International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Credit Card Fraud Detection System Using Smote Technique and Whale Optimization Algorithm

Sahayasakila.V, D. AishwaryaSikhakolli, Venkatavisalakshiseshsai Yasaswi

Abstract*: Credit cards are now being widely used all over the world for transactions, irrespective of the geographical boundaries. Therefore the range of theft and fraud transactions have widely increased. In order to detect both the fraud and non-fraud transactions Credit card fraud detection system (CCFD) is used. Credit card fraud detection system is proposed using machine learning techniques .The two main important algorithm techniques used in this system are whale optimization algorithm(WOA) and Smote(synthetic minority oversampling technique). The Smote technique is used to solve Class imbalance problem. The Whale optimization algorithm comprises mainly of three operators which are used to stimulate the search for prey, encircling prey and bubble-net scratch around the behaviour of humpback whales. It is also used to increase the efficiency of the credit card fraud detection system. The Smote technique is used to solve Class imbalance problem. Thus by using the SMOTE technique and Whale optimization algorithm credit card fraud detection system solves the problem of data imbalance, reliability, data optimization, and improvises the convergence speed.

Keywords:* credit card fraud detection, whale optimization algorithm, smote technique, kaggle datasets.

I. INTRODUCTION

Credit card fraud is a wide term for frauds dedicated involving a payment card such as credit card as a corrupted source of funds in transactions. To detect these criminal sources of transactions Credit card fraud detection system is used. There is a tremendous increase in the amount of fraud transactions taking place all over the world. Fraud transactions are the fake transactions done through credit cards without the knowledge of the original card holders. Fraudulent transactions have been done in private merchandise, online transactions, bill payments, etc. Thus credit card fraud has been a serious issue now a days that has to be solved immediately. The credit card fraud detection system is initiated by using machine learning .

The two main algorithms used in the machine learning areSMOTE(Synthetic minority oversampling technique) algorithm and the Whale optimization algorithm.

SMOTE(synthetic minority oversampling technique) solves the Machine learning problem of one class dominating the other class. This problem is termed as class imbalance .

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
Ms.Sahayasakila, V, Assistant ProfessorK. Kavya Monisha Department Of Cse Srmist Ramapuramdepartment OF CSE SRMIST Ramapuram.
D. Aishwarya Sikha kolli, DEPARTMENT OF CSE SRMIST Ramapuram
Venkatavisalakshiseshsai Yasaswi, DEPARTMENT OF CSE SRMIST Ramapuram

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

Class imbalance occurs when the occurrence of one class outnumber the occurrence of the other classes. SMOTE differentiates the instances and synthesizes the minority instances between the real instances. The SMOTE ( ) functions are used to shrink the minority instances from the real instances. The synthesis of instances are performed using SMOTE ( ) function parameters. The whale optimization algorithm is proposed based on the simulation of the behaviour of whales . The way how the whales prey ,attacks ,encircle and other processes helps in achieving whale optimization. The WOA is disdainful in solution accuracy. to improvise the convergence speed in the detection of fraud transaction the whale optimization algorithm is used. WOA increases the efficiency of the system.A detailed view of the system is explained in the following topics.

II. PROPOSED SYSTEM

The Credit card fraud detection system is initiated for detecting the fraud transactions from the number of transactions made by the card holders. The transactions done by credit card holders are derived in the form of kaggle datasets. Kaggle datasets are nothing but data that are already being posted by the companies and researchers for the purpose of machine learning and data mining.

SMOTE(Synthetic minority oversampling technique) is a machine learning technique used for classification of data. The kaggle datasets are trained by using the SMOTE technique.SMOTE technique is used to solve data imbalance problem. Using the smote technique the data, which is nothing but the transactions are trained .This technique is mainly used to differentiate the fraud transactions from the original transactions done by the card holders. Initially the transaction data are stored in a confluence form. Thus the confluence data have been trained by the SMOTE technique to synthesize the fraud transactions from the non fraud transactions. The synthetic minority oversampling technique shrinks the fraud transaction from the non-fraud transactions. The SMOTE( )function parameters synthesize the confluenced transactions.

