academicJournals

Vol. 8(27), pp. 1289-1295, 18 July, 2013
DOI 10.5897/SRE2013.5559
ISSN 1992-2248 © 2013 Academic Journals
http://www.academicjournals.org/SRE

Scientific Research and Essays

Full Length Research Paper

Customer churn prediction using a hybrid genetic programming approach

Ruba Obiedat[1], Mouhammd Alkasassbeh[2], Hossam Faris[1]* and Osama Harfoushi[1]

[1]Department of Business Information Technology, The University of Jordan, P. O. Box 11942 Amman, Jordan.
[2]Department of Information Technology, Mutah University, P. O. Box 61710, Karak, Jordan.

Accepted 3 July, 2013

A churn consumer can be defined as a customer who transfers from one service provider to another service provider. Recently, business operators have investigated many techniques that identify the customer churn since churn rates leads to serious business loss. In this paper, a hybrid technique has been used which combines K-means clustering with Genetic Programming to predict churners in telecommunication companies. First, K-means clustering is used to filter the training dataset from outliers and non representative customer behaviors then Genetic Programming is applied in order to build classification trees that are able to classify customers into churners and non churners. The proposed approach is evaluated and compared with other common classification approaches. Experimental results show that K-means clustering with Genetic Programming has promising capabilities in predicting churners' rates.

Key words: Churn consumer, churn customer, K-means clustering, Genetic Programming

INTRODUCTION

Recently, mobile telecommunications became the superior communication medium and sharing data between the callers all over the world. Sanou (2013), in the International Telecommunication Union reveals that there were around 6.8 billion mobile subscriptions, which means 96% of the world population. The mobile subscriptions here mean the SIM cards and most of the mobile users have more than one SIM card. In many countries, usually, there is more than one mobile provider who gives different services to attract new customers and keep the exciting ones; every now and then there are latest new mobile phones offers and promotions to keep the company alive and prosperous. At the same time, public regulations and the standardization of mobile communication allow customers to easily move from one provider to another, which is called a *churn consumer* or *churn customer*; as a result, churn prediction has raised a crucial mobile Business Intelligence (BI) application that aims at identifying the customer who is about to leave to a competitor or stay with the same provider. This information is very important for the managers to set-up their plans for the future, technically or commercially. To endure in the stimulating atmosphere of an international market and be competitive, organizations essentially identify and predict customer predilections and behaviors to make the most of customer retention before their rivals do so.

Customer churn management includes three steps: determination and identification of churn customers, investigating the reasons of churn and application of certain policies; as well as taking measures to deteriorate the rate of churn (Rodpysh, 2012). Some of the major data tasks in data mining are prediction and classification, which can be applied to extract knowledge by using the data available regarding customers' behaviors. Nowadays, this technique is used for customers churn management and customers relation management (Rodpysh, 2012). A study carried out by

*Corresponding author. E-mail: 7ossam@gmail.com.

Keramati and Ardabili (2011) recognizes factors that affect customer churn; the single most valued of an organization's possessions. One year's data from a call logs files belonging to 3150 clienteles were designated arbitrarily from an Iranian mobile operator call-center record. The findings from this research specify that a customer's displeasure; their amount of facility usage and certain demographic features have the greatest effect on their decision to stay or churn. The findings also suggest that customer position (active or inactive status) facilitate the association between churn and reason of churn.

Technically, classical Decision Trees and Artificial Neural Networks (ANN) are one of the most common applied techniques in the field and they revealed their capabilities in churn estimation (Tsai and Lu, 2009; Huang et al., 2012). Authors in Tsai and Lu (2009) have conducted research in which they reflect two hybrid models by joining two dissimilar neural network practices for churn expectation, which are back-propagation ANN and self-organizing maps (SOM). The findings demonstrate that the two hybrid models outperform the single neural network baseline model in terms of expectation accurateness over the testing sets.

In another research conducted by Kim and Yoon (2004), the factors of subscriber customer loyalty and churn are recognized in the Korean mobile telephony market, by means of a binomial logit model based on examination of 973 mobile users in Korea. They conclude that the unimportance of subscription extent in touching the loyalty-induced action specifies that lock-in effects are possible to be focused among the 'spuriously loyal' consumers who are not eager to churn just for the reason of switching costs.

