by CP-GPAB, as shown in Table 12. Therefore, the proposed CP-GPAB shows good prediction capabilities to accurately identify potential churners.

6 Conclusion

GP-AdaBoost algorithm in combination with PSO is presented as a churn prediction approach (Ch-GPAB) for telecom. PSO based method undersamples majority class instances and develops a balanced training set which extends improved learning to proposed churn predictor, resulting in enhanced prediction performance. Ch-GPAB approach evolves multiple GP programs per class using AdaBoost style boosting technique, which strengthens its classification capabilities. Each GP program acts as a single class classifier and a higher weighted sum of GP programs decides the final prediction of a test instance, extending more confident results. Integrated boosting in GP attains enhanced prediction performance on both of the used telecom datasets. Ch-GPAB attains 0.91 AUC and 0.86 AUC on Cell2Cell and Orange datasets, respectively. Moreover, introducing boosting also saves time because a whole new population is not created for every second GP program evolved for the same class, rather an improve GP program may already be found in the population. Besides, attaining improved prediction performance the proposed Ch-GPAB offers an intuitive interpretation of customer churn behavior. Frequent features selected during GP's evolution can be analyzed for making effective marketing decision. Consequently, the proposed Ch-GPAB churn prediction approach is believed to be beneficial for industrial use.

Acknowledgements This work is supported by the Higher Education Commission of Pakistan (HEC) as per Award No. 20-3408/R&D/HEC/14/233.

References

1. ICT: The World in: ICT Facts and Figures, Geneva (2014)
2. Shin, Y.H., David, C.Y., Hsiu, Y.W.: Applying data mining to telecom churn management. Expert Syst. Appl. **37**, 3665–3675 (2006)
3. Bock, K.W.D., Van den Poel, D.: An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. Expert Syst. Appl. **38**, 12293–12301 (2011)
4. Huang, Y., Kechadi, T.: An effective hybrid learning system for telecommunication churn prediction. Expert Syst. Appl. **40**, 5635–5647 (2013)
5. Pendharkar, P.C.: Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. Expert Syst. Appl. **36**, 6714–6720 (2009)
6. Burez, J., Van den Poel, D.: Handling class imbalance in customer churn prediction. Expert Syst. Appl. **36**, 4626–4636 (2009)
7. Mikel, G., Alberto, F., Edurne, B., Humberto, B., Francisco, H.: A review on ensembles for the class imbalance problem: bagging-boosting- and hybrid-based approaches. IEEE Trans. Syst. Man Cybern. C **42**, 463–484 (2012)
8. Verbeke, W., Dejaeger, K., Martens, D., Hur, J., Baesens, B.: New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. Eur. J. Oper. Res. **218**, 211–229 (2012)
9. Adnan, I., Muhammad, R., Asifullah, K.: Churn prediction in telecom using random forest and PSO based data balancing in combination with various feature selection strategies. Comput. Electr. Eng. **38**, 1808–1819 (2012)
10. Huang, B.Q., Kechadi, T.M., Buckley, B., Kiernan, G., Keogh, E., Rashid, T.: A new feature set with new window techniques for customer churn prediction in land-line telecommunications. Expert Syst. Appl. **37**, 3657–3665 (2010)
11. Huang, B., Buckley, B., Kechadi, T.M.: Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. Expert Syst. Appl. **37**, 3638–3646 (2010)
12. Kisioglu, P., Topcu, Y.I.: Applying Bayesian belief network approach to customer churn analysis: a case study on the telecom industry of Turkey. Expert Syst. Appl. **38**, 7151–7157 (2011)
13. Xu, H., Zhang, Z., Zhang, Y.: Churn prediction in telecom using a hybrid two-phase feature selection method. In: Third International Symposium on Intelligent Information Technology Application, 2009. IITA 2009, pp. 576–579 (2009)

