Volume 2, Number 4, 2013

International Journal of Research in Industrial Engineering

journal homepage: www.nvlscience.com/index.php/ijrie

A Novel Genetic Algorithm Based Method for Building Accurate and Comprehensible Churn Prediction Models

H. Abbasimehr[*], S. Alizadeh
Department of Industrial Engineering, K.N.Toosi University of Technology, Tehran, Iran

A R T I C L E I N F O

Article history :
Received:
September 9, 2013
Revised:
December 10, 2013
Accepted:
December 28, 2013

Keywords :
Data mining,
Churn prediction,
Feature selection,
Genetic algorithm.

A B S T R A C T

Customer churn has become a critical problem for all companies in particular for those that are operating in service-based industries such as telecommunication industry. Data mining techniques have been used for constructing churn prediction models. Past research in churn prediction context have mainly focused on the accuracy aspect of the constructed churn models. However, in addition to the accuracy, comprehensibility aspect should be considered in evaluating a churn prediction model. Being comprehensible, a model can reveal the main reasons for customer churn; thereby mangers can use such information for effective decisions making about marketing actions. In this paper, we demonstrate the application of a genetic-algorithm (GA) method for building accurate and comprehensible churn prediction model. The proposed method, GA-based method uses a wrapper based feature selection approach for choosing the best feature subset. The key advantage of this method, is taking into account the comprehensibility measure (measured as the number of rules extracted from C4.5 decision tree) in evaluating the performance of a candidate model. The GA-based method is compared to the two filter feature selection methods including Chi-squared based and Correlation based feature selection using two telecommunication churn datasets. The results of experiments indicated that the GA-based method performs better than the two filter methods in terms of both accuracy and comprehensibility

1. Introduction

Today, companies operating in service-based industries such as telecommunication, banking incur more lost from customer churn. Customer churn is defined as "the tendency of customers to stop doing business with a company in a given time period" [1]. Identification of customers who are at risk of churn is of critical importance in all industries especially in those that there is low switching cost for customers to leave the company. Churn prediction is a useful tool for identifying customers who tend to stop doing business and thereby preventing them from churning by implementing marketing programs. As expressed in [2], churn prediction is a binary classification task, which intends to build a classifier to distinct between churners and non-churners.

[*]Corresponding author
E-mail address: h.abbasimehr@mail.kntu.ac.ir

Various data mining techniques and in particular classification algorithms have been used for churn prediction including support vector machines (SVM) [3, 4], decision tree [5, 6], artificial neural network (ANN) [7, 8], Logistic regression [9, 10].

Previous research in the field of churn prediction paid much attention to the accuracy of the churn models. However, comprehensibility of a churn prediction model is another important aspect since comprehensibility of model facilitates using the model to make decisions about marketing program of the firm [11]. Therefore, in this paper, we pay attention to the both accuracy and comprehensibility aspects.

Data preprocessing and especially feature (attribute) selection is an important step in a mining process [12], since by preprocessing data, the quality of data and subsequently mining results are improved.

The comprehensibility problem is related to the choice of data mining technique; for instance decision tree generates transparent and comprehensible whereas produce a black-box model that is the model is not transparent. In addition, the size of a model is an indicator of its comprehensibility. The size of model is also dependent on the parameter of the algorithm and number of features.

To this aim, we firstly use a modeling technique that produces a comprehensible model. Decision tree is selected as modeling technique through this paper. Decision tree algorithm is a widely used classification techniques that takes a training samples and construct a decision tree by adopting divide and conquer strategy [13]. By traversing a full decision tree, a set of If-then rules is generated. Depending on the algorithm' parameters and selected features, some characteristics of tree related to comprehensibility including size of tree, number of leaves (rules) may be varied. Number of rules is used as the comprehensibility metric in this paper.

To reduce the number of rules generated from a tree, two main actions should be performed:

1- Setting the parameters of the algorithm in order to generate a small-sized model.

2- Selecting the best features that are informative.

In data mining, feature selection approaches are divided into two categories including filter and wrapper. Thus, another aim of this study is to investigate the effectiveness of filters and wrappers feature selection approaches in selecting the best features in order to build accurate and comprehensible churn prediction model.

The proposed method for building accurate and comprehensible churn prediction model uses a wrapper based feature selection approach for choosing the best feature subset. Furthermore, in the proposed method, comprehensibility (measured as the number of rules extracted from decision tree) is considered as an influencing factor in the fitness function. In other words, the performance of a model generated by using a candidate feature set is evaluated by both accuracy and comprehensibility view point. Therefore, the resulted model can be more comprehensible and provides insights about the churn drivers. Our method is applied on two churn dataset from telecommunication industry. The results indicate that our method leads to the accurate and comprehensible churn prediction models comparing to the other methods used.

