Ensemble Classification and Extended Feature Selection for Credit Card Fraud Detection

F. Fadaei Noghani and M.- H. Moattar[*]

Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

Abstract

Due to the rise of technology, the possibility of fraud in different areas such as banking has increased. Credit card fraud is a crucial problem in banking and its danger is ever increasing. This paper proposes an advanced data mining method, considering both the feature selection and the decision cost for accuracy enhancement of credit card fraud detection. After selecting the best and most effective features, using an extended wrapper method, an ensemble classification is performed. The extended feature selection approach includes a prior feature filtering and a wrapper approach using C4.5 decision tree. Ensemble classification is performed using cost sensitive decision trees in a decision forest framework. A locally gathered fraud detection dataset is used to estimate the proposed method. The method is assessed using accuracy, recall, and F-measure as the evaluation metrics and compared with the basic classification algorithms including ID3, J48, Naïve Bayes, Bayesian Network, and NB tree. The experiments carried out show that considering the F-measure as the evaluation metric, the proposed approach yields 1.8 to 2.4 percent performance improvement compared to the other classifiers.

Keywords: *Credit Card Fraud Detection, Feature Selection, Ensemble Classification, Cost Sensitive, Learning.*

1. Introduction

Generally, fraud "is the act of deceiving to gain unfair, undeserved and/or illegal financial profit" [1]. Fraud detection is an important issue in many areas including credit loans, credit cards, long distance communications, and insurance [2]. Any attempt to detect fraud in these areas is called a fraud detection process [3]. In banking, fraud happens in credit cards, online bank accounts, and call centres (telephone banking) [4]. The sooner the fraudulent transactions are detected, more damages can be prevented by stopping the transactions of counterfeit credit cards [5]. There are two main and important types of frauds related to credit cards. The first one is counterfeit fraud, which is done by organized crime gangs. The second type of credit card fraud is the illegal use of a missing or stolen credit card. Detecting fraud in a card with a larger balance is much more valuable than detecting fraud in a card with a smaller and limited balance. As a result, we have a classification problem with different costs [4].

Fraud detection is one of the best applications of data mining in the industry and the government [6]. Statistical methods of fraud detection are divided into two broad categories, supervised and unsupervised [7]. Traditional fraud detection is very costly due to expensive experts and broadness of the databases. Another deficiency is that not every human expert is able to detect the most recent patterns of fraud. Thus a data mining algorithm should analyze huge databases of transactions, and only then the expert will be able to do a further investigation about the diagnosed risky measures [3].

The purpose of this work is to propose an advanced method of data mining to detect credit card fraud. Feature selection is performed using an extension on the classical wrapper approach. This approach partitions the data prior to the wrapper algorithm. The goal is to select stable features that are independent from the size of the dataset and can be generalized on the other ones.

Also a cost-sensitive approach based on decision forest is proposed to tackle the unbalance data problem, which is an intrinsic property of fraud detection systems. The rest of this paper is organized as what follows. Section 2 introduces the previous works. In section 3, we investigate the feature selection methods and imbalanced datasets, and introduce approaches to overcome them, i.e. the cost-sensitive learning and ensemble methods. Section 4 introduces the proposed method. Section 5 deals with the results obtained and evaluations of the proposed approach.

2. Related works

Logistic models, Bayesian belief network, neural networks, and decision trees are the main data mining techniques for detecting financial frauds (credit card fraud, corporate fraud, and money laundry), and all of them provide original solutions to the problem of detection and classification of the counterfeit data. Generally speaking, approaches applied for detecting credit card fraud include neural network, data mining, meta-learning, and support vector machine [6].

Hilas and Mastorocostas (2008) [8] have proposed an approach based on the user model identification. In order to test the ability of each profile to discriminate between legitimate usage and fraud, feed-forward neural network (FF-NN) is used as classifier. Panigrahi et al. (2009) [9] have proposed a new method for detecting credit card fraud, which combines evidences of the past and present behavior. Their fraud detection system (FDS) consists of four components, which include law-based filter, Dempster-Shafer adder, transaction history database, and Bayesian learner. Duman and Ozcelik (2011) [4] have developed a method in which every transaction is marked and scored. Then based on the scores and ranks, transactions are categorized as legal or fraudulent. Their method has presented a combination of two metaheuristic techniques, namely genetic algorithms and scatter search. Bhattacharyya et al. (2011) [7] have investigated two advanced data mining techniques including support vector machines and random forests accompanied with logistic regression, for a better detection (as well as control and prosecution) of credit card fraud. Jha et al. (2012) [5] have used the strategy of collecting transactions to foresee the purchase behavior of customers. They have used these sets to estimate a model for detecting fraudulent transactions. Dheepa and Dhanapa (2013) [10] have suggested a combination of supervised and unsupervised approaches to detect fraudulent transactions. Their model includes an approach of

grouping basic behaviors benefited from patterns of animals' collective behavior to detect changes in the behavior of credit card users [10].