A similar study through experimental evaluation had been carried out by Hung et al. (2006). They link various data mining methods that can allocate a propensity-to-churn score occasionally to every subscriber of a mobile operator. The findings indicate that neural network methods can bring precise churn expectation models by means of billing information, customer demographics, call detail records, contract/service status, and service change log.

As marketplaces have become progressively saturated, corporations have accepted that their business policies need to emphasize on recognizing those consumers who are most possible to churn. To expose this problem, a study by Hadden (2008) suggests that consumers can be positioned into one of numerous profiles groups according to their relations with the service provider. Grounded on this, estimation is likely based on when the consumer can probably dismiss his/her service with the firm. Further studies also indicate that in a very good mobile telecommunication business atmosphere, marketing managers require a business intelligence model that permits them to keep the best (at least a near optimal) level of churners very successfully and professionally whereas reducing the charges through their marketing programs (Lee et al., 2011), consumers churn is the principal anxiety of most businesses, which are working in production with little switching cost (Jahromi, 2009). Furthermore, the tendency of consumers to end their affairs with service providers has required several businesses in competitive markets to move their planned focus from consumer achievement to consumer preservation (Seog Kim et al., 2012).

In this paper, we introduce a hybrid approach for predicting churners based on using K-means clustering with Genetic programming (GP). GP is an evolutionary heuristic approach which is inspired by the evolutionary theories. GP has some advantages over some other common modeling approaches like the typical artificial neural network since GP is able to develop simple models which are easy to evaluate and it can assist in analyzing the importance of the involved variables. The proposed approach in this work is conducted in two stages. In the first stage, K-means clustering is used to reduce the training dataset and to remove the unnecessary data. In the second stage, GP is applied to generate for the classification models. Finally, the proposed approach is evaluated and comparative experiments are conducted to compare the results with those obtained by other common approaches (the basic genetic programming, ANN and decision tree algorithm C4.5).

In this paper, K-means clustering and GP are first introduced, followed by the dataset set used. The proposed hybrid method is thereafter presented; the evaluation criteria used to assist the proposed method are listed and then experiments and results are discussed.

K-MEANS CLUSTERING

K-mean clustering algorithm is one of the statistical data mining techniques, based on unsupervised learning method as no predefined classes are given. It takes a large set of elements and separates them according to their features and characteristics into k different groups, each group called a cluster. Intra-cluster distances have to be minimized while the Inter-cluster distances are maximized, which means elements in one cluster have to be similar to each other but dissimilar from elements in other clusters (Han et al., 2006). The algorithm starts by randomly choosing K elements from the data set as the initial centers for each cluster; then the rest of elements are assigned to the cluster to which it is most similar based on (nearest) distance between the elements and the cluster center. The new center for each cluster was then computed using the mean of the current objects in that cluster, and the process repeated until the cluster center does not change. The commonly used distance measurement for clustering is calculated by the following squared error function (Han et al., 2006).

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2 \qquad (1)$$

Where p presents a given training example, m_i the mean of the cluster C_i and k is the number of clusters.

One of the important applications of the K-Means Clustering mentioned in the literature is data simplification. K- Means Clustering can be used to cluster input datasets into group of sets of similar data. Each of the smaller datasets will be used for training a given modeling or classification techniques independently. As reported in the literature, this application can lead to a better learning performances and reduction in training time Faraoun and Boukelif (2006).

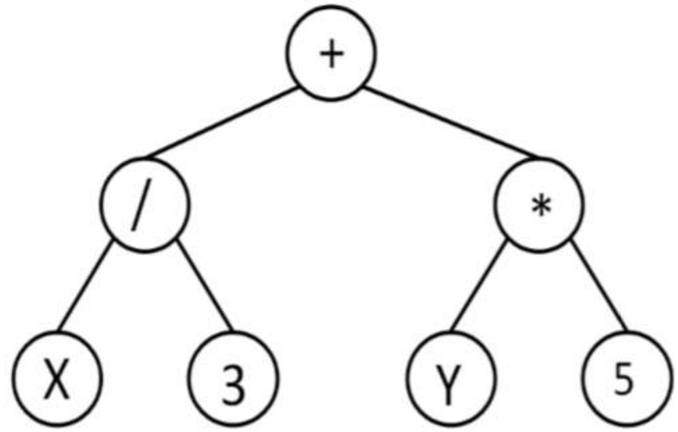

Figure 1. Simple genetic program represented as a tree.