The reminder of this paper is organized as follows. Firstly, a background about feature selection and churn prediction is presented in Section 2. The proposed method is described in Section 3. In Section 4, the datasets and experimental setup are described; besides the results of experiments are analyzed. Conclusions and future works are considered in Section 5.

2. Background

2.1.Feature Selection

Real-life datasets contain many features which many of those features (attributes) may be irrelevant or redundant. Including the irrelevant or redundant features causes the classification algorithm to be confused and thereby the results may be unsatisfactory [12] Feature selection is a well-known way for detecting irrelevant or redundant attributes [12, 13]. Benefits of employing feature selection approach are: enhancing the predictive ability of a classifier, gaining a faster and computationally effective classifier, and providing an easy to understand classification model [14]. Therefore, one main benefit of using feature selection methods in classification tasks is improving the comprehensibility of the generated models.

In general, feature selection methods fall into two main categories including filters and wrappers [14, 15]. The main difference between filters and wrappers is that filter methods evaluate a subset of features based on general characteristic of data. In the filter methods, candidate features are ranked based on certain measures of relevance prior to learning step. In other words, after the feature set is filtered according to a measure as a preprocessing step, the learning starts. In filter methods, several different measures of relevance like distance, information, dependency, and consistency are used; but as stated in [13] there is not any commonly accepted measure. Correlation-based, Chi-Squared based, consistency-based, Information gain, Relief, Symmetrical uncertainty are frequently used filter methods [13].

In wrapper methods, the learning algorithm is wrapped into the feature selection process so that the usefulness of the subset is evaluated by employing a learning algorithm [15]. In other words, these methods select an optimal subset based on performance measures resulted by a classifier created with that feature subset. Wrapper methods require a learning algorithm, an accuracy measure, and a search strategy [14, 15].

2.2.Churn prediction

Due to the fierce competition between firms almost in all industries, customer churn become a critical problem. While almost all companies incur customer churn, the telecommunication industry is the best example that experiences a higher customer churn rate. The mobile telecommunication segment undergoes an average of 20-40 percent annual churn so the profitability loss is much higher in this industry [16].

Generally, research in the churn domain fall into two categories: researches that try to identify influential factors on customer churn, building churn prediction models [9]. Our study is related to the second category which proposes a method for constructing churn prediction models. Hence, we review some research from the second category.

Research that attempt to build churn prediction models mainly focus on using prediction techniques with high performance or extracting new information about customer which by including that information the predictive performance of a modeling technique could be increased. Data mining, machine learning and soft computing techniques are used widely in churn prediction context.

In [6] C4.5 decision tree was utilized for predicting churn since this technique can generate interpretable knowledge in an understandable (comprehensible) form. Furthermore, in response to the problem of customers' demographic data unavailability in telecommunication sector, churn prediction was done based on customers' usage data extracted from call details. The new approach is capable of identifying churners better than demographic based churn prediction approaches [6].

Lemmens and Croux [10] applied bagging and boosting classification techniques to predict customer churn in wireless telecommunication. They proved that these techniques improve accuracy in predicting churn.

Coussement & Van den Poel [3] applied support vector machines (SVM) in newspaper subscription context. Their results indicated that SVM gained higher performance compared to logistic regression. Furthermore, their results showed that random forest technique outperforms SVM. Pendharkar [7] developed two genetic algorithm based neural network models for customer churn prediction in wireless network service. The two developed models perform better than statistical z-score model in terms of performance criteria. In addition, neuro-fuzzy techniques such as adaptive neuro-fuzzy inference system (ANFIS) and locally linear model tree (Lolimot) were used for churn prediction purpose [17, 18].

Two novel methods called AntMiner+ and ALBA developed in [11] to build accurate and comprehensible churn prediction models. AntMiner+ is a data mining technique based on the principal of Ant Colony optimization. ALBA is a rule induction technique that combines concepts of the SVM and rule learners such as C4.5 and RIPPER [11]. Both ALBA and AntMiner+ lead to comprehensible models with improved accuracy when applied to the churn dataset.

Real churn datasets suffers from class imbalance problem that is the proportion of churners to non-churners is small. As a result, modeling techniques faces with difficulties in discriminating between churners and non-churners. To deal with imbalance problem existed in churn data, four possible solutions were presented in [19]. Using appropriate evaluation metrics, applying cost-sensitive learning methods, employing sampling techniques including basic sampling (e.g. under-sampling and oversampling) and advanced sampling techniques such as SMOTE and using Boosting technique as the four possible solutions were tested using six real-life churn data sets [19].

