International Journal of Scientific and Technical Advancements
IJSTA
ISSN: 2454-1532

Meta Classification Technique for Improving Credit Card Fraud Detection

S. Suganya[1], N. Kamalraj[2]
[1]Department of Computer Science, Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore, India-641049
[2]Department of Information Technology, Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore, India-641049
Email address: [1]sugiselvi85@gmail.com, [2]tpkamal@gmail.com

***Abstract*—** Data mining is the process of automatic classification of cases based on data patterns obtained from a dataset. A number of algorithms have been developed and implemented to extract information and discover knowledge patterns that may be useful for decision support. Credit card frauds occur by online and offline. Due to increase in recent developments in technology fraud transactions also increased. In this work a ensemble method based on the D-TREE, SVM, KNN and FA is proposed for solving transaction data classification problems. Initial solutions are generated at random using D-TREE, SVM, KNN and the FA that tries to optimize the weights of the D-TREE, SVM, and KNN carries out the improvement. Experiments results using CREDITCARD transaction data sets show that the proposed FA-D-TREE, SVM, KNN outperforms the D-TREE, SVM, KNN on datasets. Further comparison with other approaches in the literature shows that the ensemble method is able to minimize the error rate. All results show ensemble FA-D-TREE, SVM, KNN outperforms normal methods.

***Keywords*—**Datamining; D-TREE; FA; KNN; SVM.

I. INTRODUCTION

Data Mining or Knowledge Discovery is more important to extract needful data or information from huge amount of data collection. Knowledge Discovery in data is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data [1]. Analysis and prediction are also a part of data mining process which is used to extract models with different data classes and to predict future models using extracted models by analyzing it. Classification in data mining consists of two steps namely model or classifier building and model or classifier usage for classification of huge amount of data.

Credit Card frauds is of two types as online through internet, phone and the like and offline which occurs by loss of credit card. The filter and wrapper are two forms of feature selection models for classification. Ensemble learning method is used in this work; in which collection of methods learn a target function by training a number of individual learners and combining their predictions. The reasons for using ensemble learning is to improve accuracy and efficiency

Jerzy Stefanowski et al. [2] proposed an experimental study and the main aim of this proposed method is to improve the classification accuracy. Experiments are done over various benchmark datasets and found that the combiner classifier is having higher classification accuracy than the single classifiers.

The ample number of ensemble methods is now available for researchers in that area. Different factors that distinguish ensemble methods are Inter-classifiers relationship, combining method, diversity generator, ensemble size. The various ensemble algorithms are bagging. Random subspaces, random forest, rotation forest.

II. LITERATURE REVIEW

Several taxonomies are arisen in literature which is used to categorize ensemble methods from algorithm designer point. Sharkey [5] proposed taxonomy for ensemble of neural networks which suggests three dimensions**:**

1. Two modes of ensemble members are Competitive where single member is used for classification and Co-operative where all members are selected for classification.
2. Ensemble creation is done whether Top-down where combination is based on some other feature other than output of classifiers or Bottom-up which takes members' output for their combination. Further Bottom-up methods are divided into fixed (voting) and dynamic (stacking).
3. The ensemble of components are whether modular or hybrid which make a distinction between modular and pure ensemble systems. The main theme of pure ensemble systems is the combination of classifier set, where every set solves the same unique task and to get a most accurate and reliable performance when compared to single user. The complex problem is broken down into number of solvable problems is the use of modular systems.

Decision optimization and Coverage optimization are two important categories of ensemble techniques.

Brown (4) proposes that According to the diversity whether the ensemble methods choose implicitly by randomization or explicitly by some other metric. Three factors influence in grouping techniques which are as follows, how they initialize the inducers in the hypothesis space, what the space of accessible hypotheses is, and how that space is traversed by the inducer.

Although several surveys on ensemble for classification tasks are available in the literature [3] and there are several papers which suggest taxonomy for ensemble methods [4], in this paper the four main contributions are introduced:

S. Suganya and N. Kamalraj, "Meta classification technique for improving credit card fraud detection," *International Journal of Scientific and Technical Advancements*, Volume 2, Issue 1, pp. 101-105, 2016.

