

Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods

Mohammad Fathian and Yaser Hoseinpoor

*School of Industrial Engineering, Iran University of Science and Technology,
Tehran, Iran, and*

Behrouz Minaei-Bidgoli

*School of Computer Engineering, Iran University of Science and Technology,
Tehran, Iran*

Abstract

Purpose – Churn management is a fundamental process in firms to keep their customers. Therefore, predicting the customer's churn is essential to facilitate such processes. The literature has introduced data mining approaches for this purpose. On the other hand, results indicate that performance of classification models increases by combining two or more techniques. The purpose of this paper is to propose a combined model based on clustering and ensemble classifiers.

Design/methodology/approach – Based on churn data set in Cell2Cell, single baseline classifiers, ensemble classifiers are used for comparisons. Specifically, self-organizing map (SOM) clustering technique, and four other classifier techniques including decision tree, artificial neural networks, support vector machine, and K-nearest neighbors were used. Moreover, for reduced dimensions of the features, principal component analysis (PCA) method was employed.

Findings – As results 14 models are compared with each other regarding accuracy, sensitivity, specification, F-measure, and AUC. The results showed that combination of SOM, PCA, and heterogeneous boosting achieved the best performance comparing with other classification models.

Originality/value – This study examined the performance of classifier ensembles in predicting customers churn. In particular, heterogeneous classifier ensembles such as bagging and boosting are compared.

Keywords Classification, Churn prediction, Data mining, Classifier ensemble

Paper type Research paper

1. Introduction

To survive the competitive markets or keep their competitive advantages, many organizations have step toward communication marketing and emphasized on maximizing value of the customer's life cycle and managing the customer's churn. In fact, many organizations believe that their customers are the most valuable assets they have. Therefore, keeping the customer is a long-term profitable strategy that ensures success of the organization. Along with intensification of competition, many businesses have shifted their concentration on customer relationship management (CRM). In CRM, study of the future behaviors of the customer plays a critical role. Future behavior identification of the customer can help companies to take the necessary actions in relationship with the customers as soon as possible (Nie *et al.*, 2011). Customers who stop using products/services of the company are considered as



churn customers. Therefore, by spotting them, companies may have better chance of keeping their customers. One of the main reasons to undergo the customer's preservation process is that keeping a customer is far less expensive than finding a new customer (Nie *et al.*, 2009, 2011). That is, sale cost for new customers is five times of that of old customers (Nie *et al.*, 2011; Slater and Narver, 2000; Li *et al.*, 2011). By achieving 5 percent increase in customer keeping rate a bank can enjoy 85.5 percent increase in profit (Nie *et al.*, 2011). Authors in Lin *et al.* (2011) mentioned that 5 percent increase in customer keeping rate leads to 18 percent reduction in operational costs. Intensification of competition among companies has widened the customer's options for providers of the services/products they needed. Todays, it is easy to switch from one provider to another. By predicting future behavior of the customer, firms can take required measures to prevent the customer's switch. One may say, that churn management includes developing techniques that enable the firm to keep its profitable customers and also increase loyalty of the customer (Farquad *et al.*, 2009). Many prediction problems, in the past, were resolved using statistical methods such as logistic regression or regression analysis. At the current many researches, however, has shifted toward machine learning techniques. Ensemble classifiers in machine learning, given its higher performance, are now adopted by many researchers. As recommended by the literature, ensemble classifiers have higher predicting performance comparing with other methods (Tsai, 2014; West *et al.*, 2005). Furthermore, combination of clustering and classifiers technique, as a modern technique, has drawn attention of many authors in machine learning field. In fact, clustering is pre-processing stage to eliminate outlier data and then clustering is used to develop prediction model (Tsai, 2014; Lin *et al.*, 2014; Hsieh, 2005). Studies have shown that performance of combined clustering method is better than the basic classifier. The idea of combining bagging and boosting ensemble classifiers, by this study, is first of its kind in the literature. Thereby, the present study tries to compare learning technique of advanced and basic machines. The rest of the paper is organized as follows: Section 2 brings in a review of literature about customer's churn and surveys machine learning techniques. Methodology is discussed in Section 3 following by results in Section 4. Finally, Section 5 discusses the results and recommendations.