A new learning method, called improved balanced random forest (IRBF) were developed in [20] which showed higher performance in comparison with other algorithms including neural network, SVM , and decision trees. A method called extended support vector machine (ESVM) was proposed in [4] to tackle the imbalance and nonlinear of customer churn. The proposed method applied on a churn data from an ecommerce website which demonstrated better performance than ANN, decision tree, and SVM [4].

Data used for churn prediction purpose usually extracted from structured marketing database such as customer socio-demographic, usage information, customer /company interaction however, Coussement & Van den Poel [9] used unstructured data from call center emails. They found that including textual email information improve predictive performance. Besides incorporation of network-based information for example features extracted from customers' social networks for churn predictive model building was performed in [17, 18, 21].

Some of recent researches in the churn prediction area are summarized in Table 1. The main contributions of each work as well as the domain are highlighted in Table 1. As illustrated in the table, customer churn is the focal concern of many domains especially in telecommunication, financial.

Table 1. An overview of churn prediction research

Reference	**Contribution**	**Domain**
Coussement, K., Van den Poel, D.[3]	Application of Support Vector Machines(SVM) in order to construct churn model	Newspaper publishing
Burez, J., Van den Poel, D.[22]	Defining two distinct types of churn :Commercial and financial	Pay-TV
Coussement, K., Van den Poel, [9]	Developing a higher predictive performance churn model by adding textual information obtained customers' emails to call center	Newspaper publishing
Pendharkar, P. C.[7]	proposing two genetic-algorithm (GA) based neural network (NN) models for churn prediction	cellular wireless network
Xie, Y., Li, X., Ngai, E. W. T., Ying, W. [20]	Developing and applying a new algorithm called improved balanced random forests (IBRF) for churn prediction	Bank
Tsai, C.-F., Lu, Y.-H .[8]	constructing two hybrid models using two different neural network techniques to predict churn	telecom
Glady, N., Baesens, B., Croux, C. [23]	Proposing a solution to detect the churner customers based on the customer lifetime value(CLV)	Financial service company
Huang, B., Buckley, B., Kechadi, T. M. [5]	Using NSGA-II for feature selection in churn prediction context	telecom
Tsai, C.-F., & Lu, Y.-H. [24]	Application of association rules for selecting important variable in churn prediction	Multimedia on demand
Yu, X., Guo, S., Guo, J., Huang, X. [4]	Presenting An extended support vector machine (ESVM) to address the imbalance and nonlinear of customer churn.	ecommerce
Verbeke, W., Martens, D., Mues, C., &Baesens, B. [11]	Developing two novel data mining algorithms, AntMiner+ and ALBA that induce accurate and comprehensible rules in churn prediction context	telecom
Verbeke, W., Dejaeger, K., Martens, D., Hur, J., Baesens, B. [25]	Developing a novel profit-based metric to assess the performance of churn prediction models. Also conducting an extensive experiments to test performance of various classification techniques using eleven real-life datasets	telecom
Idris, A., Rizwan, M., Khan, A. [26]	Presenting an undersampling method based on Particle Swarm Optimization (PSO) to balance data distribution of telecom churn dataset	telecom
Abbasimehr, H., Setak, M., Soroor, J. [17]	Identification of High-value customers along with employing neuro-fuzzy techniques to predict churn of those customers	telecom
Haenlein, M. [21]	Investigation of the effect of social interaction on customer churn behavior using call detail record	telecom

3. Proposed method

The proposed method used in this paper for building accurate and comprehensible churn prediction models is shown in Figure 1. As mentioned before, the aim of the proposed method is finding the best feature set so that the classification model built using that feature set become acceptable from both accuracy and comprehensibility measures which are two important aspects of a churn models emphasized in the past research (e.g. [11]). The proposed method takes a dataset (in this study a churn dataset) as input and uses it for building classifiers during its running. Usually, in a classification task, the input dataset is divided into train and test datasets, the training part used for building the classifier and the test part is used for assessing the performance of the model concerning evaluation metrics. Therefore, the input dataset is divided into two separate train and test datasets using the hold-out split [12]. Actually, the method is composed of two main components: the GA part and the evaluation part. The GA part generates candidate feature sets using genetic algorithm(GA) [27] and the evaluation part assesses the quality of each feature set through building classifier using train set extracted dynamically based on that feature set.

3.1.GA Part:

The GA Part of the method firstly produces an initial GA population. Each member of initial population is evaluated based on their cost function by evaluation part.