1. All noteworthy ensemble methods are to be categorized into a fresh unified taxonomy. As noted in [3], a structure is now only gradually under development by numerous attempts. A fresh taxonomy suggests systemizing present taxonomies into logical and unified taxonomy.
2. The updated survey of ensemble learning is proposed since it is an lively research field.
3. Efficient and mature ensemble methods are covered in this paper which does not fit in mainstream.
4. To choose apt ensemble method, numerous criteria for selection is given from practitioner's view of point.

A. Neural Network

Raghavendra Patidar, Lokesh Sharma proposed a work by neural networks on credit card fraud detection. Different techniques are available to detect fraud but very few are able to find fraudulent one in evolution. The short time to conclude to accept or to reject and enormous amount of transactions in credit card to be processed are atypical characteristics of fraud detection in credit card usage.

a) Working principle (Pattern recognition)

Neural network is alike a working of human brain. Computer is made to imagine like a human brain which gains knowledge from past experience that is used to solve a daily problems in life. The user of the credit card will use some standard pattern which is trained for previous two or three years on neural network. Other information such as huge purchase frequencies and the like are stored. Credit card pattern usage is trained with different other faces of credit card provided by the particular bank. Prediction algorithm is used to discriminate fraudulent and non-fraudulent transactions from usage pattern of credit card. The trained original card holder's pattern is matched with illegal user's pattern, when they are same then the final conclusion on that transaction is genuine.

b) Fraud detection

The small variations in pattern matching can be accepted and if variations are big then the transaction is fraud or illegal. Neural network's output will be stuck between 0 and 1. When the output lies below .6 or .7 then the transaction is legal. When the output is above .7 the transaction's probability of being illegal is high. There are some situations when card holders (legal users) make different pattern transaction which resembles illegal users' transaction and vice versa. Card holders will use credit card according the amount limit specified by the bank while fraudster won't be the like because he will try to use card for more amount before legal action is taken by the card holder. History descriptors have some details such as payment and card details, date of issue and so on [6].

B. Decision Tree and SVM

Y. Sahin and E. Duman developed a work with real dataset for the comparison of performance between SVM and Decision trees algorithms. The result of comparison is decision tree models are more enhanced in final result than SVM Models. SVM models have a problem of over fitting of training dataset during result comparison. The identification of more number of fraud transactions is the success factor of this problem. Despite of whether transaction is true fraud or true normal assignment, the accuracy shows true assignment rate.

According to accuracy, the performance of SVM models becomes as equal as decision tree model and performance is compared as the increase in number dataset for training, over fitting becomes less. Decision tree models caught more number of fraudulent transactions than SVM models. The accuracy is not matched accordingly with performance metric in this problem. Among more models C5.0 is finest but C&RT models gets more fraud transactions from samples. [7]

C. Based on Frequent Item Set Mining

K. R. Seeja and Masoumeh Zareapoor worked on highly imbalanced and anonymous credit card transaction datasets to detect frauds.

The class with imbalance problem is handled by frequent item set mining which is used for identifying legal and illegal patterns. A matching algorithm is developed for identifying legal or illegal pattern from the incoming transactions. Pattern identification is done by equally treating every attribute without giving attention on attributes to manage anonymous nature of transaction. This work has fewer false alarms when compared to state of the art classifiers, rate of fraud detection is high, classification rate is balanced, Matthews correlation coefficient.

"Fraud Miner" is the key technique proposed. Patterns are created for legal and illegal transactions for each customer in frequent item set mining in the training phase respectively. The value "0" is returned if legal pattern is matched with incoming transaction. If the value returned by algorithm is "1" then the incoming transaction is fraudulent [8].