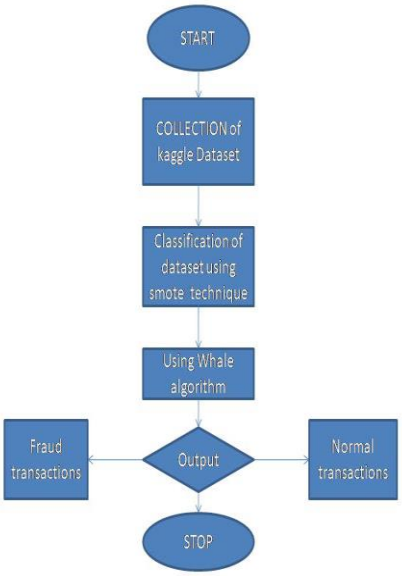


Fig.1 Flowchart

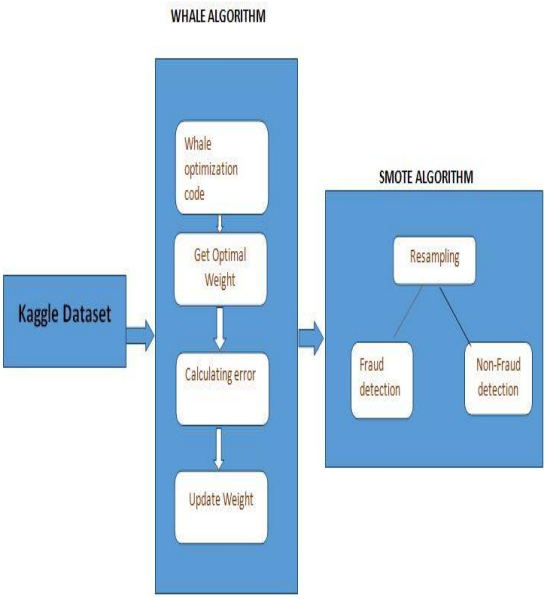


Fig.2 system architecture

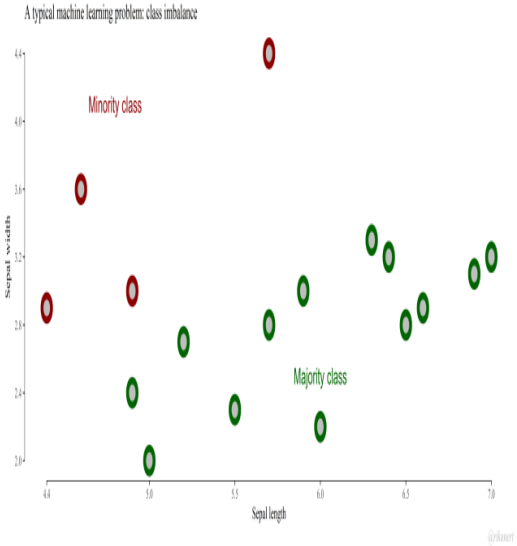


Fig.3 Smote technique

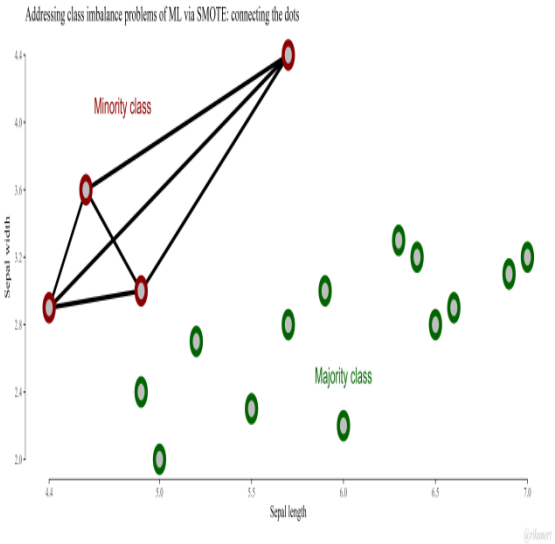


Fig.4 synthesis of minority class

As shown in the figure 3, the minority class refers to the fraud transaction that are detected from the datasets. The majority class refers to the non-fraud transactions in the datasets that are differentiated from the fraud transactions.