2. Literature review

2.1 Machine learning techniques

2.1.1 Classification. Classification is one of the most popular supervised learning methods in data mining (Tsai, 2014). The method helps us in predicting future behavior of the customer through classification records of the database into a set of classes defined based on specific measures (Chen *et al.*, 2003). Every sample or record in a database includes a set of features and classification process finds a model that gives predictable features or classification of every sample based on a function of other features of the sample. Four classification techniques to predict the customer' churn was used in this study including support vector machine (SVM), artificial neural networks (ANNs), decision tree, and K-nearest neighbors (KNN).

2.1.2 Clustering. Clustering divides a heterogeneous population into more homogeneous clusters (Ngai *et al.*, 2009). To be more specific, clustering means grouping data into K different clusters so that the data in one cluster are more similar and data in different clusters are different. Among the most common clustering techniques are self-organizing map (SOM) and K-means clustering (Tsai, 2014; Duda *et al.*, 2012).

In this paper we employ the SOM clustering. A SOM is a type of ANN that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples, called a map. The main properties of SOM can be stated as Lin *et al.* (2014):

- (1) The map allocates different numbers of nodes to inputs based on their occurrence frequencies. If different input vectors appear with different frequencies, the more frequent one will be mapped to larger domains at the expense of the less frequent ones.
- (2) The distance relationships between the input data are preserved by their images in the map as faithfully as possible. While some distortion is unavoidable, the mapping preserves the most important neighborhood relationships between the data items, i.e., the topology of their distribution.

2.1.3 Principal component analysis (PCA). PCA is a multivariate statistic technique that determines key features needed to achieve the best description of variance in a set (Lin *et al.*, 2014). Despite other methods that use a subset of features to reduce dimensions of data, PCA combines principle features to create a subset equal or smaller than the primary features. Concerning computation load, this method suits dispersed or diagonal data (Han *et al.*, 2011). The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set.

2.1.4 Ensemble classifiers. The algorithm takes a set of ineffective classifiers and generates a classifier with higher performance through combining the several classifiers (Zaijane, 1999). There are several algorithms recommended in the literature such as majority vote, Borda Count, Bayesian, Bagging, Boosting, and so on (Tsai *et al.*, 2011; Kim *et al.*, 2003). The present study focusses on and compares bagging and boosting. Figure 1 shows general architecture of ensemble classifiers (Tsai, 2014).

2.1.4.1 Bagging. According to the bagging technique, several classifiers are trained independently on different sets of data through bootstrap method (Tsai *et al.*, 2011). Bootstrap method generates k subsets out of training data set through sampling with replacement. Afterward, k classifiers are performed on each subset and eventually

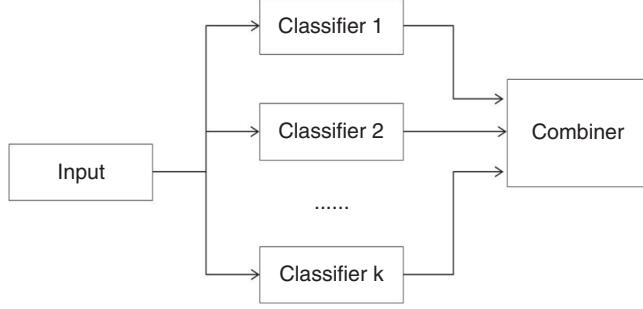


Figure 1.
Architecture of
ensemble classifiers

results of these k classifiers are combined. Finally, when an unknown instance is presented to each individual classifier, a majority or weighted vote is used to infer the class (Galar *et al.*, 2012).

2.1.4.2 Boosting. Boosting algorithm was first introduced by Schapire in 1990 that AdaBoost is the most representative algorithm in this family and it has been appointed as one of the top ten data mining algorithms (Galar *et al.*, 2012). The algorithm employs whole data set to train each classifier serially, but after each round, it gives more focus on difficult instances, with the goal of correctly classification samples in the next iteration that were incorrectly classified during the current iteration. After each iteration, more emphasis is put on the hard data to achieve better classification. Despite bagging, the records have equal weight at first iteration and their weight change after each iteration; so that weight of wrongly ranked samples increases and that of correctly ranked samples decreases. Finally, when a new instance is submitted, each classifier gives a weighted vote, and the class label is selected by majority.