GENETIC PROGRAMMING APPROACH

Genetic Programming (GP) is an evolutionary algorithm-based methodology for automatically solving problems inspired by biological evolution Koza (1991, 1982). GP has many advantages when used to model nonlinear systems Kotanchek et al. (2003). They include:

1) Simple summary models: Models generated by GP are more simple and easier to evaluate compared to other soft computing techniques like ANN and FL.
2) Variables impact analysis: GP is able to identify the significant variables according to their appearance during the evolutionary process.

GP has been applied successfully to a large number of complex problems like statistical modeling, electronic circuitry, pattern recognition, computational finance, and picture generation. The evolutionary process of the GP starts by generating some initial individuals where each individual is a hierarchical computer program and all the individuals form one population. Each program is tree structured and can be seen as graphical representations of so-called S-expressions of the programming language LISP (Affenzeller et al., 2009). Figure 1 shows an example of a very simple genetic program represented as a tree. After generating the initial population, the fitness of each individual in this population is computed. While stopping criterion is not yet reached we do the following:

a) Select individual for reproduction using some selection mechanisms such as tournament, rank, etc.
b) Create an offspring using reproduction operators which make small random changes to the construction of the individuals. Reproduction operators include the following:

i) Crossover refers to producing two new individuals by selecting a random subtree in each of the two parents and swapping the resultant subtrees, with the new individuals being the offspring (Figure 3).
ii) Mutation operates on one individual by replacing a subtree below a random chosen point by a randomly generated subtree (Figure 4). The probability of crossover and mutation are selected based on the application but typically the probability of mutation is much smaller that crossover.

c) Compute the new generation. This process will end either when the optimal solution is found or the maximum number of generations is reached.

By this process the individual programs evolves and have better fitness values by time. The whole process is described as a flow chart in Figure 2. In this work a symbolic regression model is developed by GP to fit the given sample data. J. Koza identifies symbolic regression which is also called (function identification) as “finding a mathematical expression, in symbolic form, that provides a good, best, or perfect fit between a given finite sampling of values of the independent variables and the associated values of the dependent variables.” In our case, the variables are real-valued; therefore the symbolic regression involves finding both the functional form and the numeric coefficients for the model. The mathematical expression as a result can be seen as a computer program generated by the GP evolution that takes the values of the independent variables as input and produces the values of the dependent variables as output. In general, the main goal of GP in symbolic regression is to find a composition of the functions, input variables, and coefficients that minimizes the error of the function with respect to the empirical values (Affenzeller et al., 2009).

METHODOLOGY

Dataset description

The churn dataset of investigation in this research is available in the companion website[1] to "*Discovering Knowledge in Data: An*

[1] http://www.dataminingconsultant.com/DKD.htm

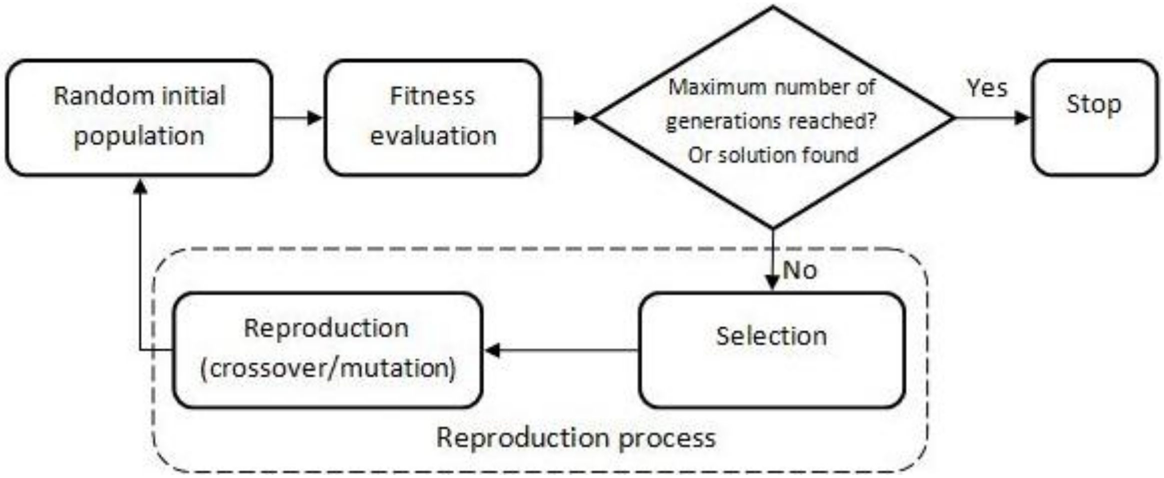

Figure 2. Flow chart of the genetic programming approach.