Sahin et al. (2013) [11] have proposed a cost-sensitive decision tree approach, which minimizes the total cost of incorrect classifications. The performance of this approach is compared with the traditional classification models in real datasets of credit cards. The cost of incorrect categorization is considered variable in this approach. Wei et al. (2013) [12] have suggested a model for an efficient online banking fraud detection, which combines several data mining techniques, cost-sensitive decision tree, and decision forest. Soltani Halvaiee and Akbari (2014) [13] have proposed a distributed model, considering a new method for credit card fraud detection using artificial immune system (AIS).

Santiago et al. (2015) [14] have proposed an approach to address the market fraud problem in on-line payment services. They have presented a model based on the history of entities involved in a transaction and extracted features to classify the transaction as fraud or legal. Kulkarni and Ade (2016) [15] have suggested a framework using logistic regression to tackle the problem of unbalanced data in credit card fraud detection. They have used an incremental learning approach for fraud modeling and detection. Bahnsen et al. (2016) [16] have expanded the transaction aggregation strategy, proposing a new feature based on the periodic behavior of a transaction. A valuable review on datamining approaches for credit card scoring can be found in [17].
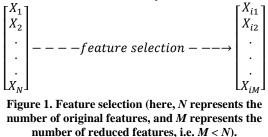
Regarding the fact that datasets of credit cards include many features, there is an urgent need for selecting the best discriminating features. Also, datasets of credit cards for fraud detection include two classes that are not balanced. However, this has been overlooked in the previous works. Thus, it is necessary to pay special attention to the mentioned issues for proposing a practical framework for credit card fraud detection.

3. Material and methods
3.1. Feature selection methods
Feature is a unique and measurable characteristic of a process that is visible [18]. Any time a credit card is used, the transaction data including a number of features (such as credit card ID, amount of the transaction, etc.) are saved in the database of the service supplier [19]. Precise features strongly influence the performance of a fraud detection system [20]. Feature selection is the process of selecting a subset of features out of a larger set, and leads to a successful

classification. The whole search space contains all possible subsets of features, meaning that its size is 2^N, in which N is the number of features. Thus feature selection is an NP-hard problem [21]. Figure 1 depicts the concept of feature selection [22].

In classification, a dataset usually includes a large number of features that may be relevant, irrelevant or redundant. Redundant and irrelevant features are not useful for classification, and they might even reduce the efficiency of the classifier regarding the large search space, which is the so-called curse of dimensionality [23].

$$\begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_N \end{bmatrix} ----feature\ selection----\rightarrow \begin{bmatrix} X_{i1} \\ X_{i2} \\ . \\ . \\ . \\ X_{iM} \end{bmatrix}$$

Figure 1. Feature selection (here, *N* represents the number of original features, and *M* represents the number of reduced features, i.e. *M* < *N*).

The benefits of feature selection include reducing the computational costs, saving storage space, facilitating model selection procedures for accurate prediction, and interpreting complex dependencies between variables [24]. The features that are well selected not only optimize the classification accuracy but also reduce the number of required data for achieving an optimum level of performance of the learning process [25, 26]. Feature selection methods usually include search strategy, assessment measure, stopping criterion, and validation of the results. Search strategy is a search method used for producing a subset of candidate features for assessment. An assessment measure is applied for evaluating the quality of the subset of candidate features. The objective of the stopping criterion is to determine when a decision process should stop, and validation is the study of validity of the selected features with the real world datasets. It is obvious that search strategy and assessment measure are the two key factors in the feature selection process [27]. Filter and Wrapper methods are the most important methods of feature selection [25].

