

Analysis on Credit Card Fraud Identification based on KNN and Outlier Detection

¹Darshna Parmar, ²Tanaya Ganguly

¹Assistant Professor, ²Assistant Professor

^{1,2}Department of Computer Science & Engineering,

^{1,2}Parul Institute of Engineering & Technology, Parul University, Vadodara, India.

Abstract: Popular payment mode which is credit card is acquired both offline and online that provides us with its great advantage of cashless transaction. A credit card is a convenient financial product that can be used for everyday purchases such as gas, groceries, and other goods and services. It becomes very easy, suited and groovy to make payments and other transactions through credit card. Credit card fraud is also growing along with the development in technology. Along with improvement in the global communication the economic fraud is remarkably increasing in the global communication. It is being recorded every year that the loss due to these fraudulent actions is billions of dollars. These activities are carried out so gracefully which looks similar to authentic transactions. Therefore, to have a systematized method of fraud detection has become a need for all banks in order to minimize heu and cry and bring order in place. Thus techniques of KNN and outlier detection are implemented to optimize the best result for the fraud detection problem. These methods are proved to slash the false alarm rates and enlarge the fraud detection rate. KNN and Outlier detection are quiet familiar area of research. Outlier Detection is absolutely important task in various application domains. Earlier outliers were considered as noisy data and now it has become severe difficulty in various areas of research. The discovery of outlier is useful in detection of unpredicted and unidentified data in certain areas like fraud detection of credit cards, calling cards, discovering computer intrusion and criminal behaviors etc. Whereas KNN (k-Nearest Neighbor) is used for classification because of its interpretation and low calculation time.

IndexTerms - Credit card fraud, KNN, Outlier detection.

I. INTRODUCTION

Due to popularity of online shopping, credit card fraud becomes one of the biggest issue for customers. According to an ACNielsen study conducted in 2005, one-tenth of the world's population is shopping online [1]. Retailers like Wal-Mart typically handle much larger number of credit card transactions including online and regular purchases [3]. In past, people was only doing online transaction via net banking. But Nowadays, there are various apps available like paytm, google pay, Square cash, Zelle, various UPIs and etc. In General, customers are using online transaction because of good discount, cashback and fastest way of completing transaction. Mostly, customers using online mode of transaction for shopping purpose. There are two ways for shopping physical card online and virtual card online. In physical card online, theft steal credit card and set up signature for purchase. In virtual card online, attackers steal some confidential data of credit card like card number, expiry date and cvv etc. Due to credit card hacking customer will loss valuable good and money. Hence, an effective system is required to detect fraudulent. There are various machine learning techniques are introduced to find fraudulent. Normally, KNN, SVM, HMM and GA techniques are used. In this paper, we have surveyed credit card techniques using KNN and outlier detection.

II. LITERATURE SURVEY

Detection of fraud is quite a complex task, there is no such system which will correctly predicts any fraudulent.

The properties for a good fraud detection system are:

1. Should accurately identify the frauds.
2. Should detect the frauds quickly.
3. Should not misjudged a genuine transaction as fraud.

Techniques used in fraud detection can be divided into two:

- 1) Supervised techniques - Supervised techniques where past known fraud cases are used to build a model which will produce a suspicion score for the new transactions.
- 2) Unsupervised techniques - Unsupervised are those where there are no prior sets in which the state of the transactions is known to be fraud or legitimate.

2.1 Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering [3].

The author uses the ranges of transaction amount as an attribute in the HMM. The author has suggested method for finding the spending profile of cardholders. It is also discussed how the HMM can identify the fraudulent transactions.

Advantage:

- The detection of the fraud use of the card is found much faster that the existing system.
- In case of the existing system even the original card holder is also checked for fraud detection.
- We can find the most accurate detection using this technique.
- This reduces the tedious work of an employee in the bank.
- Scalable for large data set as well.

Dis-advantage:

- It has large number of unstructured parameters.
- HMM cannot express dependencies between hidden states.

2.2 Credit Card Nearest Neighbor Based Outlier Detection Techniques [4].

KNN algorithm and outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These approaches are proved to minimize the false alarm rates and increase the fraud detection rate. Any of these methods can be implemented on bank credit card fraud detection system, to detect and prevent the fraudulent transaction.

Advantage:

- The K-Nearest Neighbor (KNN) Classifier is a very simple classifier that works well on basic recognition problems.

Dis-advantage:

- The main disadvantage of the KNN algorithm is that it is a lazy learner.
- The main disadvantage of this approach is that the algorithm must compute the distance and sort all the training data at each prediction, which can be slow if there are a large number of training examples.
- Another disadvantage of this approach is that the algorithm does not learn anything from the training data, which can result in the algorithm not generalizing well and also not being robust to noisy data.

2.3 Fraud Detection of Credit Card Payment System by Genetic Algorithm [5].

It creates a test data and through which the fraudulent activities are detected. This algorithm is also called as an optimization technique based on genetic and natural selection in high computational problems. The author proposes a method to detect credit card fraud and the results are validated using principles of this algorithm. The purpose of detecting fraud cases is to declare it to the client and the service provider.

Advantage:

- It can find fit solutions in a very less time.
- The random mutation guarantees to some extent that we see a wide range of solutions.