14. Owczarczuk, M.: Churn models for prepaid customers in the cellular telecommunication industry using large data marts. Expert Syst. Appl. **37**, 4710–4712 (2010)
15. Verbeke, W., Martens, D., Mues, C., Baesens, B.: Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Syst. Appl. **38**, 2354–2364 (2011)
16. De Bock, K.W., Van den Poel, D.: Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. Expert Syst. Appl. **39**, 6816–6826 (2012)
17. Espejo, P.G., Ventura, S., Herrera, F.: A survey on the application of genetic programming to classification. IEEE Trans. Syst. Man Cybern. C **40**, 121–144 (2010)
18. Khan, G.M., Arshad, R., Mahmud, S.A., Ullah, F.: Intelligent bandwidth estimation for variable bit rate traffic. IEEE Trans. Evol. Comput. **19**, 151–155 (2015)
19. Bhowan, U., Johnston, M., Mengjie, Z., Xin, Y.: Reusing genetic programming for ensemble selection in classification of unbalanced data. IEEE Trans. Evol. Comput. **18**, 893–908 (2014)
20. Rokach, L.: Ensemble-based classifiers. Artif. Intell. Rev. **33**, 1–39 (2009)
21. Sorokina, D.: Application of additive groves ensemble with multiple counts feature evaluation to KDD Cup ’09 small data set. In: Presented at the JMLR Workshop and Conference Proceedings, Paris (2009)
22. Adnan, I., Asifullah, K., Lee, Y.S.: Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification. Appl. Intell. **39**, 659–672 (2013)
23. Amin, A., Anwar, S., Adnan, A., Nawaz, M., Howard, N., Qadir, J., et al.: Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. IEEE Access **4**, 7940–7957 (2016)
24. Yang, P., Xu, L., Zhou, B.B., Zhang, Z., Zomaya, A.Y.: A particle swarm based hybrid system for imbalanced medical data sampling. BMC Genom. **10**, S34–S34 (2009)
25. Au, W.H., Chan, K.C.C., Xin, Y.: A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Trans. Evol. Comput. **7**, 532–545 (2003)
26. Hadden, J., Tiwari, A., Roy, R., Ruta, D.: Computer assisted customer churn management: state-of-the-art and future trends. Comput. Oper. Res. **34**, 2902–2917 (2007)
27. Wang, P., Emmerich, M., Li, R., Tang, K., Back, T., Yao, X.: Convex Hull-based multiobjective genetic programming for maximizing receiver operating characteristic performance. IEEE Trans. Evol. Comput. **19**, 12 (2015)
28. Langdon, W.B., Harman, M.: Optimizing existing software with genetic programming. IEEE Trans. Evol. Comput. **19**, 118–135 (2014)
29. Lemmens, A., Croux, C.: Bagging and boosting classification trees to predict churn. J. Mark. Res. **43**, 276–286 (2006)
30. Bose, I., Chen, X.: Hybrid models using unsupervised clustering for prediction of customer churn. J. Org. Comput. Electron. Commer. **19**, 133–151 (2009)
31. Lima, E., Mues, C., Baesens, B.: Monitoring and backtesting churn models. Expert Syst. Appl. **38**, 975–982 (2011)
32. Chen, Z.-Y., Fan, Z.-P., Sun, M.: A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. Eur. J. Oper. Res. **223**, 461–472 (2012)
33. Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., Abbasi, U.: Improved churn prediction in telecommunication industry using data mining techniques. Appl. Softw. Comput. **24**, 994–1012 (2014)
34. Kyoungok, K., Chi-Hyuk, J., Jaewook, L.: Improved churn prediction in telecommunication industry by analyzing a large network. Expert Syst. Appl. **41**, 6575–6584 (2014)
35. Ning, L., Hua, L., Jie, L., Guangquan, Z.: A customer churn prediction model in telecom industry using boosting. IEEE Trans. Industr. Inform. **10**, 1659–1665 (2012)
36. Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., et al.: Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing **237**, 242–254 (2017)
37. Amin, A., Khan, C., Ali, I., Anwar, S.: Customer churn prediction in telecommunication industry: with and without counter-example. In: Gelbukh, A., Espinoza, F.C., Galicia-Haro, S.N. (eds.) Proceedings on Nature-Inspired Computation and Machine Learning: 13th Mexican International Conference on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, Mexico, November 16–22, 2014, Part II, pp. 206–218. Springer International Publishing, Cham (2014)
38. Amin, A., Rahim, F., Ramzan, M., Anwar, S.: A prudent based approach for customer churn prediction. In: International Conference: Beyond Databases, Architectures and Structures, pp. 320–332 (2015)
39. Miller, H., Clarke, S., Lane, S., Lonie, A., Lazaridiz, D., Petrovski, S., et al.: Predicting customer behaviour: the University of Melbourne’s KDD Cup report. In: Presented at the JMLR Workshop and Conference Proceedings, Paris (2009)
40. Huang, B., Kechadi, M.T., Buckley, B.: Customer churn prediction in telecommunications. Expert Syst. Appl. **39**, 1414–1425 (2012)
41. Idris, A., Khan, A., Lee, Y.S.: Genetic programming and adaboosting based churn prediction for telecom. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1328–1332 (2012)
42. Busa-Fekete, R., Kegl, B.: Accelerating AdaBoost using UCB. In: Presented at the JMLR Workshop and Conference Proceedings, Paris (2009)

Adnan Idris received his M.S. degree in Computer System Engineering from GIK Institute of Engineering Sciences and Technology Topi, Pakistan in 2006. Prior to that, he has earned his master degree in software engineering from COMSTATS Institute of IT, Islamabad in 2002. In 2015, he completed his Ph.D. from Pakistan Institute of Engineering and Applied Sciences Islamabad. Further, he has more than 15 years of research & teaching experience at university level. Currently he is working as an Assistant Professor in Department of Computer Sciences & IT, The University of Poonch Rawalakot. His research areas include Customer Churn Prediction, Machine Learning and Evolutionary algorithms.

Aksam Iftikhar completed his B.S. degree in computer engineering from University of Engineering and Technology, Lahore in 2007. He received M.S. degree in computer science from the same university in 2010. In 2014, he completed his Ph.D. at Pakistan Institute of Engineering and Applied Sciences, Islamabad. Nowadays he is associated as Assistant Professor with COMSATS Institute of I.T Lahore campus. His primary research areas include Image Processing (especially medical image processing and analysis), Computer Vision, Machine Learning and Pattern Recognition.

Zia ur Rehman has been currently pursuing postdoctoral research from Institute for Molecular Medicines Finland (FIMM) for last three years in the field of computational biology. He has strong expertise in machine learning and databases. In past, he has extensively worked on feature selection, undersampling and oversampling techniques, especially for biological data.