3.1.1. Filter methods
Filter approaches are independent from learning algorithm, and are cheaper and more general than the wrappers from the computational cost viewpoint [23]. Filter methods only evaluate the relation between features, and are independent from the classification and use measures such as distance, information, dependency, and compatibility. Filter methods are classified into

the feature subset selection (FSS) and feature ranking (FR) methods [25]. This classification is based on whether these methods evaluate the relation between the features separately or through feature subsets. In feature ranking methods, each feature is ranked separately, and then the features are ranked based on their relation with the objective variable. The subset selection methods explore all the subsets of features using a certain assessment measure [25].

3.1.2. Wrapper methods
Wrapper methods use the classifier as a black box and its performance as objective function for features subset assessment [18]. Wrapper approaches include a learning algorithm as assessment function [23]. Feature selection criterion in wrapper methods is a forecasting function that finds a subset with the highest performance [18]. Sequential backward selection (SBS) and sequential forward selection (SFS) are two common wrapper methods. SFS (SBS) starts without any features (or all features), and then the candidate features are, respectively, added to (or omitted from) until adding or omission does not increase the classification performance [23].

Comparing the two classes of feature selection approaches, we can say that the filter methods can be considered as preprocessing, which ranks features independent from the classifier. These approaches have a lower computational complexity, and are more generalizable (due to the classifier independence). When the number of initial features is high, the filter approaches are usually time-efficient, and can achieve an acceptable performance. However, their performance depends on the ranking measure and the factors that are taken under consideration (i.e. discrimination power, correlation, class relation, and so on). On the other side, the wrapper approaches span a large search space, and therefore, their time complexity is high, which is negligible in offline systems. Also since features are selected based on the classifier performance, the wrapper methods usually have a better performance on the evaluation datasets. However, the optimality of the wrapper approaches is classifier-dependent, and both their efficiency and generalizablity depend on the classifier. However, the frameworks such as the one proposed in this paper may lead to a better stability of the performance.

3.2. Unbalanced datasets
One of the main problems involved in data mining is the problem of classes being unbalanced. In

some classification problems, the number of samples of each class can be very different. The imbalance problem appears especially when facing a dataset with only two classes [2]. The problem of unbalanced datasets is very important in real world applications such as medical diagnoses, detecting software deficiencies, financial issues, finding drugs, and bioinformatics. In these issues, a class with fewer samples is more important from the learning viewpoint [27, 28], and when its detection is not done properly, the decision costs increase [28]. The methods dealing with the problem of unbalanced datasets can be grouped into three categories [28, 29]:

Data level methods work during pre-processing, and directly on the data and try to re-balance the class distributions. These methods are independent from the real classification stage, and can be used flexibly. The most famous approaches use the oversampling strategy. A popular approach is the synthetic minority oversampling technique (SMOTE), though, recently, better options have been suggested such as adaptive synthetic sampling (ADASYN), which investigates the most difficult objectives for learning or ranked minority sampling (RAMO), which uses the direct probabilistic method. However, the oversampling methods can cause other problems such as changes in class distribution in higher iterations.

The classifier level methods try to make the existing algorithms consistent with the problem of unbalanced dataset, and enrich them towards the minority group. Here, a deeper knowledge is required about the nature of forecasters and the reasons of their defeat in detecting the minority group. The cost-sensitive methods are able to use data correction (by adding a certain cost to the wrong classification) and correcting learning algorithms for making them compatible with the possibility of wrong classification. The higher cost of the wrong classification, which is dedicated to the classification of the minority class, reduces the overall cost. The solutions are based on the cost-sensitive learning combine data methods and the algorithm level.

3.3. Cost-sensitive trees

The induction of a decision tree is an important and active topic in data mining and machine learning. Major algorithms in inducing decision trees such as ID3, CART, and C4.5 are widely and successfully used in different applications [30]. The existing algorithms optimize the classifying decision trees with the objective of maximizing precision in classification or minimizing wrong

classifications. The traditional methods of decision trees are designed under the supposition that all the classification mistakes are considered as equal costs. In fact, in actual applications, different classification errors usually lead to different costs. For example, false negative costs are very different from false positive costs in medical diagnoses. Thus it leads to the creation of cost-sensitive learning (CSL) search area [32].

3.4. Ensemble methods

Ensemble methods are very compatible with unbalanced areas, and have demonstrated a great performance [19]. The Accuracy of the fraud detection model is a critical factor for a proper categorization of fraudulent or legal case [33]. Advancements in machine learning suggest using a classifier ensemble instead of a single forecaster. Many researches indicate that an ensemble of classifiers will have better results than a single classifier. Bagging, boosting, and random forests are the most well-known examples of these methods. Random forests are very efficient for the classification and regression problems [34]. A random forest is a collection of decision trees. The reputation of random forest is due to its high performance compared with the other algorithms [29].