D. Meta Classification Strategy

Joseph Pun, Yuri Lawryshyn followed the meta-learning techniques introduced by Chan and Stolfo [10] in their proposed work. The results of various learners to prediction accuracy are combined by Meta learning method. And also pros and cons of methods are complimented between the methods. Arbiter and combiner are the two methods of combiner algorithms. Combiner methods are most effective than arbiter is found by the experiments. The attributes and correct classifications are used to train base classifier in combiner method. Meta-level classifier gets an input from resulting predictions. The training data for Meta classifier is fed from the combination of original attributes, base classifier prediction and correct classifications for all instances which is a "combined" dataset. Meta-level classifier's prediction is the final prediction in the combiner strategy [9].

E. Hidden Markov Model

Fraud Detection System (FDS) is runned by the bank which issues credit card. Every incoming transaction is verified by FDS. Genuine and fake transactions are identified from card details and purchase details by FDS. Comparison of spending details, address for delivery and the like is checked by FDS for finding the difference. If the transaction is identified as fake or fraud then that particular transaction is declined [10].

S. Suganya and N. Kamalraj, "Meta classification technique for improving credit card fraud detection," *International Journal of Scientific and Technical Advancements*, Volume 2, Issue 1, pp. 101-105, 2016.

a) HMM model for credit card transaction processing

HMM should determine observation symbols to handle credit card transactions. Control the x values of purchase into M price ranges such as V1, V2,…, VM, establishes observation symbols to bank.

b) Generation of observation symbols

Each credit card holder is trained by HMM. The observation symbols of each card holder's transaction are got by the clustering algorithm. Issuing bank consists of many attributes in the database.

c) Checking spending profile

Three types of spending details of card holders are low, high, and medium.

d) Model parameter estimation and training

Few transactions are used to train proposed model and is developed with further enhancements for future references to detect frauds efficiently.

e) Fraud detection

Initial symbol sequence is formed from the symbols taken from the cardholder's training data after learning HMM factors [10].

F. Bayesian Classification

Sam Maes, Karl Tulys, Bram Vanschoenwinkel, Bernard Manderick [11], had done a work on fraud detection using Bayesian Belief Networks which received a high score from STAGE algorithm. The experimental results from four features of dataset results in 68% of fraudulent transactions are correctly identified and 10% of genuine transactions are falsely classified as fraudulent transactions. Another experimental result from ten features of dataset is 15% of fraudulent transactions are incorrectly classified and & 73% of transactions are fraudulent.

III. RESEARCH METHODOLOGY

The system is first loaded with the .csv format Italian Government Employee Transaction Dataset which is then preprocessed and converted into array format. The preprocessed data is then produced as input for base classifiers (D-Tree, SVM, KNN) separately. The data is trained and classified by each algorithm separately which is given as input to meta classifier where ensemble technology is implemented to find the optimal classified solution.

A. KNN Algorithm

Closest training examples in the feature space are used for classifying objects by KNN Algorithm. Two types of sets is divided by KNN as test set and training set. Euclidean Distance is used to find training sets objects for each row of the test set and majority vote is used for classification. If **k**th nearest vector has ties, all candidates are included to vote.

The transaction date is taken as a feature for classification. Accurate predictions are done about unknown data after trained on known data. (w1, w2… wn, v) are given as training tuples. In testing part (classification) only (w1, w2… wn) is given, the main aim is to find 'v' with more accuracy. Euclidean distance is calculated by using the formula,

$$D(\vec{w},\vec{v}) = \sqrt{\sum_i^n (w_i - v_i)^2}$$

Steps:

1. Randomly select some 'k' number of transaction dates.
2. Using test set find the classes where data in training set is classified using distance function.
3. Calculate the labeled data with (+/-) difference.
4. Draw bisectors.
5. Extend & join all bisectors.