3.1.1. Chromosome representation

For a problem with n features, we consider one bit for each feature; 1 means that the feature is selected and 0 means it is not selected. Therefore, we have an n-bit representation for all chromosomes (feature subsets) where each bit represents the omission or inclusion of the associated feature.

For example, considering a problem with four features, a member of population is illustrated in Figure 2. As seen from the Figure 2, this member of population only contains two features F_1 and F_4.

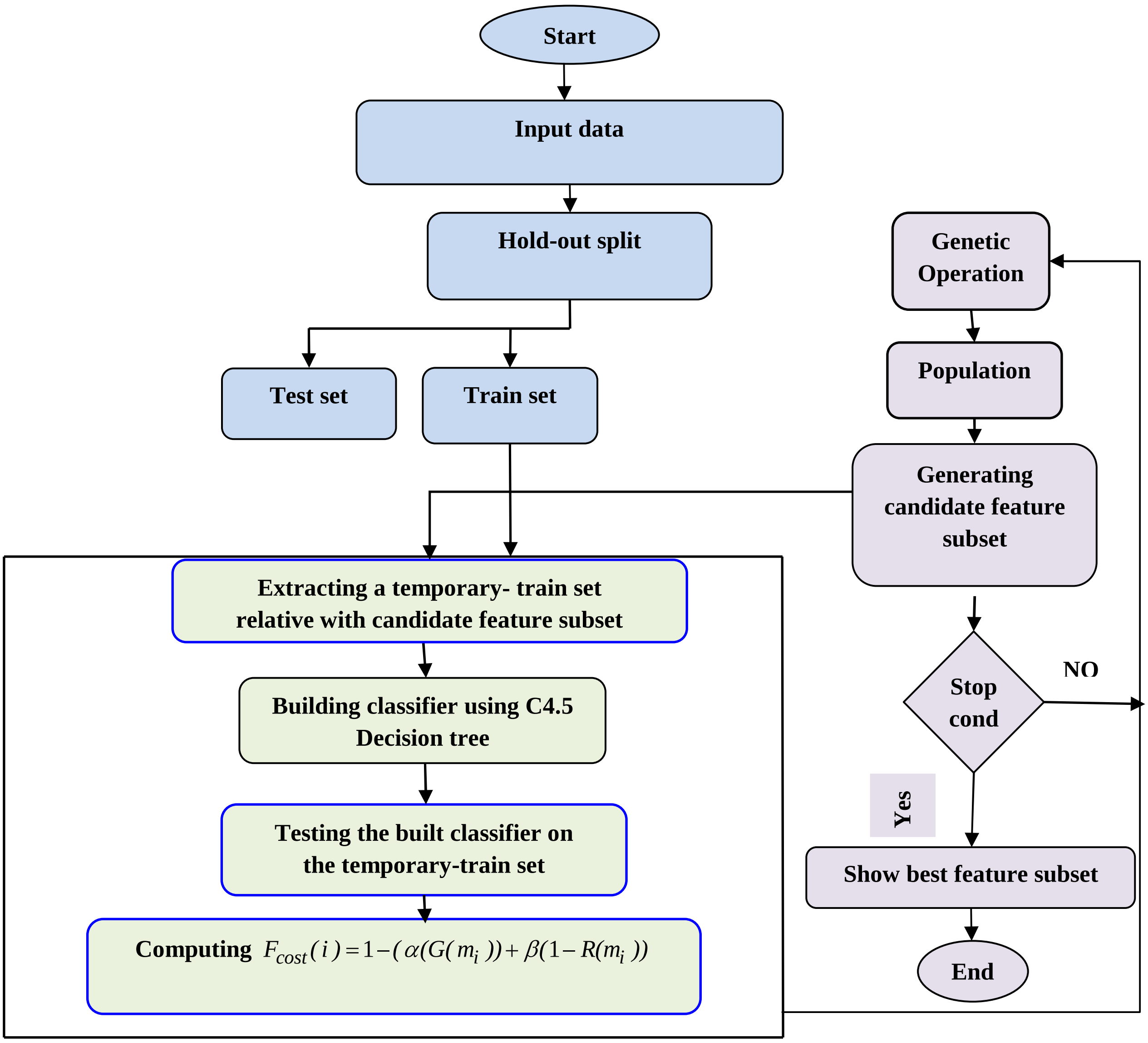

Figure 1. The procedure of the GA-based method

F_1	F_2	F_3	F_4
1	0	0	1

Figure 2: representation of a choromosome

3.1.2. Cost Function

As mentioned previously, the proposed method aims at building accurate and comprehensible churn prediction models. Churn datasets suffer from class imbalance. For example, the number of instances of churner class is much less than that of non-churner class. So, we should choose an appropriate cost function for evaluating the merit of each feature set. Several measures can be used as cost function. For example, the error rate (1-accuracy) can be considered as a cost function. However, in this study, based on our aim, we used a cost function that is a combination of both accuracy and comprehensibility in terms of Geometric mean (G) and number of generated rules (R) respectively. Geometric mean, which is defined as $G = \sqrt{Specificity \times Sensitivity}$ [28] is a good metric to assess the performance of a classifier in applications suffer from class imbalance problem such as churn prediction. Geometric mean (G) is measure that contains both sensitivity (true positive rate) and specificity (true negative rate) [12].

Thus, the cost function is defined as:

$$F_{cost}(i) = 1-(\alpha(G(m_i)) + \beta(1-R(m_i)) \quad , \quad \alpha+\beta=1 \ , \quad 0 \le G(m_i) \le 1 \ , \ 0 \le \mathrm{R}(m_i) \le 1 \qquad (1)$$

where, feature set i, is a member of current population, m_i is the model built using feature set i, and $G(m_i)$ is the geometric mean of m_i and $R(m_i)$ is the normalized number of rules generated by m_i.

$R(m_i)$ is computed using the min-max normalization method [12] as follows.

$$R(m_i) = \frac{r - min\_rule}{max\_rule - min\_rule} \qquad (2)$$

where r is the number of induces rules using classifier.

3.1.3. Selection, Cross over and mutation

Genetic algorithms (GA) rely on the use of three main operators including selection, cross over and mutation [27].

Selection is the process by which two parents are chosen from population based on their cost for cross over. Roulette wheel selection is one of the commonly used selection techniques in GA. In this study we have utilized this technique [27].

Cross over is the process that takes two parents from population and generate a child from them. There are various cross over techniques that have been mentioned in literature. In this paper, we have employed uniform cross over [27].

Mutation is a GA operator that prevents the algorithm from falling into local minimum. The Flip bit mutation operator which takes a genome and inverts the bits was used in this paper [27].