4. Proposed method

As Figure 2 suggests, the proposed method consists of two main parts, namely feature selection and decision forest construction. The first part of the proposed method includes division of the datasets and an extended wrapper method that leads to selecting the best and the most efficient features. The second part of the suggested method consists of dividing the dataset to several parts, making a decision tree for each part, scoring each tree, and choosing the best tree with the highest score in the decision forest.

4.1. Extended wrapper-based feature selection

In this stage, to provide stability on the best features for the final experiments, different subsets of training dataset are created. Thus first, the training dataset is divided into 5 different subsets, which include different percentages of the available data (i.e. 80%, 85%, 90%, 95%, and 100%). First, the features are ranked based on the Chi-squared filter, gain ratio, and ReliefF. These filters are known as appropriate and efficient filters for feature rankings [20]. The Chi-squared filter is based on the χ^2 statistics, and evaluates each feature based on the class labels separately. The objective of the gain ratio filter is to

maximize information gain. ReliefF is a sample-based filter that determines the volubility of a feature by repeated sampling and considering the value of a feature for discriminating a sample from a neighboring sample of a similar or a different class. Equations (1), (2), and (3) denote the Chi-squared, gain ratio, and ReliefF filters, respectively [35]:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{1}$$

$$GR = \frac{IG}{H(X)} \tag{2}$$

$$W[A] = W[A] - \sum_{j=1}^{k} \frac{diff\left(A, R_i, H_j\right)}{m.k}$$
$$+ \sum_{C \neq class(R_i)} \frac{P(C)}{1 - P\left(class(R_i)\right)} \sum_{j=1}^{k} diff(A, R_i, M_j(C)) \tag{3}$$
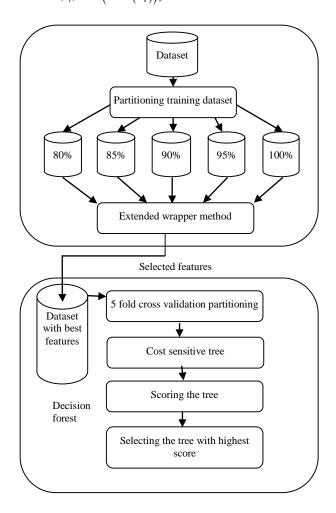

Figure 2. Block diagram of proposed method.

In (1), O_{ij} is the resulting output when E_{ij} is the target output. In (2), IG denotes the information gain [35]. The ReliefF measure, as denoted in (3), randomly selects an instance R_i and its k same-class nearest neighbors, denoted by H_j and k different-class nearest neighbors denoted by

$M_j(C)$. Then the ReliefF measure is updated for attribute *A* using the above-mentioned subsets. The contribution for each class is weighted with the prior probability of that class *P(C)*. The second term is to ensure that the contribution of each class is in the range of [0, 1] and sums to 1 [35].

Then the feature sets created by the Chi-squared, gain ratio, and ReliefF filters are integrated, and for each training subset, a candidate feature set is made. Form the feature sets made by the three filters, a feature with the highest rank is selected. In order to choose the best features of different subsets, the candidate features of each subset that are selected based on their rank, are, respectively, classified by the C4.5 decision tree. Then the accuracy of the classifier is determined. In case the accuracy of the classifier does not decrease, the feature is selected. However, if the feature being studied decreases the accuracy of the classifier, the feature is not selected, and the next feature is investigated (pseudo-code 1).

Pseudo code 1: Selecting best features for each phase

Input: Data set selected from candidate feature set;
Output: Selecting the best features

1. Steps 2 to 6 are repeated until all of candidate features are investigated.
2. Classifier (C4.5) is made for each feature in candidate feature set.
3. Classifier accuracy is calculated.
4. The feature is selected and added to the best feature set if classifier accuracy does not decrease.
5. Else the next feature is investigated.
6. End.

In the last stage, the best features that are common between all the subsets are selected. After performing the first part of the proposed method, the irrelevant features are discarded. The selected features are expected to include the precise features that enhance the accuracy of the classifier.