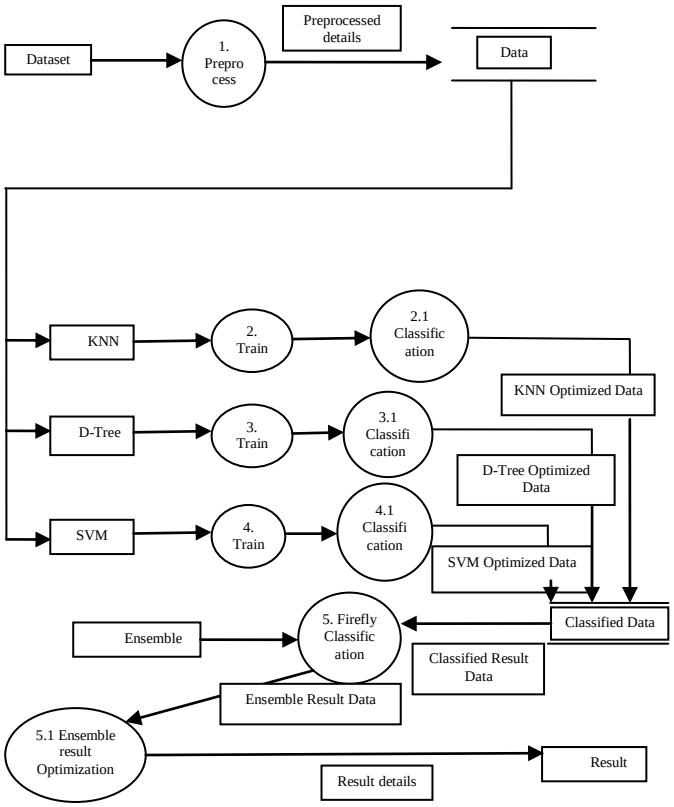

Fig. 1. Framework of the system.

B. Decision Trees

Decision tree is build by greedy method using lowest disorder tests. The creditor is taken as a key feature based on which classification is made. In credit card dataset context, each feature 'fi' is the test count of credit card, 'vi' is result amplitude and 'di' is amplitude. Entropy is calculated by considering if a random variable 'fi' can take 'vi' different values then the **ith** value with probability 'pi', entropy is calculated as

$$H = -\sum_i p_i \log_2 p_i$$

Class entropy is calculated and is used to construct decision tree. Root of the tree can be any feature test that usually maximally distinguishes class labels.

C. Support Vector Machine (SVM)

SVM classifies the credit card dataset based on the services feature. Data is given as input to SVM and function is

S. Suganya and N. Kamalraj, "Meta classification technique for improving credit card fraud detection," *International Journal of Scientific and Technical Advancements*, Volume 2, Issue 1, pp. 101-105, 2016.

given as output which is used to predict the future data's feature. The aim of SVM is to use optimal hyper plane for linearly separable patterns. For not linearly separable by transformations of original data to map into new space called kernel function. Credit card fraud detection uses linear separability, since these are high dimensions it needs hyper plane.

Steps:

1. Consider (x1,….xn) is data set and yi(1,-1) be class label of xi.
2. Find the boundary for the data set.
3. The decision boundary can be found by finding the solution for constrained optimization problem. (f(x) = wx + b).
   w-> weight factor, b-> bias.

D. Ensemble Classifier – Firefly

Yang (2009) developed FA which is based on population. The flashing light produced by fireflies is main aspect of FA. The light intensity makes the fireflies to attract each other and for other activities also used. Minimum intensity fireflies are easily attracted towards maximum intensity fireflies. This concept is used as an optimization algorithm; the flashing light of fireflies is mapped to fitness function which is to be optimized.

In this study, the FA is employed to optimize the weights of the D-TREE, SVM, KNN model, denoted as FA-D-TREE, SVM, KNN, to obtain the optimal parameter settings for training the network of D-TREE, SVM, KNN and to minimize the error rate. The quality of transaction is measured on the error rate which is calculated on the basis of confusion matrix.

```
Begin
        Generate the initial solution randomly
        Evaluate each individual in the population f(x)
based on error rate
        Find the best solution from the population
        While (stopping criterion satisfied)
                For i = 1 to n do
                        For j = 1 to n do
                                If (f (xj) < f (xi))
        Calculate attractive fireflies by eq.1
        Calculate the distance between each fireflies i
                        and   j by eq.
        Move all firefly (xi) to the best solution (xj)  by
eq.3
                End if
                End for j
                End for i
                Moves best solution randomly by eq.4
                Find the best solution from the new
population
        End while
Return best (TP), (TN), (FP), and (FN)
End of the algorithm
```