The smote technique synthesizes all the fraud transactions from the original non-fraud transactions as shown in the figure4.

The synthesized transactions are again resampled to check the data accuracy. The synthesized fraud transactions are optimized by using whale optimization algorithm. The WOA is disdainful in solution accuracy and optimization.TheWOA improvises the convergence speed and reliablity. The whale optimization code gets the optimal weight of the transaction. Whale algorithm increases the efficiency of the system. The WOA formula optimizes the fraud transactions, detects the error and updates the weight. Structure of whale optimization algorithm is shown in Figure2.

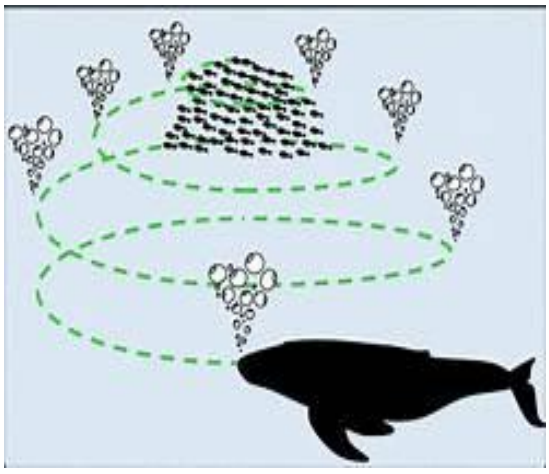


Fig.5 Whale optimization algorithm

III. WHALE OPTIMIZATION FORMULA

The WOA is a new probing optimization algorithm influenced by whale hunting. The main algorithm is as follows:

© Copyright: All rights reserved.

- For surrounding the Prey

$$D=|CY*(t)-Y(t)|$$
$$Y(t+1)=Y*(t)-AD$$

- Hunting Behavior

$$Y(t+1)=Y*(t)-AD, p<Pi$$
$$Y(t+1)=Y*(t)+DPeblcos(2\pi l), p\geq Pi$$

- Searching prey

$$D=|CYrand-Y(t)|$$
$$Y(t+1)=Yrand-AD$$

IV. KAGGLE DATASETS

The kaggle is an online community that allows the user to find and publish the datasets. The datasets used in the CCFD system contains transactions made by credit cards by credit card holders.

V. ARCHITECTURAL DESIGN

The architectural design of credit card fraud detection system mainly comprises of two parts training the data and testing the data. The training part is divided into two sub processes . The first process is that the card holders transaction amount is converted into observation symbols to calculate threshold from the sequence of amount. The second process is clustering . clustering groups the data into clusters. This helps in unsophisticated data retrieval . clustering is technique used for splitting the data into related components. Thus orders and patterns are visible. The data is tested using the smote technique. The smote technique is used to do minor sampling of the datasets. The is inferred in the form of receiver operating characterstics. Confusion matrix table is used to illustrate the performance of the classifier, over a set of test dataset for which the true values are determined. The architectural flow of the CCFD is shown in the fig. 6