2.1.5 Hybrid classifiers. In general, hybrid classifiers are based on combining two or more machine learning techniques (Tsai and Hung, 2014). The techniques are clustering and classification so that clustering is performed at one stage of pre-processing and then the results of clustering are used as input of classification. Through this, the data that are hardly fit in the clusters are discarded (Figure 2) (Tsai and Chen, 2010). When the classifier is trained enough, it can classify new inputs automatically.

2.2 Evaluating classifiers models

The evaluation criterion is a key factor both in the assessment of the classification performance and guidance of the classifier modeling (Galar *et al.*, 2012). Five measures were usually used for classifiers evaluation in data mining and consequently in customer churn prediction including accuracy, sensitivity, AUC, specificity, and F-measure. These measures can be formulated based on confusion matrix as shown in Table I and formulas 1-7. In many of the studies, the accuracy rate has been the most commonly used empirical measure. However, in imbalanced data sets, accuracy is no longer a proper measure, since it does not distinguish between the numbers of correctly classified examples of different classes. Hence, it may lead to inaccurate conclusions. Also for enrich to comprehensive analysis and comparing with other studies (Idris and Khan, 2012; Idris *et al.*, 2012, 2013), we use variety of validation criteria. The accuracy is the proportion of true results (both true positives

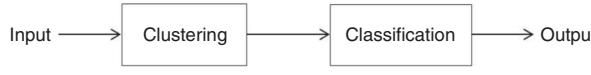


Figure 2.
General architecture of hybrid classifier

		Predicted	
		Positive	Negative
Actual	Positive	True positive(TP)	False negative(FN)
	Negative	False positive(FP)	True negative(TN)

Table I.
Confusion matrix

and true negatives) among the total number of cases examined. Sensitivity (also called the true positive rate (TPR) or the recall) measures the proportion of positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition), and is complementary to the false negative rate. Specificity (also called the true negative rate) measures the proportion of negatives which are correctly identified as such (e.g. the percentage of healthy people who are correctly identified as not having the condition), and is complementary to the false positive rate. Precision (also called positive predictive value) is the ratio of true positives to combined true and false positives. FPrate is the percentage of negative instances misclassified and TPR or recall is the percentage positive instances correctly classified.

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score. AUC provides a single measure of classifier's performance that determines which model is better on average:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$FP_{rate} = \frac{FP}{FP + TN} \quad (3)$$

$$Sensitivity = TP_{rate} = \frac{TP}{TP + FN} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$F-measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (7)$$

2.3 Related works

In this section, a summary of the related works in the field of churn customer and other fields provided. In Lin *et al.* (2014) a combination of SOM, association rules, and PCA was used as data pre-processing method and then the whole process was combined with neural networks. The results showed higher performance achieved by SOM+PCA+ANN. Two combined approaches (SOM+ANN and ANN+ANN) were used in Tsai and Lu (2009) to develop a predication model. The results showed that ANN+ANN outperformed the other combination with respect to performance, accuracy and Types I and II errors. A combined approach based on K-means and SOM was used (Tsai, 2014) to develop a prediction model as data pre-processor combined with decision tree

classifiers, logistic regression, and multilayer perceptron (MLP). The results indicated that combination of SOM and ensemble of MLP classifier achieved highest accuracy, and lowest Types I and II errors. In Tsai *et al.* (2011) three classifiers such as decision tree, neural network, and regression logistic were used as basic classifiers. To combine the classifiers, majority vote, and bagging were used through homogeneous and heterogeneous methods. The results showed that the ANNs with homogeneous majority vote approach outperformed other methods. Bagging and boosting methods were used in Lemmens and Croux (2006) and the results showed that bagging method outperformed boosting and CART decision tree. In Keramati *et al.* (2014) KNN, ANN, decision tree, and SVM were combined and increase in performance was reported. To develop a more efficient model, three data mining techniques (back propagation ANNs, SOM, α -cut fuzzy c-means (α -FCM), Cox proportional hazards regression model) were combined in Mohammadi *et al.* (2013). The combined models were FCM+ANN+Cox, SOM+ANN+Cox, and ANN+ANN+Cox and the results showed that the former outperformed the two others. Another combined approach was used in Coussement and De Bock (2013) to predict churn based on poker players' information on a website. They also used a combination of CART decision tree and generalized additive models (GAM) as an ensemble model (GAM+CART). To evaluate performance of the predictor model, TDL measure and lift index were used. A combined model based on neural network and Cox regression was used in Bahmani *et al.* (2013). The proposed model was a combination of neural network to determine unrelated data and Cox regression to predict future events.