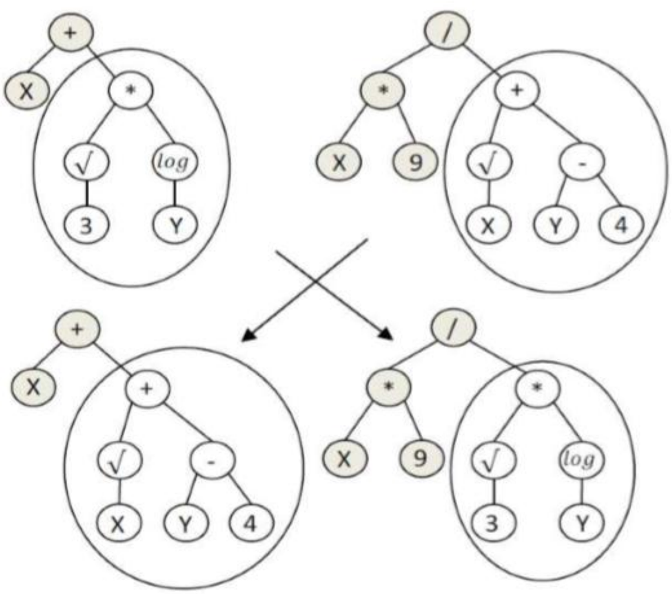

Figure 3. Example of GP crossover operation.

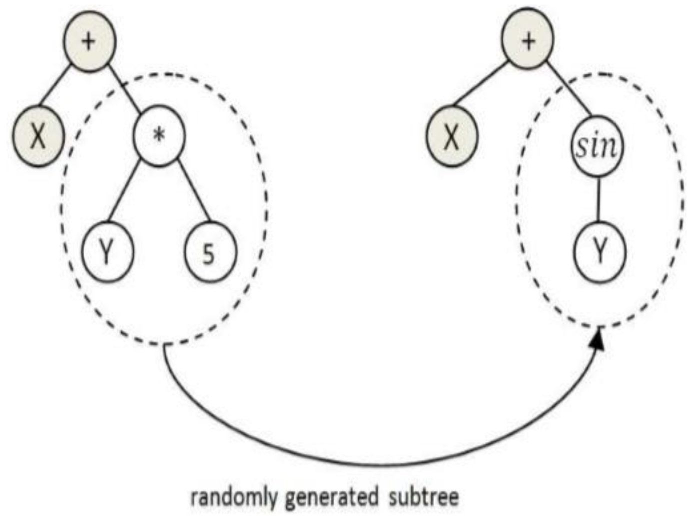

Figure 4. Example of GP mutation operation.

Introduction to Data Mining" by Daniel T. Larose. The data set contains 20 attributes of 3333 customers, along with an indication of whether or not that customer churned (left the company). The total number of churners is 483.

• State: categorical variable, for the 50 states and the district of Columbia.
• Account length: integer-valued variable for how long account has been active.
• Area code: categorical variable.
• Phone number: essentially a surrogate key for customer identification.
• International Plan: dichotomous categorical having yes or no value voice.
• Mail Plan: dichotomous categorical variable having yes or no value.
• Number of voice mail messages: integer-valued variable.
• Total day minutes: continuous variable for number of minutes customer has used the service during the day.
• Total day calls: integer-valued variable.
• Total day charge: continuous variable based on foregoing two variables.
• Total evening minutes: continuous variable for minutes customer has used the service during the evening.
• Total evening calls: integer-valued variable.
• Total evening charge: continuous variable based on previous two variables.
• Total night minutes: continuous variable for storing minutes the customer has used the service during the night.
• Total night calls: integer-valued variable.
• Total night charge: continuous variable based on fore-going two variables.
• Total international minutes: continuous variable for minutes customer has used service to make international calls.
• Total international calls: integer-valued variable.