3.2.Evaluation Part

The procedure used by evaluation part for assessment of each member generated in the GA-Part is as follows:

1- A temporary train set related to the member (candidate feature set) is extracted from the original train set.

2- A classification technique is employed in order to build a model. Here we use the c4.5 decision tree due to its high performance. Furthermore, C4.5 decision tree reveals the model

3- The built model is tested on the temporary train set so that the cost function is computed for that candidate feature set. As demonstrated in Figure 1, the cost function used in this study considers both accuracy and comprehensibility when assessing the merit of each

candidate feature set. In the next section we will provide more information regarding cost function and its associated parameters.
The value of cost function associated with each member is returned to the GA part of algorithm in order to generate new candidate feature sets through genetic operations including selection, crossover, and mutation. The algorithm stops when the defined conditions are met.

4. Empirical Analysis and Results

For testing the usefulness of the proposed procedure, we have compared it with two well-known filter feature selection techniques including correlation-based feature selection (CFS), Chi-squared based feature selection. CFS selects a feature subset by considering the predictive value of each feature individually, along with the degree of redundancy among them (Witten & Frank, 2005). Chi-squared based feature selection assesses attributes by computing the chi-squared statistic with respect to the class.
As mentioned earlier, for carrying out experiments, we used two churn datasets, Telecom1 and Telecom2. These datasets are described in the following subsection.

4.1. Data set description

In this research, we have used two churn datasets including Telecom1 and Telecom2. These datasets come from the telecommunication industry. Telecom1 is a telecom dataset used in the Churn Tournament 2003, organized by Duke University. This database contains datasets of mature subscribers (i.e. customers who were with the company for at least six months) from a major U.S. Telecommunication Company. For Telecom1, any features having too many missing values were omitted. Therefore, 94 features remained for analysis. Telecom2 is another churn dataset obtained from Duke University website (The Center for Customer Relationship Management). This dataset provided by Cell2Cell Company, a major wireless company in US. Telecom2 dataset contains 75 features.
The characteristic of each dataset is presented in Table 2. For each dataset, a sample of 10000 customers was selected with the same proportion of churners and non-churners (50%/50%). This sample is used as the training set. Besides, for each dataset a second sample of 5000 customers was selected as the test set. For Telecom1 and Telecom2, the percentages of churners are 1.8% and 2% respectively. In other words, in each test set the percentage of churners is equal to the real churn rate.