In FA, the form of attractiveness function of a firefly is depicted by the following:

$$\beta(r) = \beta_0 \exp(-\gamma r^2) \quad (1)$$

where,

r = the distance between any two fireflies

β0 = the initial attractiveness at r = 0 and set to 1 in this study

γ = an absorption coefficient which controls the decrease of the light intensity and also set to 1 in this study

$$r_{ij} = \|x_{i-} x_j\| = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (2)$$

$$x_i = x_i + \beta_0 * \exp(-\gamma r_{ij}^2) * (x_{j-} x_i) + \alpha * \left(rand - \frac{1}{2}\right) \quad (3)$$

Attractiveness Distance

Attraction part Randomly part

$$x_i = x_i + \alpha * \left(rand - \frac{1}{2}\right) \quad (4)$$

IV. IMPLEMENTATION AND RESULT

A. Benchmark Datasets

This experiment is performed on datasets that can be freely downloaded from the CREDITCARD Transaction data Classification Homepage: www.cs.CreditCard.edu/~eamonn/time_series_data. The data contains data sets, which come from different domains (Table I).

TABLE I. Instances and features of algorithm.

Algorithm	Instance#	Feature#
KNN	2424	100
D-Tree	2397	617
SVM	2000	500
Firefly	2000	649

All the CREDITCARD data sets are categorized as having similar complexity to real-world data sets with the data sets based on several criteria. All the benchmark CREDITCARD transaction data sets have a moderate to high transaction data length that ranges from 1996 to 2637 transaction data length.

The results clearly indicate that the hybrid method (FA- D-TREE, SVM, and KNN) has outperformed the D-TREE, SVM, KNN algorithm on all datasets. For example, in the Gun-Point dataset the D-TREE, SVM, KNN has achieved 11.33% error rate, while the proposed FA-D-TREE, SVM, KNN obtained 00.08% of error rate. It is due to capability of the FA which incorporated into D-TREE, SVM, KNN to find the optimal weights for the D-TREE, SVM, KNN and consequently increase the performance of the D-TREE, SVM, and KNN. This is believed that fireflies come together more closely around the optimal solution. In other words, it has good exploitation capability and can find better solutions as many candidates (fireflies) are gathered near optimal solution.

S. Suganya and N. Kamalraj, "Meta classification technique for improving credit card fraud detection," *International Journal of Scientific and Technical Advancements*, Volume 2, Issue 1, pp. 101-105, 2016.

B. Comparison with state-of-the-Art

Table II shows the comparison of the results of FA-D-TREE, SVM, KNN and other available approaches in terms of error rate classification using credit card datasets. The best results are presented in bold.

The experimental results indicate that the proposed hybrid method (FA-D-TREE, SVM, KNN) outperforms other approaches on credit card datasets. FA-D-TREE, SVM, KNN is able to classify the Wafer with error rate of 0.004%. This capability is supported by the feature of the attractiveness i.e., the density of the light that caused the fireflies to be brighter (is determined by the value of the objective function) and attract to the location of near optimal solutions.

C. Experimental Results

Table II presents the comparison of the error rate (%) between FA-D-TREE, SVM, KNN and D-TREE, SVM, KNN transaction data classification techniques with credit card datasets.