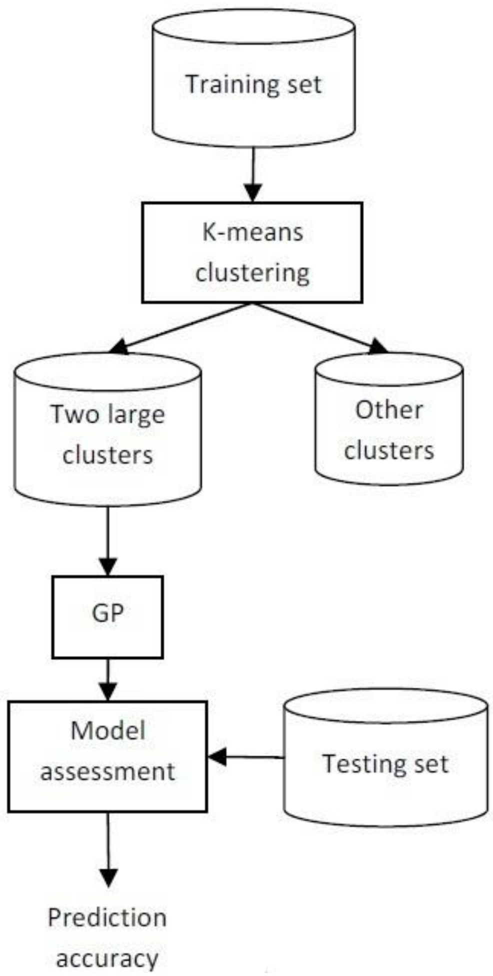

Figure 5. The process of using K-means clustering and GP for predicting churn.

Model development

The model developed in this research is based on using K-means clustering with GP in two stages. At the first stage, data reduction is performed by applying K-means algorithm on the selected training dataset. This process will split the training dataset into a number of smaller sets (clusters). Two clusters with highest number of churners and non-churners are selected for the second stage. This selection method was presented in Du and Zhang (2008). At the second stage, GP is applied on the selected clusters in order to develop the final classification model. Finally, the developed GP model is assisted using left aside dataset. The overall process of the proposed approach is depicted in Figure 5.

Evaluation criteria

In order to evaluate the developed model, we refer to the confusion matrix shown in Table 1 which is the primary source for accuracy estimation in classification problems. Two criteria are used based on the confusion matrix. First one is the accuracy rate shown in Equation 2 which identifies the percentage of the total number of predictions that were correctly classified. The second criteria is the actual churners rate which identifies the percentage of churners that were correctly identified

Table 1. Confusion matrix.

Actual / Predicted	Non-churners	Churners
Non-churners	A	B
Churners	C	D

Table 2. GP parameters.

Parameter	Value
Mutation probability	15%
Population size	1000
Maximum generations	100
Selection mechanism	Tournament selector
Operators	{+ ,- , *,>,<,AND,OR,NOT,If Then Else}

$$Accuracy = \frac{A+D}{A+B+C+D} \quad (2)$$

$$Actual\ churners\ rate = \frac{D}{B+D} \quad (3)$$

EXPERIMENTS AND RESULTS

In order to give a better indication of how well the developed classification model will do when it is asked to classify new data, three folds cross validation is applied. In this process, the selected data set is split into three mutually exclusive subsets, and then each subset is used as testing while the rest is used as training. The process is repeated three times. Eventually, the entire data set is used for both training and testing. In the first stage of the experiments, K-means clustering is applied on each training dataset using WEKA[2]. Different number of clusters where applied in order to select the best number of clusters (that is, 4, 5 and 6 respectively). It was found that clustering the dataset into four clusters leads to better performance. Two clusters among the four which have the highest number of churners, and non-churners are selected and the other two are discarded. The discarded clusters are expected to have outliers of churner and non churner customers. In the second stage, the selected clusters in the previous stage are loaded into Heuristiclab framework[3] as a training set; then classification via GP is applied with parameters set as shown in Table 2. Accuracy and actual churners' rate

[2] WEKA is an open-source software implemented in JAVA, developed at the University of Waikato, New Zealand. WEKA provides a set of machine learning algorithms for data mining tasks. http://www.cs.waikato.ac.nz/ml/weka/