3. Research methodology

3.1 The data set

The data set is provided by Duke University (www.fuque.duke.edu/) referred as Cell2Cell data set. The data set has as many as 40,000 instances, for which labels are readily available. Cell2Cell data set is already processed and provided in balanced shape with 40K instances. Features of the data set included 76 features including eight nominal features and 68 numerical features. The variable "churn" was the variable to be predicted.

3.2 Evaluation

In order to evaluate the performance of a predictive model objectively, we use fivefold cross-validation model. The data are randomly partitioned into five nearly equal size sets; during each iteration one set is used as the validation data, while the remaining four sets of data are used for training. In other words, the validation process is repeated five times; each of the five data sets is used exactly once as the validation data. The five validation results are then averaged to produce the final result. The fivefolds cross-validation testing is adopted for analyzing the performance attained during the experimentation using accuracy, sensitivity, AUC, specificity, and F-measure based on performance measures.

3.3 Prediction models

3.3.1 Basic classifiers. In this paper, ANNs decision tree, SVM, and KNN are used as the baseline prediction model. To construct ANN, there are two important parameters need to be setup in order to avoid the overfitting or overtraining problem. Here we discuss two key parameters including learning rate and three different learning epochs. we designed three different learning epochs (including 300, 500, and 1,000) and three

different numbers of learning parameter (including 0.2, 0.3, and 0.5). With respect to KNN, number of neighbors (K) was set at 3, 5, 7, and 9; and in the end default values were considered for decision tree and SVM. Default values for decision tree and SVM are according to Tables II and III.

3.3.2 Ensemble classifiers. To build an ensemble classifier, heterogeneous ensemble classifier were considered. To design heterogeneous classifier ensembles, the four types of single classifiers described above are combined. That is, we use the best neural network, KNN, decision tree, and SVM models according to their fivefold cross-validation results, respectively. In addition, to combine the classifier heterogeneously, bagging and boosting methods were used. Number of iteration for construct bagging and boosting is 10. To this end, ten sets of data were taken into account and neural network, SVM, decision tree, and KNN were implemented on 3, 2, 2, and 3 sets of data, respectively.

3.3.3 Hybrid classifiers. In this section, at first, SOM method was selected for clustering stage and afterward, different dimensions of SOM were examined to achieve proper clusters needed for classification stage. These dimensions were 2×2 , 3×3 , 4×4 , 5×5 , and 6×6 that generated 4, 9, 16, 25, and 36 clusters. As there are churn and non-churn groups, clusters corresponding to the two groups, respectively, of each SOM, which provide the highest rate of accuracy over the other clusters are selected as the clustering result. Furthermore the other clusters whose data are not well classified or difficult to be classified with SOM are discarded.

3.3.4 Combining clustering approach with ensemble classifiers. This paper considers a more sophisticated approach by combining the clustering method with classifier ensembles instead of single ensembles. Instead of using basic classifiers with clustering method, combination of SOM with bagging, and boosting ensemble methods were used heterogeneously. Therefore, the aim of this paper is to examine the prediction performance by comparing single and advanced machine learning techniques. In particular, the combination of the clustering method and classifier ensembles is compared with four well-known single classifiers, i.e. ANNs decision tree, SVM, and KNN, as the baseline classifiers.

Table II.
Default value for
decision tree

Criteria	Value
Maximal depth	20
Confidence	0.25
Minimal size for split	4
Criterion	Gain ration
Minimal leaf size	2
Minimal gain	0.1

Table III.
Default values
for SVM

Criteria	Value
Kernel cache	200
Convergence ϵ	0.001
Max iterations	100,000

3.4 Evaluation methods

Five measures (accuracy, sensitivity, specificity, F-measure, and AUC) were used to evaluate classification models based on fivefold cross-over evaluation. In this paper 5×5 SOM showed the highest performance of clustering of churn and non-churn groups. On the other hand, ANN with learning rate of 0.2 and 1,000 iterations showed the best performance based on the adopted measures. Eventually, in the case of KNN, value of K was set 9.