TABLE II: Comparison of algorithms error rate

Algorithm	Instance#	Feature#	Proposed algorithm	Actual	Predicted
KNN	2424	100	73.15±7.41	61.38±5.09	61.89±4.11
D-Tree	2397	617	90.56±1.02	89.77±1.02	90.01±1.03
SVM	2000	500	67.72±3.36	55.63±3.29	55.1±3.47
Firefly (Ensemble)	2000	649	96.11±1.3	97.9±0.9	97.9±0.92

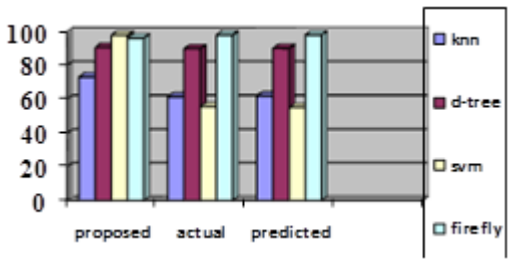

Fig. 2. Comparison of error rate between algorithms.

Note#: Instance is the number of rows of data. Feature is the initial result of the algorithm. Proposed Algorithm indicates the result obtained by ensemble of D-Tree with FA, KNN with FA, FA with SVM. Actual column shows the actual classes in the selected instance of data. Predicted shows the results obtained by the ensemble algorithms with FA.

V. Conclusion and Future Work

In this work an ensemble method based on the D-TREE, SVM, KNN and FA is proposed for solving transaction data classification problems. Initial classification results are generated at random instances of data using D-TREE, SVM, KNN and the improvement is carried out by the FA that tries to optimize the weights of the D-TREE,SVM,KNN using ensemble mechanism. Experiments results using benchmark CREDITCARD transaction data sets show that the proposed FA-D-TREE, SVM, KNN outperforms the D-TREE, SVM, KNN on all dataset instances. Further comparison with other approaches in the literature shows that the ensemble method is able to minimize the error rate with new best results on instances. As an extension of this study, further investigation will be devoted to validate the hybridization between FA with local search algorithm for the purpose of creating a balance between the exploration and exploitation during the optimization process and to avoid the premature convergence.

References

[1] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: bagging, boosting, and variants," *Machine Learning*, vol. 35, pp. 1-38, 1999.

[2] C. E. Brodley, "Recursive automatic bias selection for classifier construction," *Machine Learning*, vol. 20, pp. 63-94, 1995.

[3] T. Dietterich, "Ensemble methods in machine learning," *First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science*, Springer-Verlag, pp. 1-15, 2000.

[4] G. Brown, J. L. Wyatt, and P. Tino, "Managing diversity in regression ensembles," *The Journal of Machine Learning Research*, vol. 6, pp. 1621-1650, 2005.

[5] A. Sharkey, *Combining Artificial Neural Networks: Ensemble and Modular Multi-Net Systems*, Springer-Verlag, pp. 1-30, 1999.

[6] R. Patidar and L. Sharma, "Credit card fraud detection using neural network," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 1, issue-NCAI2011, 2011.

[7] Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support vector machines," *Proceedings of the International Multiconference of Engineers and Computer Scientists*, Hong Kong, vol. I, 2011.

[8] K. R. Seeja and M. Zareapoor, "FraudMiner: A novel credit card fraud detection model based on frequent item set mining," *The Scientific World Journal*, Hindawi Publishing Corporation, vol. 2014, article ID 252797, pp. 1-10, 2014.

[9] J. Pun and Y. Lawryshyn, "Improving credit card fraud detection using a meta-classification strategy," *International Journal of Computer Applications*, vol. 56, no. 10, pp. 41-46, 2012.

[10] A. Thakur, B. Shaikh, V. Jain, and A. M. Magar, "Hidden markov model in credit card fraud detection," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, issue 2, pp. 997-1000, 2015.

[11] S. Maes, K. Tulys, B. Vanschoenwinkel, and B. Manderick, "Credit card fraud detection using bayesian and neural networks," Vrije Universiteit Brussel–Department of Computer Science, Computational Modeling Lab (COMO), Pleinlaan 2, B-1050 Brussel, Belgium.

S. Suganya and N. Kamalraj, "Meta classification technique for improving credit card fraud detection," *International Journal of Scientific and Technical Advancements*, Volume 2, Issue 1, pp. 101-105, 2016.