Dis-advantage:

- It might not find the most optimal solution to the defined problem in all cases.
- It is also hard to choose parameters like number of generations, population size etc. When we are working even though our heuristic was right we were not realizing it because we were running for a fewer generation.

2.4 Analysis on Credit Card Fraud Detection Methods [6].

Proposed a fraud detection method which involves monitoring the activities of populations to observe and predict undesirable behavior. Undesirable behavior is a set of several habits like intrusion, fraud, delinquency and defaulting. This research speaks on several credit card fraud detection and telecommunication fraud and different techniques which help in resolving the discussed problems.

Advantage:

- The paper explains us the various methods of fraud detection like Hidden Markov Model, Genetic Algorithm, Decision tree, Support vector machine, Neural Network which gives us the clear idea to use a specific algorithm for fraud detection.

Dis-advantage:

- Not given any algorithm in deep.

2.5 Credit card fraud detection using anti-k nearest neighbor algorithm [2].

The author uses concept of data stream outlier detection algorithm which is based on anti k-nearest neighbors for credit card fraud identification. Whereas traditional methods need to scan the database many times to find the fraudulent transaction, which is not suitable for data stream surroundings. This method makes easier to stop fraudulent transaction happens by Lost and stolen card and Credit card validation checks and detects errors in a sequence of numbers which also helps to detect valid and invalid numbers easily.

Advantage:

- Lost and stolen card feature makes it easier to stop fraudulent transactions
- Credit card validation checks detects errors in a sequence of numbers hence detects valid and invalid numbers easily.

Dis-advantage:

- Needs to improve outlier detection speed.

III. COMPARATIVE ANALYSIS TABLE

Sr. No	Paper Name	Year	Algorithm Used	Advantage	Disadvantage	Future Scope
1.	Card Fraud Detection System Using Hidden Markov Model and K Clustering. ^[4]	2014, IJAR C	Hidden Markov Model (HMM).	Scalable Faster Accurate	Unstructured parameters. Dependent	To take more amount of observation of card holder and then consider the states of HMM.

2.	Credit Card Nearest Neighbor Based Outlier Detection Techniques. ^[5]	2018, IJCT	K-Nearest Neighbor (KNN).	Very simple classifier that works well on basic recognition problems.	Not Robust Time Consuming	Detecting the presence of outliers from a large amount of data via an online updating technique.
3.	Fraud Detection of Credit Card Payment System by Genetic Algorithm. ^[6]	2012, IJSER	Genetic Algorithm.	Faster. Wide range of solution.	Do not find the best solution. Hard to choose parameter.	It aims to obtain the better and optimal solutions.
4.	Analysis on Credit Card Fraud Detection Methods. ^[7]	2014, IJCTT	Brief about Decision trees, Support vector Machine, Neural Network, Genetic Algorithm, Hidden Markov Model.	Clear idea to use a specific algorithm for fraud detection.	Not given any algorithm in deep.	Applying any combination of algorithm into bank credit card fraud detection system to detect fraud.
5.	Credit card fraud detection using anti-k nearest neighbor algorithm. ^[8]	2012, IJCSE	Anti K-Nearest Neighbor (KNN).	Easier to stop fraudulent transactions. Detects valid and invalid numbers easily.	Needs to improve outlier detection speed.	To improve the detection of outlier in very less time so that the fraud of credit card is detected in early time.

TABLE 1: Comparative Analysis Table.

IV. CONCLUSION

The scam of credit card has become much more immense. In order to headway with safety measures of the monetary transaction systems in a persistent and effectual way, structure a precise and well-organized credit card scam detection system is one of the essential functions in today's world. By performing over sampling and extracting the principal direction of the data we can use our KNN method to determine the anomaly of the target instance. Hence the KNN method can ensemble for detecting fraud with the limitation of memory. By the mean time outlier detection mechanism helps to detect the credit card fraud using less memory and computation requirements. Especially outlier detection works fast and well on online large datasets. But compared with power methods and other known anomaly detection methods, experimental results prove that the KNN method is more accurate and coherent.

REFERENCES

- [1] "Global Consumer Attitude Towards On-Line Shopping," http://www2.acnielsen.com/reports/documents/2005_cc_online_shopping.pdf, Mar. 2007.
- [2] Venkata Ratnam Ganji," Credit card fraud detection using Anti-k Nearest Neighbor Algorithm",International Journal on Computer Science and Engineering (IJCSE) Vol. 4 ,06 June 2012,(1035-1039).
- [3] Abhinav Srivastava, Amlan Kundu, Shamik Sural, "Credit Card Fraud Detection Using Hidden Markov Model", IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 5, NO. 1, JANUARY-MARCH 2008.
- [4] Mrs.C.Navamani, S.Krishnan," Credit Card Nearest Neighbor Based Outlier Detection Techniques" Volume 5 Issue-2, Mar-April 2018.
- [5] K.RamaKalyani, D.UmaDevi" Fraud Detection of Credit Card Payment System by Genetic Algorithm" Volume 3, Issue 7, July-2012.
- [6] Renu, Suman" Analysis on Credit Card Fraud Detection Methods" volume 8 number 1– Feb 2014.