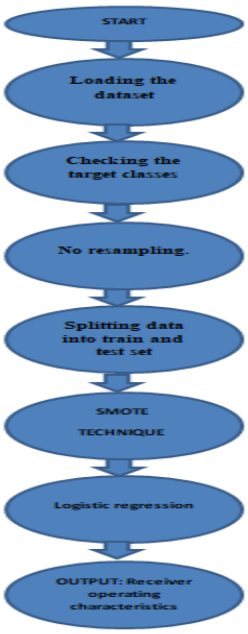


Fig.6 algorithmic design

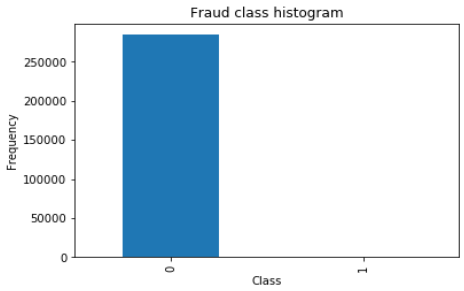


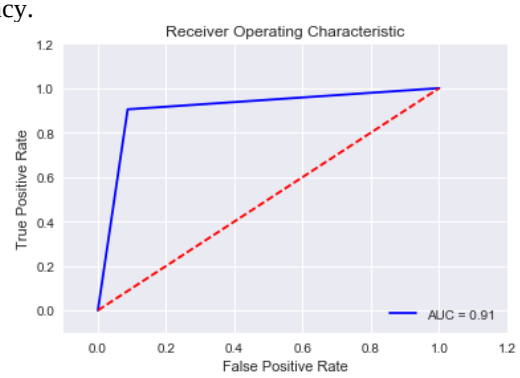
Fig.7 fraud data histogram

VI. EXISTING SYSTEM

Since the credit card fraud detection system is a highly researched field, there are many different algorithms and techniques for performing the credit card fraud detection system. One of the earliest system is CCFD system using markov model. Some other various existing algorithms used in the credit cards fraud detection system includes Cost sensitive decision tree (CSDT), support vector machine (SVM), Random forest,etc. credit card fraud detection(CCFD) is also proposed by using neural networks. The existing credit card fraud detection system using neural network follows the whale swarm optimization algorithm to obtain an inceptive value. It the uses BP network to rectify the values which are found error. All of these techniques has some serious disadvantages such as decreasing accuracy levels, lack of efficiency, sometimes classifying the normal transactions as fraud transactions and vise versa. These disadvantages are overcomed in this credit card fraud detection system using whale algorithm and smote technique.

VII.EXPERIEMENTAL RESULT AND ANALYSIS

The experimental analysis of Credit card fraud detection system using whale optimization and SMOTE technique is observed to be much more efficient than BP neural networks. The problem of class imbalance is overcome using smote technique. Thus by using the simulation of the behaviour of the whales the convergence speed is improvised and the efficiency of the system is increased , more than that of the existing system. The system differentiates all the fraud transactions from the authentic transactions. Thus the module is disdainful in solution accuracy.



Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)
© Copyright: All rights reserved.

VIII. CONCLUSION

Credit card fraud detection system using whale optimization algorithm and SMOTE (Synthetic minority optimization technique) aims in indentifying the fraud transactions occurring during the transactions made by the card holder. The system also aims to improve the convergence speed and solves the data imbalance. The receiver operating characterstics(ROC) shows that the relation between the true positive rate and false positive rate.

REFERENCES

1. D. P. Deepti, M. K., Sunita, M. W. Vijay, J. A. Gokhaleand S. H. Prasad, Computer Science and Network Security, vol. 10, no. 8, (2010).
2. S. O. Falaki, B. K. Alese and W. O. Ismaila, Practical Mathematics and Computing, vol. 1, no. 2, (2010).
3. S. Esakiraj and S. Chidambaram, "A predictive approach for fraud detection using hidden markov model" International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 1, January-2013 C Suman ndNutan "Review paper on credit card fraud detection", International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 7–July 2013.
4. V. Bhusari, and S. Patil, "Study of Hidden Markov Model in Credit Card Fraudulent Detection", International Journal of Computer Applications (0975 – 8887) Volume 20–No.5, April 2011.
5. P. Jayant, Vaishali and D. Sharma "Survey on Credit Card Fraud Detection Techniques" International Journal ofEngineering Research & Technology (IJERT) Vol. 3 Issue 3,March – 2014.
6. Ghosh, S., and Reilly, D.L., 1994. Credit Card Fraud Detection with a Neural Network, 27th Hawaii International l Conference on Information Systems, vol. 3 (2003), pp. 621- 630.
7. Syeda, M., Zhang, Y. Q., and Pan, Y., 2002 Parallel Granular Networks for Fast Credit Card Fraud Detection, Proceedings of IEEE International Conference on Fuzzy Systems, pp. 572-577 (2002).
8. Stolfo, S. J., Fan, D. W., Lee, W.,Prodromidis, A., and Chan, P. K., 2000. CostBasedModeling for Fraud and IntrusionDetection: Results from the JAM Project,Proceedings of DARPA Information SurvivabilityConference and Exposition, vol. 2 (2000), pp.130-144.

© Copyright: All rights reserved.