4. Experimental results

Due to the issues presented in Section 3, in this section the final results of this study will be presented according to the parameters considered. Totally 14 prediction models were developed which were classified in four categories:

- (1) basic classifier;
- (2) classifier with SOM + basic classifier;
- (3) classifier with SOM + reducing features with PCA + basic classifier; and
- (4) classifier with SOM + reducing features with PCA + bagging and boosting ensemble classifier.

4.1 Basic classifiers

Table IV shows the experimental results the classifiers on the whole data without any pre-processing. Regarding performance, SVM is relatively better than the other classifiers, although, specificity of SVM is less than some of the classifiers. In addition, results of the classifiers are not as promising as expected and except for SVM and ANN, accuracy is not satisfactory. Clearly, SVM has the highest level of accuracy and decision tree is at the bottom line in this regard. However pre-processing on the data at the next section, improves performance of the classifiers.

4.2 Classification by SOM+basic classifiers

Table V shows performance of the classifiers after implementing SOM clustering method. As the results show, with accuracy of 92.79 percent, KNN classifier

AUC	Accuracy	F-measure	Sensitivity	Specificity	Algorithm
50	49.48	54.07	60	40	DT
60.56	57.59	58.37	59.45	55.74	ANN
60.8	57.82	60.83	65.52	50.11	SVM
58.9	56.22	55.29	54.14	58.3	KNN

Table IV.
Performance of
the classifiers

AUC	Accuracy	F-measure	Sensitivity	Specificity	Algorithm
92.9	92.28	92.29	92.26	92.39	DT
93.9	92.43	92.39	92.6	92.26	ANN
93.4	89.07	89.13	92.71	85.6	SVM
93.9	92.79	92.77	93.19	92.41	KNN

Table V.
Classification
with SOM

outperforms other classifiers. In addition, ANN and decision tree classifiers show acceptable accuracy. Next rank is SVM classifier with accuracy of 89.07 percent. It is notable that of KNN with nine neighbors exceeds the expectations. Another point is that except decision tree, for other classifiers, sensitivity is higher than specificity. This means that classifiers are better in determining churn group comparing with non-churn group. This indicates that the classifiers function as expected as we try to determine churn customers. In this regard, AUC by KNN and ANN was highest (93.9 percent).

4.3 Classification by SOM+PCA+basic classifiers

Table VI provides a comparison of these four classifiers. Accuracy of ANN (93.32 percent) is higher than other classifiers. In addition, classifiers KNN, decision tree, and ANN have the highest accuracy. Furthermore, comparison of the results with those of the previous experiment (Table V) clarifies that combination of SOM and PCA gives better results. Therefore, one can optimize the data set to achieve better classifier by doing pre-processing, which cuts the data volume and features. For AUC, KNN and ANN achieved the best result (94 percent) and decision tree achieved lowest AUC (92.8 percent). It can be noted that the ANN model outperforms the other three models in terms of average accuracy and other criteria. However, there is no big difference between ANN and KNN.

4.4 Classification by SOM+PCA+bagging ensemble classifier

Table VII shows the results of this approach implemented heterogeneously. By using the bagging method, the heterogeneous classifier ensembles perform better than the basic classifier regarding all the measures. Accuracy is satisfactorily 95.86 percent and the good results are repeated regarding all other measures.

4.5 Classification by SOM+PCA+boosting ensemble classifier

Table VIII shows the results for heterogeneous classifier via boosting. Clearly, this approach yields better results comparing with bagging method regarding all measures.

Table VI.
Classification with
SOM and PCA

	AUC	Accuracy	F-measure	Sensitivity	Specificity	Algorithm
	92.8	92.47	92.43	92.3	92.64	DT
	94	93.32	93.3	93.38	93.26	ANN
	93.9	92.33	92.29	92.53	92.13	SVM
	94	92.91	92.89	93.15	92.68	KNN

Table VII.
Heterogeneous
bagging

	Algorithm	AUC	Specificity	Sensitivity	F-measure	Accuracy
	Bagging (Hetero)	95.9	95.12	96.1	95.32	95.86

Table VIII.
Heterogeneous
boosting

	Algorithm	AUC	Specificity	Sensitivity	F-measure	Accuracy
	Boosting (Hetero)	96.38	95.92	96.86	95.85	96.13

Accuracy of the proposed method reaches 96.13 percent. One may conclude that boosting method outperforms bagging method as to most of the measures.