[3] HeuristicLab is a framework for heuristic and evolutionary algorithms that is developed by members of the Heuristic and Evolutionary Algorithms Laboratory (HEAL), http://dev.heuristiclab.com

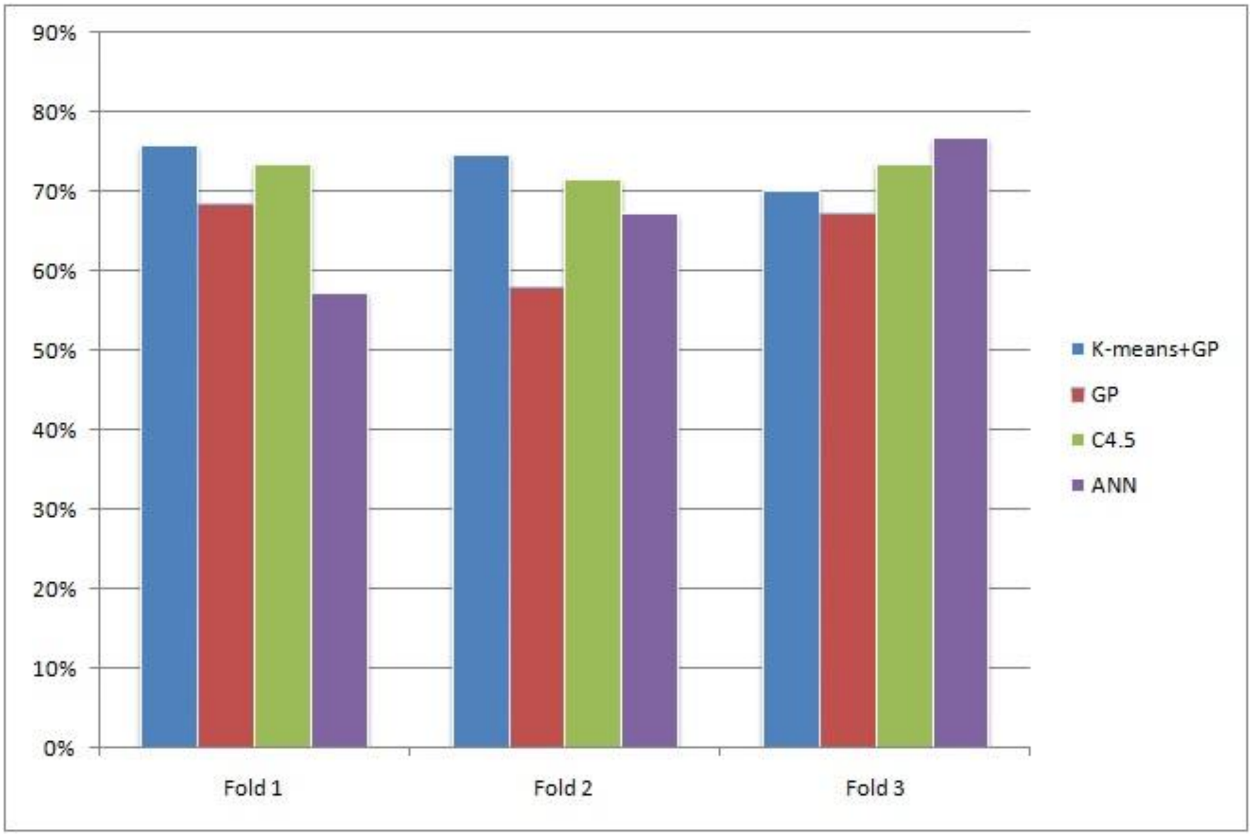

Figure 6. Actual churners' rates for K-means+GP, GP, ANN and C4.5 algorithm.

Table 3. Prediction results for GP and K-means+G.

Parameters	Fold 1		Fold 2		Fold 3	
	Accuracy (%)	Actual churners rate (%)	Accuracy (%)	Actual churners rate (%)	Accuracy (%)	Actual churners rate (%)
K-means+GP	8.67	75.80	82.30	74.60	91.40	70.20
GP	92.70	68.50	92.60	58.00	92.10	67.40
C4.5	94.10	73.60	94.10	71.70	94.10	73.60
ANN	92.20	57.30	93.50	67.40	94.80	76.90

were calculated for GP classification trees in each cross validation fold (Figure 6). Results are compared to those obtained using basic GP, ANN with backpropagation and C4.5 algorithm in Table 3. WEKA was also used for the later two techniques. Training parameters were set for the ANN as follow: learning rate =0.2, epochs = 1000 and number of neurons in the hidden layer =9.

