International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 1 | Issue – 6

Credit Card Fraud Detection using Fire Fly Algorithm

S. Senthil Kumar
Assistant Professor
Dr. SNS Rajalakshmi College of Arts And Science (Autonomous), Coimbatore, Tamil Nadu, India

Ms. D. Nivya
PG Student
Dr. SNS Rajalakshmi College of Arts And Science (Autonomous), Coimbatore, Tamil Nadu, India

ABSTRACT

Data Mining or Knowledge Discovery is needed to make sense and use of data. Knowledge Discovery in data is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data [1]. Data mining consists of more than collection and managing data; it also includes analysis and prediction. People are often do mistakes while analyzing or, possibly, when trying to establish relationships between multiple features. This makes it difficult for them to find solutions to certain problems.

Classification models predict categorical class labels; and prediction models predict continuous valued functions. For example, we can build a classification model to categorize bank loan applications as either safe or risky, or a prediction model to predict the expenditures in dollars of potential customers on computer equipment given their income and occupation.

Keywords: *Data Mining, Credit card fraudulent, Classification*

INTRODUCTION

The feature selection process is embedded into a classification algorithm, in order to make the feature selection process sensitive to the classification algorithm. This approach recognizes the fact that different algorithms may work better with different features. The strategy of combining the classifiers generated by an induction algorithm. The simplest combiner determines the output solely from the outputs of the individual inducers.

WORKING OF CLASSIFICATION

With the help of the bank loan application that we have discussed above, let us understand the working of classification. The Data Classification process includes two steps,

- Building the Classifier or Model
- Using Classifier for Classification

Building the Classifier or Model

This step is the learning step or the learning phase. In this step the classification algorithms build the classifier. The classifier is built from the training set made up of database tuples and their associated class labels. Each tuple that constitutes the training set is referred to as a category or class. These tuples can also be referred to as sample, object or data points.

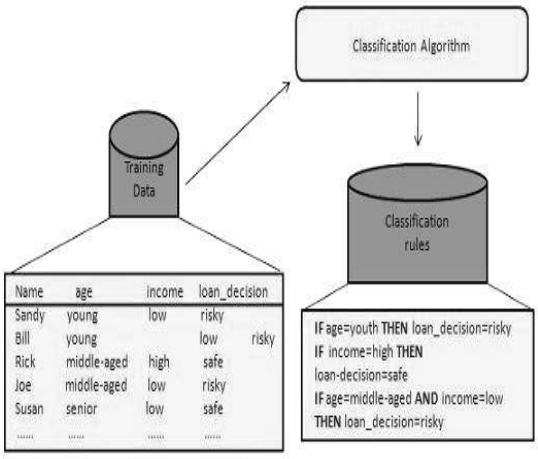

Fig: (a) Classifier Builder

Using Classifier for Classification

The classifier is used for classification. Here the test data is used to estimate the accuracy of classification rules. The classification rules can be applied to the new data tuples if the accuracy is considered acceptable.

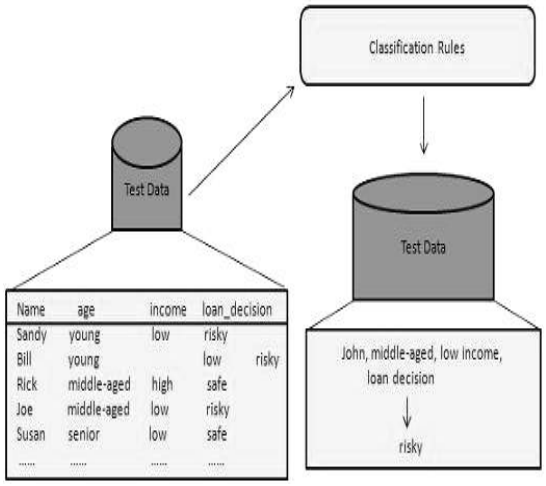

Fig: (b) Classifier Usage for Classification

PROBLEM DEFINITION

The studies on the use of transaction data classification in actual applications is lacking in the literature. The dataset for credit card applications is minimum and not available more. The main problem is to find whether the transaction in the dataset is fraudulent or not. Only single classifier is used for detecting fraudulent transaction in existing literature but in the proposed work three base classifiers and one meta classifier is used.

METHODOLOGY

The main idea of ensemble methodology is to combine a set of models, each of which solves the same original task, in order to obtain a better composite global model, with more accurate and reliable estimates or decisions than can be obtained from using a single model. The idea of building a predictive model by integrating multiple models has been under investigation for a long time. The ensembles techniques are divided into two main categories are Decision optimization and Coverage optimization.

The proposed model is trained with few transactions so that it will be easier to detect frauds and which is further developed with corrections for future references to efficiently detect the fraud. The main aim of this proposed method is to improve the classification accuracy.

CREDIT CARD DATASET CLASSIFICATION USING KNN

The transaction date is taken as a feature for classification. The general goal is to make accurate predictions about unknown data after being trained on known data. Data comes in form of examples with the general form: w1, .., wn are also known as features, inputs or dimensions v is the output or class label. Both wi and vs can be discrete (taking on specific values) {0, 1} or continuous (taking on a range of values) [0, 1].

In training we are given (w1, .., wn, v) tuples. In testing (classification), we are given only (w1, .., wn) and the goal is to predict v with high accuracy.