4.6 Comparing the proposed method with similar studies

Table IX shows the results obtained by the proposed method and similar methods proposed by other studies. Clearly, the proposed method in this work outperforms other methods regarding all measures. Highest AUC (96.38 percent) is obtained by our proposed method.

5. Discussion and conclusion

Predicting churn of customers in CRM is one of the main concerns of businesses including service businesses such as banks and telecommunication companies. This means that the business owners need to know their customers and through this predict their behaviors pattern. Through this, they have better chance of keeping their customers. Classifier ensembles have been examined in many pattern recognition problems. They have shown better performances than single classifiers. Dimensionality reduction and data reduction are the two important data pre-processing steps in the data mining. The main goal of two steps is to make a given data set more representatives by filtering out irrelevant features and data samples for provide good results. However, for classifier ensembles in particular, much related works are limited to exclusively construct homogeneous classifier ensembles. This study examined the performance of classifier ensembles in predicting customer churn. In particular, heterogeneous classifier ensembles are compared. The experimental results show that combining SOM with classifier ensembles by the PCA approach can provide the best prediction result. In summary, classifier ensembles perform better than single classifiers do. Totally 14 different models were examined based on machine learning techniques and measures such as accuracy, sensitivity, specificity, F-measure, and AUC. The results showed that combination of clustering methods with bagging ensemble and boosting ensemble classifiers led to better results. On the other hand, boosting method showed better performance comparing with bagging. Therefore, we can conclude that in customer churn prediction data pre-processing by SOM and PCA can produce a better data set to construct an optimal prediction model. Future studies can focus on the following issues:

- using other methods for reducing features such as heuristic methods;
- using other ensemble methods such as random forests, rotation forests, and so on;
- using other clustering methods as data pre-processing; and
- using other data sets to compare performance of predicting classifiers.

	AUC	Sensitivity	Specificity
mRMR with RotBoost (Idris and Khan, 2012)	81.6	76.5	74.6
GP with AdaBoost (Idris <i>et al.</i> , 2012)	96	81	89
mRMR with RotBoost (Idris <i>et al.</i> , 2013)	81.6	76.5	74.6
Proposed method (Boosting (Hetero))	96.38	96.86	95.92

Table IX.
Comparing results of
the proposed method
with similar studies

References

- Bahmani, B., Mohammadi, G., Mohammadi, M. and Tavakkoli-Moghaddam, R. (2013), "Customer churn prediction using a hybrid method and censored data", *Management Science Letters*, Vol. 3 No. 5, pp. 1345-1352.
- Chen, Y.L., Hsu, C.L. and Chou, S.C. (2003), "Constructing a multi-valued and multi-labeled decision tree", *Expert Systems with Applications*, Vol. 25 No. 2, pp. 199-209.
- Coussement, K. and De Bock, K.W. (2013), "Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning", *Journal of Business Research*, Vol. 66 No. 9, pp. 1629-1636.
- Duda, R.O., Hart, P.E. and Stork, D.G. (2012), *Pattern Classification*, John Wiley & Sons, New York, NY.
- Farquad, M.A.H., Ravi, V. and Raju, S.B. (2009), "Data mining using rules extracted from SVM: an application to churn prediction in bank credit cards", *Proceedings of the 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, Springer-Verlag Berlin, Heidelberg, pp. 390-397.
- Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. and Herrera, F. (2012), "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches", *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 42 No. 4, pp. 463-484.
- Han, J., Kamber, M. and Pei, J. (2011), *Data Mining: Concepts and Techniques: Concepts and Techniques*, Morgan Kaufmann Publishers Inc., San Francisco, CA.
- Hsieh, N.C. (2005), "Hybrid mining approach in the design of credit scoring models", *Expert Systems with Applications*, Vol. 28 No. 4, pp. 655-665.
- Idris, A. and Khan, A. (2012), "Customer churn prediction for telecommunication: employing various various features selection techniques and tree based ensemble classifiers", *15th International IEEE in Multitopic Conference (INMIC)*, The Center for Customer Relationship Management, Duke University, Durham, NC, pp. 23-27, available at: www.fuqua.duke.edu/
- Idris, A., Khan, A. and Lee, Y.S. (2012), "Genetic programming and adaboosting based churn prediction for telecom", *Conference on IEEE International in Systems, Man, and Cybernetics (SMC)*, pp. 1328-1332.
- Idris, A., Khan, A. and Lee, Y.S. (2013), "Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification", *Applied Intelligence*, Vol. 39 No. 3, pp. 659-672.
- Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M. and Abbasi, U. (2014), "Improved churn prediction in telecommunication industry using data mining techniques", *Applied Soft Computing*, Vol. 24, November, pp. 994-1012.
- Kim, E., Kim, W. and Lee, Y. (2003), "Combination of multiple classifiers for the customer's purchase behavior prediction", *Decision Support Systems*, Vol. 34 No. 2, pp. 167-175.
- Lemmens, A. and Croux, C. (2006), "Bagging and boosting classification trees to predict churn", *Journal of Marketing Research*, Vol. 43 No. 2, pp. 276-286.
- Li, F., Lei, J., Tian, Y., Punyapathanakul, S. and Wang, Y.J. (2011), "Model selection strategy for customer attrition risk prediction in retail banking", *Proceedings of the Ninth Australasian Data Mining Conference*, Australian Computer Society, Vol. 121, pp. 119-124.
- Lin, C.S., Tzeng, G.H. and Chin, Y.C. (2011), "Combined rough set theory and flow network graph to predict customer churn in credit card accounts", *Expert Systems with Applications*, Vol. 38 No. 1, pp. 8-15.