Table 2: Description of datasets

Dataset	**Case description**	**# Instances**	**Churn (%)**	**# features**	**#Train**	**#Test**
Telecom1	Telecom1 is a telecom dataset used in the Churn Tournament 2003, organized by Duke University	15000	1.8	93	10000	5000
Telecom2	Telecom2 is a dataset provided by Cell2Cell Company	15000	2	75	10000	5100

4.2.Implementation details

In this study, the filter methods (Chi-squared, and CFS) were executed using the default parameters of Weka 3.6.4 (Witten & Frank, 2005). The GA-based method was implemented in Matlab version 7.10. Also, the classes of Weka 3.6.4 (Witten & Frank, 2005) were used in Matlab for the implementation of the evaluation component. It is worth to note that the evaluation component takes a feature subset and extracts a temporary training set relative to that feature set. Then it builds a classifier using a classification technique such as C4.5. Table 3 shows the value of parameters during the experiments. The parameters min_rule and max_rule are set to 5 and 50 respectively. Furthermore, we set $\alpha = 0.9$ and $\beta = 0.1$.

As illustrated in Figure 1, the C4.5 decision tree is used as a classifier. To obtain a high-performance model we need to adjust C4.5 parameters according to the characteristics of each dataset. For Telecom1 the parameters of C4.5 are as follows:
“confidence factor”=0.001 and “minimum number of instances per leaf”=120
For Telecom2 the parameters of C4.5 are as follows:
“confidence factor”=0.001 and “minimum number of instances per leaf”=12.

Table 3: Values of the Parameters

Parameters Values		
Cost function	**GA**	**C4.5**
α =0.9	Population size = 100	Confidence factor=0.001
β =0.1	Pcrossover = 0.8	For Telecom1: minimum number of instances per leaf=120
$min\_rule = 5$	Pmutation= 0.3	For Telecom2: minimum number of instances per leaf =12
$max\_rule = 50$	Number of Iterations = 20	

4.3.Experimental results and analysis

The results of experiments are illustrated in the Tables 4 and 5. As seen in the tables, the performance of models were evaluated considering the accuracy (Acc), sensitivity (Sens), specificity (Spec), G (Geometric mean) and the number of generated rules (R).

Table 4: Performance of different approaches on telecom1

method	#features	Acc	Sens	Spec	G	R
Original	92	0.6848	0.4000	0.69	0.5261	33
Chi -squared	86	0.6026	0.5556	0.6035	0.5790	31
CFS	16	0.6932	0.4333	0.6980	0.5500	17
GA-based	54	0.6904	0.5000	0.6939	**0.5890**	**12**

Table 5: Performance of different approaches on telecom2

method	#features	Acc	Sens	Spec	G	R
Original	74	0.4376	0.7300	0.4316	0.5613	6
Chi -squared	32	0.4376	0.7300	0.4316	0.5613	6
CFS	7	0.3480	0.7900	0.3390	0.5175	3
GA-based	32	0.5912	0.6100	0.5908	**0.6003**	12

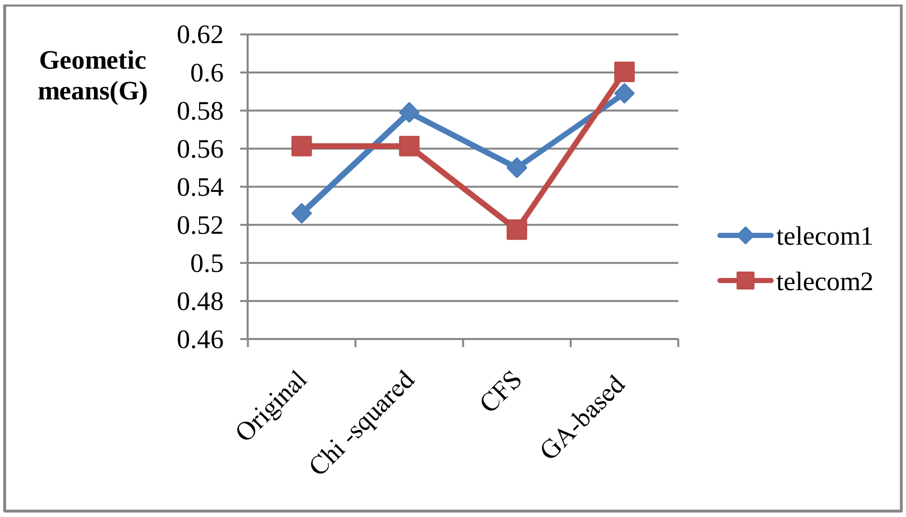

Figure 3: performance of used methods in terms of Geometric means (G)