4.2. Decision forest

In this phase, the dataset with the selected features is divided into several parts in order to make a decision forest (the number of parts in each decision tree is different) and with no overlap. Then for each part, a cost-sensitive decision tree is made; each tree is ranked (based on precision and F-Measure). In the decision forest, the tree with the highest score is selected as the best one. For creating a cost-sensitive tree, the cost of each feature is calculated using CS-Gini [11]. The false negative and false positive decision costs are calculated using (4) and (5), respectively:

$$C_N = \left(\sum_{i=1}^{f} (C_{FN})_i \right) * \left(\frac{f}{n+f} \right)^2 \qquad (4)$$

$$C_P = n * C_{FP} * \left(\frac{n}{n+f} \right)^2 \qquad (5)$$

In (4), C_N is the total cost of wrong classification of legal transactions (non-fraudulent), f shows the number of frauds, n is the number of non-fraudulent transactions, and the cost of a wrong classification of fraudulent transactions (C_{FN}) is equal to 1. In (5), C_P is the total cost of a wrong classification for determining the transaction known as fraudulent. Also the cost of a wrong classification of legal transactions (C_{FP}) is 1.

After calculating the total cost of wrong classifications, the least cost of a wrong classification is selected as the feature cost using (6) [11]:

$$Cost(A) = \min(C_N, C_P) \qquad (6)$$

Then the gain ratio is calculated for each feature A using (7) [36]. In this equation, W shows the importance level of the feature:

$$Rate(A) = 2^{Gain(A)} - 1 / (Cost(A) + 1)^W \qquad (7)$$

By calculating the gain ratio of each feature, the feature with the highest gain ratio is selected as the root of the tree. In the next stage, using the algorithm of the cost-sensitive decision tree, children of the root node are created. The algorithm is repeated for each child as well (pseudo-code 2).

Pseudo code 2: Decision forest

Input: Credit card dataset with best features
Output: The best cost-sensitive decision tree for fraud detection

1. Steps 2-6 are repeated until a tree with a high score is selected.
2. Training dataset is divided into several parts.
3. Cost-sensitive decision tree is made for each part of dataset.
4. Each tree made in the decision forest is ranked based on precision and F-Measure.
5. The tree with a high score between trees of decision forest is selected.
6. End

---

5. Evaluations

To evaluate the proposed method, the dataset from the second robotic & artificial intelligence festival of Amirkabir University was applied (http://araif2013.aut.ac.ir/index/). Table 1 shows the characteristics of the dataset. In this dataset, 3.75% of instances are fraudulent transactions, and 96.25% are non-fraudulent. As seen, the fraudulent and non-fraudulent classes are obviously imbalance.

Table 1. Characteristic of dataset.

Non-fraudulent	Fraudulent	Features	Instance
28012	1092	20	29104

5.1. Assessment measures

In the presented article, four assessment measures were used, namely recall, precision, F-measure, and accuracy. Regarding that the mentioned measures are calculated based on the confusion matrix, this matrix is depicted in table 2. The confusion matrix shows the performance of the classification algorithm when assigning input data to different classes [11].

Table 2. Confusion matrix.

	Positive (Fraud)	Negative (Non-Fraud)
Positive (Fraud)	True Positive (TP)	False Negative (FN)
Negative (Non-Fraud)	False Positive (FP)	True Negative (TN)

The recall measure (8) shows the efficiency of the classifier in detecting the actual fraudulent transactions.

$$Recall = \frac{TP}{FN + TP} \qquad (8)$$

The precision measure (9) shows how much the output of the classifier is reliable.

$$Precision = \frac{TP}{TP + FP} \qquad (9)$$

Finally, the F-measure (10) is the harmonic mean of recall and precision measures.

$$F - Measure = \frac{2 * Recall * Precision}{Recall + Precision} \qquad (10)$$

Also accuracy (as in (11)) denotes the total performance of a classifier. It shows that how many of the total experimental records have been classified correctly by the designed classifier.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (11)$$

However, the F-measure is a more trustable measure for evaluating the data mining systems with imbalance classes because it is the harmonic mean of Recall and Precision measures. Therefore, using the F-measure, both the TP and TN measures are equally important when we have an unbalanced dataset. We used a 5-fold cross scheme approach to evaluate the proposed approach.

5.2. Efficiency of decision forest

The efficiency of the decision forest with different numbers of trees and based on the F-measure and precision is depicted in figures 3 and 4. In these experiments, increasing the number of trees in the

decision forest continued up to the level that both the F-measure and accuracy approach their maximum value and become stable.
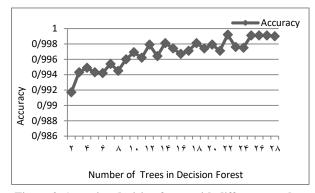

Figure 3. Assessing decision forest with different number of trees using accuracy.