- Lin, W.C., Tsai, C.F. and Ke, S.W. (2014), "Dimensionality and data reduction in telecom churn prediction", *Kybernetes*, Vol. 43 No. 5, pp. 737-749.
- Mohammadi, G., Tavakkoli-Moghaddam, R. and Mohammadi, M. (2013), "Hierarchical neural regression models for customer churn prediction", *Journal of Engineering*, Vol. 2013, Article ID 543940, p. 9.
- Ngai, E.W., Xiu, L. and Chau, D.C. (2009), "Application of data mining techniques in customer relationship management: a literature review and classification", *Expert Systems with Applications*, Vol. 36 No. 2, pp. 2592-2602.
- Nie, G., Rowe, W., Zhang, L., Tian, Y. and Shi, Y. (2011), "Credit card churn forecasting by logistic regression and decision tree", *Expert Systems with Applications*, Vol. 38 No. 12, pp. 15273-15285.
- Nie, G., Wang, G., Zhang, P., Tian, Y. and Shi, Y. (2009), "Finding the hidden pattern of credit card holder's churn: a case of China", *Lecture Notes in Computer Science*, Vol. 5545 No. 1, pp. 561-569.
- Slater, S.F. and Narver, J.C. (2000), "Intelligence generation and superior customer value", *Journal of the Academy of Marketing Science*, Vol. 28 No. 1, pp. 120-127.
- Tsai, C.F. (2014), "Combining cluster analysis with classifier ensembles to predict financial distress", *Information Fusion*, Vol. 16, March, pp. 46-58.
- Tsai, C.F. and Chen, M.L. (2010), "Credit rating by hybrid machine learning techniques", *Applied Soft Computing*, Vol. 10 No. 2, pp. 374-380.
- Tsai, C.F. and Hung, C. (2014), "Modeling credit scoring using neural network ensembles", *Kybernetes*, Vol. 43 No. 7, pp. 1114-1123.
- Tsai, C.F. and Lu, Y.H. (2009), "Customer churn prediction by hybrid neural networks", *Expert Systems with Applications*, Vol. 36 No. 10, pp. 12547-12553.
- Tsai, C.F., Lin, Y.C., Yen, D.C. and Chen, Y.M. (2011), "Predicting stock returns by classifier ensembles", *Applied Soft Computing*, Vol. 11 No. 2, pp. 2452-2459.
- West, D., Dellana, S. and Qian, J. (2005), "Neural network ensemble strategies for financial decision applications", *Computers & Operations Research*, Vol. 32 No. 10, pp. 2543-2559.
- Zaiane, O.R. (1999), "Principles of knowledge discovery in databases", Department of Computing Science, University of Alberta, CMPUT 690, Alberta.

Corresponding author

Mohammad Fathian can be contacted at: fathian@iust.ac.ir

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com