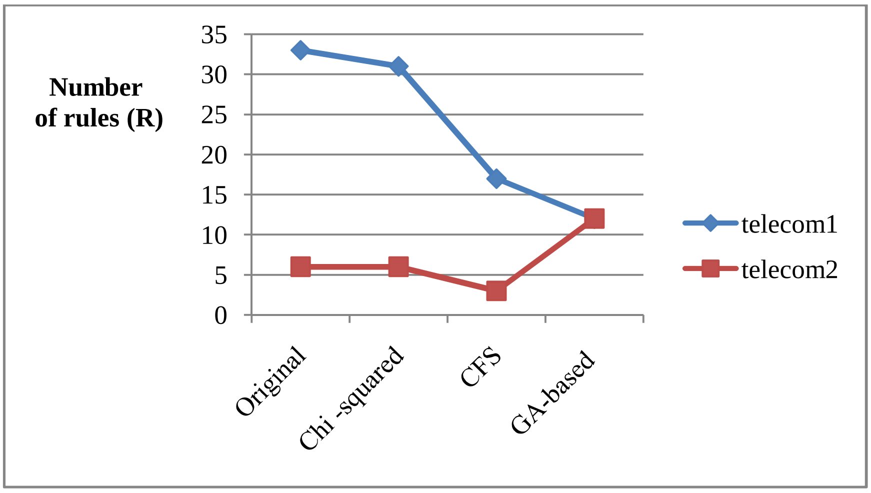

Figure 4: Performance of used methods in terms of number of generated rules(R)

A useful churn prediction model should balance between specificity and sensitivity measures. Therefore, we consider geometric mean (G) to assess the accuracy of the resulted models. In general the smaller the number of generated rules (R) the better the comprehensibility is. Figures 3 and 4 show the performance of utilized methods in terms of geometric mean (G) and generated rules (R) respectively.

For telecom1, as depicted in Table 4 and Figures 3,4 the best G is obtained using GA-based method (G=0.5890). Also, the Chi-squared based method doesn’t perform considerably worse. In addition, the model generated using the CFS method is less accurate with (G=0.55). Both GA-based and CFS methods induce comprehensible models however among the all models the best comprehensible model is gained using GA-based method with R=12. Chi-squared based method generates less comprehensible model (R=31) comparing to the GA-based method. In fact, in terms of R, the Chi-squared based model doesn’t make significant improvement.

Considering the above mentioned, generated model using the GA-based method is considered as the most accurate and comprehensible model among the all other models.
For telecom2 as table 5 and Figure 3, 4 show, the best G is achieved using GA-based method (G=0.6003). Other models perform worse in terms of G that is they produce less accurate model. The number of generated rules(R) lies between 3 and 12. Therefore, the comprehensibility of all models is acceptable. Regarding both accuracy and comprehensibility, the GA-based method performs well among the all models.
In sum, the results of experiments show that the models created using the GA-based method on both churn datasets perform well in terms of comprehensibility and accuracy. It generates more comprehensible models with increased performance in terms of geometric mean (G). As stated in literature, a comprehensible churn prediction model reveals the churn drivers of customers. Therefore marketing manager of a firm can gain insight about churn reasons. As a result they can develop effective marketing actions.

5. Conclusion

Customer churn is a major problem for firms operating in competitive markets. As a result of customer churn, many firms face profit losses. Churn prediction is an important tool which is used to identify customers who are at risk of churn. Previous research in the context of churn prediction mainly concentrated on the accuracy of the produced model. However, recent research argued that a churn prediction model should be both accurate and comprehensible. In this paper, we proposed the GA-based method which uses genetic algorithm (GA) to select the best feature set. The main advantage of this method is considering comprehensibility in calculating the cost function of each feature set. This method was compared to the two well-known filter feature selection methods including Chi-squared based and Correlation based feature selection using two telecommunication churn datasets. The results of experiments indicated that the GA-based method performed better than the two filter methods in terms of both accuracy and comprehensibility (measured as number of generated rules). By employing the proposed method, the marketing managers of a firm can easily understand the reasons for customer churn and accordingly they can design effective marketing actions to prevent customers from churning. As a future work, we can use the presented method in other domains such as financial industry.

Acknowledgment:

We thank the Teradata Center for Customer Relationship Management at Duke University for providing churn datasets.

References

[1] Neslin, S.A., Gupta, S., Kamakura, W., Lu, J. and Mason, C. (2006), Defection detection: Measuring and understanding the predictive accuracy of customer churn models, *Journal of Marketing Research*, Vol. 43, pp. 204-211.