Figure 4 denotes the changes in F-measure with increasing the number of trees in the decision forest. F increases to 0.9975 with increasing the number of trees in the decision forest from 2 to 4. This measure decreases from 0.9975 to 0.9971 when the number of trees rises to 6. Then by increasing the number of trees from 6 to 23, changes in the F-measure has a growth-decline process and it is repeated. This measure reaches a constant level of 0.9988 in a decision forest with 23 and 24 trees, and increases with 25, 26, and 27 trees and reaches 0.9996.
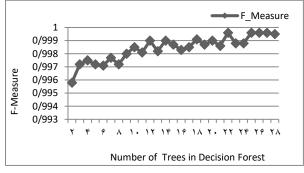

Figure 4. Assessing decision forest with different numbers of trees using F-measure.

Similarly, regarding figure 3, it is obvious that the accuracy increases to 0.9949 by increasing the number of trees from 2 to 4. This measure decreases from 0.9949 to 0.9942 by the increment of the trees from 4 to 6. Then by increasing the trees from 6 to 23, the changes in accuracy become a growth-decline process. This measure increases to 0.9991 in a forest with 25, 26, and 27 trees, and becomes stable.

5.3. Comparison with other classification approaches

The results of the proposed method (with 27 trees with maximum precision and F-measure) was

compared with some basic classifiers including ID3 tree, J48 tree, Naive Bayesian, Bayesian network, and NBT tree using recall, precision, and F measure, as depicted in figures 5, 6, and 7.
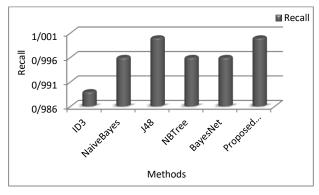

Figure 5. Comparison between proposed method (with 27 trees in decision forest) and basic classification algorithms based on recall.
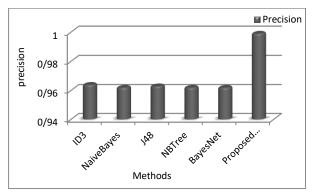

Figure 6. Comparison between proposed method (with 27 tees in decision forest) and basic classification algorithms based on precision.
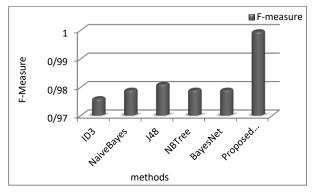

Figure 7. Comparison between proposed method (with 27 tees in decision forest) and basic classification algorithms based on F-measure.

These figures show that the proposed approach (with 27 trees in the decision forest) is superior to the mentioned algorithms. Based on the F-measure, the proposed method had 2.4%, 1.8%, 2%, 2%, and 2% absolute advantage over ID3 tree, J48 tree, Naive Bayesian, Bayesian network, and NB Tree, respectively.

Also, as depicted in figure 7, the error rate of the proposed approach was about 0.2%, while the

error rate of ID3, NaïveBayes, J48 tree, NB Tree, and BayesNet approaches are 2.6%, 2.2%, 2.0%, 2.2%, and 2.2%, respectively. Therefore, the relative error rate of the proposed approach decreased by 92.3%, 90.9%, 90%, 90.9%, and 90.9% as compared with the above-mentioned approaches. This shows a considerable decrement in the detection error, which is mainly due to the precision of the approach resulting from a cost-sensitive paradigm.

6. Conclusion

Along with the recent advances of technology, credit cards have been accepted as one of the most important cases of pay systems. Due to the deficiencies in the security of credit card systems, fraud is increasing, and millions of dollars are lost every year. Thus, credit card fraud detection is a highly important issue for banks and credit card companies. The sooner the fraudulent transaction is detected, the more damages can be prevented. The proposed approach benefited from the extended wrapper method for selecting good features that are efficient for decreasing the run time and increasing the accuracy of the classifier. Then using the decision forest that consists of cost-sensitive decision trees, each tree was scored regarding accuracy and F-measures, and later, the tree with the highest score was chosen. The results obtained indicated that the proposed method is superior to the basic classification algorithms including ID3 tree, J48 tree, Naive Bayesian, Bayesian Network, and NB tree. The precision of the proposed method was 99.96 percent based on the F-measure.