FRAUD DETECTION

Pattern matching is not necessarily to be exact rather small variations can be accepted and if there exists big difference in pattern, then chances that particular transaction is illegal transaction is more. The output of neural network will be in between 0 and 1. If the output is below .6 or .7 it implies transaction legal and if output is above .7 then probability of a illegal transaction is high. in some occasions legal users may make transaction that will be quite different and sometimes fraudster make transactions that matches the pattern trained by neural network. Due to limitation problems, legal users will use card for

limited amount but fraudster will try to do big purchase before the action taken by the credit card holder which will be a mismatch with the trained pattern by neural network. The process of business will be present always in neural network pattern recognition systems design. History descriptors provide details usage details of card and payments made. Other descriptors have information about date if issue and so on [1].

WORKING PRINCIPLE (PATTERN RECOGNITION)

Neural network based fraud detection is similar to the human brain working. Neural network made a computer to think as human brain that learns through past experience. The learning experience or knowledge is used to solve and make decision in problems in day today life. The same method is for credit card fraud detection. The consumer use fixed pattern of credit card use. This pattern is taken for past one or two years to train a neural network. The different other categories of information can also be stored like location for kids purchase, frequencies of huge purchase and so on in limited time. Neural network trains the various faces of credit card fraud along with credit card usage pattern which is provided by bank. Credit card usage pattern is taken by the prediction algorithm to differentiate fraudulent and non-fraudulent. Unauthorized user's pattern is matched with original card holder's pattern which is trained by neural network, and if pattern is same the decision made as genuine transaction.

BASED ON FREQUENT ITEM SET MINING

K. R. Seeja and Masoumeh Zareapoor proposed a credit card fraud detection model that detects fraud from highly imbalanced and anonymous credit card transaction datasets.

Frequent item set mining is used for finding legal and illegal patterns of transactions which handles the imbalance problem in class. To find whether the incoming transactions of the customers belongs to legal or illegal pattern , a matching algorithm is proposed and according that transaction closer to the patterns are identified and decisions are made. No special attention on attributes is given to manage the anonymous nature of transaction data and every attribute is treated equally for pattern finding. On UCSD Data Ming Contest 2009 Dataset, Evaluation of performance for this model is done and found to have less false alarm rate compared to state of the art classifiers, rate of fraud detection is high, classification rate is balanced, Matthews correlation coefficient.

ENSEMBLE CLASSIFIER - FIREFLY

The firefly algorithm follows three rules:

- Fireflies must be unisex.
- Lighter firefly is attracted towards the randomly moving brighter fireflies.
- The brightness of every firefly symbolizes the quality of the solutions.

The diversity in the Firefly Algorithm (FA) optimization is depicted by the random movement component, whilst the intensification is unconditionally manipulated by the attraction of various fireflies and the strength of attractiveness. As opposed to the other meta-heuristics, the association between exploration and exploitation in FA are relatively inter-connected; this might be a significant factor for its success, in solving multi-objective and multi-modal optimization problems.

```
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
                    Calculate the distance between each fireflies i
                        and  j
                    Move all firefly (xi) to the best solution (xj)

                End if
            End for j
        End for i
        Moves best solution randomly
        Find the best solution from the new population
    End while
Return best
End of the algorithm
```

At the first, it generates the initial population of candidate solutions for the given problem (here, the weights of the D-TREE, SVM, and KNN. After that, it calculates the light intensity for all fireflies and finds the attractive firefly (best candidate) within the population. Then, calculate the attractiveness and distance for each firefly to move all fireflies towards the attractive firefly in the search space. Finally, the attractive firefly moves randomly in the search space. This process is repeated until a termination criterion is met i.e., the maximum number of generations is reached.

Normally, the quality of the transaction is measured based on the error rate that can be calculated based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

The procedure starts from an initial population of randomly generated individuals. The quality of each individual is calculated using Eq. (1) and the best solution among him or her is selected.

CONCLUSION

In the above paper we discussed potential usefulness of ensemble methods; it is not surprising that a vast number of methods are now available to researchers and practitioners. There are several factors that differentiate between the various ensembles methods. For future work, we can introduce Machine Learning Algorithms for Class Imbalance problems.

Machine learning can often be successfully applied to these problems, improving the efficiency of systems and the designs of machines. There are several applications for Machine Learning (ML), the most significant of which is data mining.

REFERENCES

[1]Opitz D. and Shavlik J., Generating accurate and diverse members of a Neural network ensemble. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, Advances in Neural Information Processing Systems, volume 8.

[2]Gary M.Weiss, Bianca.Zadrozny, Maytal.Saar-Tsechansky, "Guest editorial: special issue on utility-based data mining". Data Mining Knowledge Discovery 17:129–135, (2008).

[3]Raghavendra Patidar, Lokesh Sharma, "Credit Card Fraud Detection Using Neural Network"International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011.

[4]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 2011 Vol I, IMECS 2011, March 16 -18, 2011,Hong Kong.

[5]K. R. Seeja and Masoumeh Zareapoor, "FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining", Hindawi Publishing Corporation ,The Scientific World Journal , Volume 2014, Article ID 252797, 10 pages,http://dx.doi.org/10.1155/2014/252797.