Although the accuracy rate for the K-means with GP is a little bit lower than the other approaches, the actual churners' rate for the proposed approach shows significant improvement over the basic GP with 7, 16 and 5% for the folds 1, 2 and 3 respectively. A disadvantage of the accuracy rate is that it does not give any information about how well non-churners and churners are classified. It just measures the overall accuracy.

Therefore, using actual churners' rate gives more information since churner customers lead to business loss for telecommunication companies. Moreover, results in Table 2 show that K-means with GP performed better than ANN and C4.5 algorithm for first and second folds.

Finally, we can conclude that using K-means clustering technique in order to filter the GP training dataset from unusual data for some customer behaviors, can lead to better churner prediction rates.

Conclusion

Churn prediction is very important for any telephone provider in the competitive market. The main goal of churn management is to predict possible churners and to take proactive actions before the customer leaves the enterprise which affects the companies' profit directly. Therefore, building a very accurate customer churn prediction model is an important issue for any mobile provider. In this work, a new hybrid GP based approach was proposed for predicting churners' rates. The proposed approach is conducted in two stages; in the first stage the K-means algorithm is used to reduce the data set and remove the unnecessary data. In the second stage, GP is used to build classification trees. The experimental results indicate that the K- means with GP model outperform the basic GP model in term of prediction accuracy.

REFERENCES

Affenzeller M, Winkler S, Wagner S, Beham A (2009). Genetic Algorithms and Genetic programming-Modern Concepts and Practical Applications. CRC Press.

Du H, Zhang N (2008). Application of evolving Takagi-Sugeno fuzzy model to nonlinear system identification. Appl. Soft Comput. 8(1):676–686.

Faraoun K, Boukelif A (2006). Neural networks learning improvement using the K-means clustering algorithm to detect network intrusions. Int. J. Comput. Intell. 3(2):161–168.

Hadden J (2008). A Customer Profiling Methodology for Churn Prediction. PhD Thesis, Cranfield University.

Han J, Kamber M, Pei J (2006). Data Mining, Second Edition: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems Elsevier Science.

Huang B, Kechadi MT, Buckley B (2012). Customer churn prediction in telecommunications. Expert Syst. Appl. 39(1):1414–1425.

Hung SY, Yen DC, Wang HY (2006). Applying data mining to telecom churn management. Expert Syst. Appl. 31(3):515–524.

Jahromi AT (2009). Predicting Customer Churn in telecommunications Service Providers. Lulea University of Technology, Sweden.

Keramati A, Ardabili S (2011). Churn analysis for an Iranian mobile operator. Telecommun. Pol. 35(4):344–356.

Kim HS, Yoon CH (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommun. Pol. 28(9):751–765.

Kotanchek M, Smits G, Kordon A (2003). Industrial Strength Genetic Programming. In: R.L. Riolo and B. Worzel, eds. Genetic Programming Theory and Practice. Kluwer, Chapter 15, pages 239–256.

Koza J (1991). Evolving a Computer Program to Generate Random Numbers Using the Genetic Programming Paradigm. In: Proceedings of the Fourth International Conference on Genetic Algorithms Morgan Kaufmann. La Jolla.CA.

Koza JR (1982). Genetic Programming. MIT Press.

Lee H, Lee Y, Cho H, Im K, Kim YS (2011). “Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model,” Decision Support Systems. 52(1):207–216.

Rodpysh KV (2012). Model to Predict the Behavior of Customers Churn at the Industry. Int. J. Comput. Appl. 49(15):12–16.

Sanou B (2013). ICT Facts and Figures. International Telecommunications Union.

Seog Kim Y, Lee H, Johnson J (2012). Churn management optimization with controllable marketing variables and associated management costs. Expert Syst. Appl. 40(6):2198–2207.

Tsai CF, Lu YH (2009). Customer churn prediction by hybrid neural networks. Expert Syst. Appl. 36(10):12547–12553.