Further works are suggested on using other methods such as majority voting for selecting features and applying other cost-sensitive learning approaches. One can study approaches such as sampling methods to overcome the class imbalance problem. In addition, since it is claimed that the distance metric learning (DML) approaches are robust against class imbalance, their applicability can be studied as a future work.

References

[1] Humpherys S. L., et al. (2011). Identification of fraudulent financial statements using linguistic credibility analysis, Decision Support Systems, vol. 50, no. 3, pp. 585-594.

[2] Kim J., et al. (2012). Classification cost: an empirical comparison among traditional classifier, cost-sensitive classifier, and meta-cost, Expert Systems with Applications, vol.39, no.4, pp. 4013-4019.

[3] Aral K. D., et al. (2012). A prescription fraud detection model. Computer Methods and Programs in Biomedicine, vol. 106, no. 1, pp. 37-46.

[4] Duman, E. & Ozcelik, M. H. (2011). Detecting credit card fraud by genetic algorithm and scatter search. Expert Systems with Applications, vol. 38, no. 10, pp. 13057-13063.

[5] Jha, S., Guillen, M. & Westland, J. C. (2012). Employing transaction aggregation strategy to detect credit card fraud. Expert Systems with Applications, vol. 39, no. 16, pp. 12650-12657.

[6] Ngai E. W. T., et al. (2011). The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. Decision Support Systems, vol. 50, no. 3, pp. 559-569.

[7] Bhattacharyya S., et al. (2011). Data mining for credit card fraud: a comparative study. Decision Support Systems, vol. 50, no. 3, pp. 602-613.

[8] Hilas, C. S. & Mastorocostas, P. As. (2008). An application of supervised and unsupervised learning approaches to telecommunications fraud detection. Knowledge Based Systems, vol. 21, no.7, pp. 721-726.

[9] Panigrahi S., et al. (2009). Credit card fraud detection: a fusion approach using Dempster–Shafer theory and Bayesian learning. Inform Fusion, vol. 10, no. 4. pp. 354-363.

[10] M. Thampi S., et al. (2013). Hybrid approach for improvising credit card fraud detection based on collective animal behavior and SVM. Security in Computing and Communications, Mysore, India: Springer, pp. 293-302.

[11] Sahin, Y., Bulkan, S. & Duman, E. (2013). A cost-sensitive decision tree approach for fraud detection. Expert Systems with Applications, vol. 40, no. 15, pp. 5916-5923.

[12] Wei W., et al. (2012). Effective detection of sophisticated online banking fraud on extremely imbalanced data. 21st International World Wide Web Conference, Lyon, France.

[13] Soltani Halvaiee, N. & Akbari, M. K. (2014). A novel model for credit card fraud detection using Artificial Immune Systems. Applied Soft Computing, vol. 24, pp. 40-49.

[14] Santiago, G. P., Pereira, A. C. M. & Hirata, R. (2015). A modeling approach for credit card fraud detection in electronic payment services. 30th Annual ACM Symposium on Applied Computing, New York, USA.

[15] Kulkarni, P. & Ade, R. (2016). Logistic regression learning model for handling concept drift with unbalanced data in credit card fraud detection system. 2th International Conference on Computer and Communication Technologies.

[16] Bahnsen A. C., et al. (2016). Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, vol. 51, pp.134-142.

[17] Sadatrasoul, S. M., Gholamian, M. R., Siami, M. & Hajimohammadi, Z. (2013). Credit scoring in banks and financial institutions via data mining techniques: A literature review. Journal of AI and Data Mining, vol. 1, no.2, pp. 119-129.

[18] Chandrashekar, G. & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, vol. 40, no. 1, pp. 16-28.

[19] Dal Pozzolo A., et al. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, vol. 41, no. 10, pp. 4915-4928.

[20] Chang, W. H. & Chang, J. S. (2012). An effective early fraud detection method for online auctions. Electronic Commerce Research and Applications, vol. 11, no. 4, pp. 346-360.

[21] Tabakhi, S., Moradi, P. & Akhlaghian, F. (2014). An unsupervised feature selection algorithm based on ant colony optimization. Engineering Applications of Artificial Intelligence, vol. 32, pp. 112-123.

[22] Zhang Y., et al. (2014). Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. Knowledge-Based Systems, vol. 64, pp. 22-31.

[23] Xue, B., Zhang, M. & Browne, W. N. (2014). Particle swarm optimization for feature selection in classification: Novel initialization and updating mechanisms. Applied Soft Computing, vol. 18, pp. 261-276.