[2] Coussement, K., Benoit, D.F. and Van den Poel, D. (2010), Improved marketing decision making in a customer churn prediction context using generalized additive models, *Expert Systems with Applications*, Vol. 37, pp. 2132-2143.

[3] Coussement, K. and 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*, Vol. 34, pp. 313-327.

[4] Yu, X. (2011), An extended support vector machine forecasting framework for customer churn in e-commerce, *Expert Systems with Applications*, Vol. 38, pp. 1425-1430.

[5] Huang, B., Buckley, B. and Kechadi, T.M. (2010), Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications, *Expert Systems with Applications*, Vol. 37, pp. 3638-3646.

[6] Wei, C. and Chiu, I. (2002), Turning telecommunication call details to churn prediction: a data mining approach, *Expert Systems with Applications*, Vol. 23, pp. 103-112.

[7] Pendharkar, P.C. (2009), Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services, *Expert Systems with Applications*, Vol. 36, pp. 6714-6720.

[8] Tsai, C.F. and Lu, Y.H. (2009), Customer churn prediction by hybrid neural networks, *Expert Systems with Applications*, Vol. 36, pp. 12547-12553.

[9] Coussement, K. and Van den Poel, D. (2008), Integrating the voice of customers through call center emails into a decision support system for churn prediction, *Information & Management*, Vol. 45, pp. 164-174.

[10] Lemmens, A. and Croux, C. (2006), Bagging and boosting classification trees to predict churn, *Journal of Marketing Research*, Vol. 43, pp. 276-286.

[11] Verbeke, W. (2011), Building comprehensible customer churn prediction models with advanced rule induction techniques, *Expert Systems with Applications*, Vol. 38, pp. 2354-2364.

[12] Han, J. and Kamber, M. (2006), *Data mining concepts and techniques*, Morgan Kaufmann. San Francisco, CA.

[13] Witten, I.H. and Frank, E. (2005), *Data mining: Practical machine learning tools and techniques*, Morgan Kaufmann. San Francisco, CA.

[14] Guyon, I. and Elisseeff, A. (2003), An introduction to variable and feature selection, *Journal of Machine Learning Research*, Vol. 3, pp. 1157–1182.

[15] Kohavi, R. and John, G.H. (1997), Wrappers for Feature Subset Selection, *Artificial Intelligence*, Vol. 97, pp. 273-324.

[16] Ahn, J.H., Han, S.P. and Lee, Y.S. (2006), Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry, *Telecommunications Policy*, Vol. 30, pp. 552-568.

[17] Abbasimehr, H., Setak, M. and Soroor, J. (2013), A framework for identification of high-value customers by including social network based variables for churn prediction using neuro-fuzzy techniques, *International Journal of Production Research*, Vol. 51, pp. 1279-1294.

[18] Karahoca, A. and Karahoca, D. (2011), GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system, *Expert Systems with Applications*, Vol. 38, pp. 1814-1822.

[19] Burez, J. and Van den Poel, D. (2009), Handling class imbalance in customer churn prediction, *Expert Systems with Applications*, Vol. 36, pp. 4626-4636.

[20] Xie, Y. (2009), Customer churn prediction using improved balanced random forests, *Expert Systems with Applications*, Vol. 36, pp. 5445-5449.
[21] Haenlein, M. (2013), Social interactions in customer churn decisions: The impact of relationship directionality, *International Journal of Research in Marketing*, Vol. 30, pp. 236-248.
[22] Burez, J. and Vandenpoel, D. (2008), Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department, *Expert Systems with Applications*, Vol. 35, pp. 497-514.
[23] Glady, N., Baesens, B. and Croux, C. (2009), Modeling churn using customer lifetime value, *European Journal of Operational Research*, Vol. 197, pp. 402-411.
[24] Tsai, C.F. and Chen, M.Y. (2010), Variable selection by association rules for customer churn prediction of multimedia on demand, *Expert Systems with Applications*, Vol. 37, pp. 2006-2015.
[25] Verbeke, W. (2012), New insights into churn prediction in the telecommunication sector: A profit driven data mining approach, *European Journal of Operational Research*, Vol. 218, pp. 211-229.
[26] Idris, A., Rizwan, M. and Khan, A. (2012), Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies, *Computers & Electrical Engineering*, Vol. 38, pp. 1808-1819.
[27] Sivanandam, S.N. and Deepa, S.N. (2008), *Introduction to Genetic Algorithms*, Springer.
[28] Barandela, R. (2003), Strategies for learning in class imbalance problems, *Pattern Recognition*, Vol. 36, pp. 849-851.