[24] Feng, D., Chen, F. & Xu, W. (2014). Supervised feature subset selection with ordinal optimization. Knowledge-Based Systems, vol. 56, pp. 123-140.

[25] Bouaguel, W., Mufti, G. B. & Limam, M. (2013). A fusion approach based on wrapper and filter feature selection methods using majority vote and feature weighting. International Conference on Computer Applications Technology (ICCAT), Sousse.

[26] Wang, G., Ma, J. & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. Expert Systems with Applications, vol. 41, no. 5, pp. 2353-2361.

[27] Zhao, H. & Qin, K. (2014). Mixed feature selection in incomplete decision table. Knowledge-Based Systems, vol. 57, pp. 181-190.

[28] López V., et al. (2012). Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification, open problems on intrinsic data characteristics. Expert Systems with Applications, vol. 39, no. 7, pp. 6585-6608.

[29] del Río S., et al. (2014). On the use of MapReduce for imbalanced big data using random forest. Information Sciences, vol. 285, pp. 112-137.

[30] Krawczyk, B., Woźniak, M. & Schaefer, G. (2014). Cost sensitive decision tree ensembles for effective imbalanced classification. Applied Soft Computing, vol. 14, pp. 554-562.

[31] Kotsiantis, S. B. (2011). Decision trees: a recent overview. Artif Intell Rev, vol. 39, pp. 261-283.

[32] Zhang, S. (2012). Decision tree classifiers sensitive to heterogeneous costs. Journal of Systems and Software, vol. 85, no. 4, pp. 771-779.

[33] Louzada, F. & Ara, A. (2012). Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool. Expert System Applications, vol. 39, pp. 11583-11592.

[34] Elghazel, H., Aussem, A. & Perraud, F. (2011). Trading-off diversity and accuracy for optimal ensemble tree selection in random forests. Ensembles in Machine Learning Applications, Chennai, India, pp. 169-179.

[35] Robnik-Šikonja, M. & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learning, vol. 53, no. 1, pp. 23-69.

[36] Jiawei, H., Micheline, K. & Jian, P. (2012). Data Mining concepts and techniques. 3nd ed. USA, Elsevier.

دسته‌بندی گروهی و انتخاب ویژگی توسعه یافته جهت تشخیص تقلب کارت اعتباری

فهیمه فدائی نوغانی و سید محمدحسین معطر*

گروه کامپیوتر، دانشگاه آزاد اسلامی واحد مشهد، مشهد، ایران.

چکیده:

با توجه به توسعه فناوری، امکان تقلب در حوزه‌های مختلف از جمله تقلب در بانکداری گسترش یافته است. تقلب کارت اعتباری یک مشکل مهم و رو به افزایش در بانکداری محسوب می‌شود. این مقاله، یک روش داده‌کاوی پیشرفته را با در نظر گرفتن همزمان انتخاب ویژگی و هزینه تصمیم برای افزایش صحت تشخیص تقلب کارت‌های اعتباری پیشنهاد می‌کند. پس از انتخاب برترین و مؤثرترین ویژگی‌ها با استفاده از روش پوششی توسعه یافته، یک دسته‌بندی گروهی انجام می‌شود. روش انتخاب ویژگی توسعه یافته، شامل یک فیلترینگ ویژگی اولیه و سپس روش پوشش با استفاده از درخت تصمیم C4.5 می‌باشد. دسته‌بندی گروهی با استفاده از درخت‌های تصمیم حساس به هزینه در قالب یک جنگل تصمیم انجام می‌شود. جهت ارزیابی روش پیشنهادی از یک مجموعه داده تشخیص تقلب محلی استفاده شده است. روش پیشنهادی با استفاده از معیارهای دقت، بازخوانی و F ارزیابی می‌شود و با الگوریتم‌های پایه دسته‌بندی شامل ID3، J48، بیزین ساده، شبکه بیزین و NBTree مورد مقایسه قرار گرفته است. نتایج نشان می‌دهد که بر اساس معیار ارزیابیF، روش پیشنهادی از ۱/۸ تا ۲/۴ درصد نسبت به الگوریتم‌های دسته‌بندی ذکر شده بهبود دارد.

کلمات کلیدی: تشخیص تقلب کارت اعتباری، انتخاب ویژگی، دسته‌بندی گروهی، یادگیری حساس به هزینه.