International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

Advancement in Fraud Detection Mechanism of Credit Application

Mr. A. J. Shakadwipi
[1]*Department of Computer Engineering, SNJB's COE Chandwad, amolshakadwipi@gmail.com*

Abstract—Due to market uncertainties, declining economic growth and significant growth of online e-commerce makes fraud widespread. Rapid advancement in the electronic commerce technology, the use of credit cards has increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of credit card fraud also rising. The applications for these credit cards are based on internet or manual applications by the customers who require the smart cards and various loans. The applications in above cases found fraud and is a specific case of identity crime. Identity crime has emerged as a serious problem for credit card customers and banks. Fraudsters steal customer's identity and obtain credit cards. This posses a major threat to the customers and banks. The existing non-data-mining detection system of business rules and scorecards, and known fraud matching has limitations. To address these limitations this paper provides an approach in identifying fraudsters at the time of application submission i.e. in real-time. The paper presents a new multi-layer fraud detection system based on data-mining algorithms. The detection system utilizes the two algorithms: communal detection (CD) and spike detection (SD) which complements each other, improving the fraud detection systems accuracy, time and cost. The credit card application undergoes validations at the time of submission before issuing the card. Better account for varying legal actions, and remove the redundant attributes and to store the fraudulent datum in blacklist using CBR algorithm. CBR algorithm analysis using retrieval, diagnosis and resolution to make the data more secure and to find the fraudulent data.

Keywords- Communal Detection, Data Mining, Data mining-based fraud detection, Identity crime, Spike Detection.

I. INTRODUCTION

As Fraud Detection is imperative for every business or organization as it impacts by increasing the cost of doing their businesses. Identity theft is a form of stealing someone's identity in which someone pretends to be someone else by assuming that person's identity, usually as a method to gain access to resources or obtain credit and other benefits in that person's name. The victim of identity theft (here meaning the person whose identity has been assumed by the identity thief) can suffer adverse consequences if they are held responsible for the perpetrator's actions. Identity theft occurs when someone uses another's personally identifying information like their name, identifying number, or credit card number, without their permission, to commit fraud or other crimes. Credit application fraud is a specific case of identity crime [1]. Identity crime has become a more common approach as there is so much real identity data available on the Web, and confidential data accessible through unsecured mailboxes. It has also become easy for fraudster to hide their true identities. Customers fill the credit application form online or using manual paper based work. When online forms are not closed properly, or when the paper forms are not filled or discarded safely, it becomes easy for the fraudsters to misuse the customer information. It leads to the level of deciding the economy of the country.

Most of the businesses implemented intelligent analytical techniques to control fraudulent activities to reduce their increased cost. One of the intelligent techniques called data mining, substantiated to be a competitive tool for controlling fraudulent activities. Data Mining can extract valuable patterns

or knowledge from large volumes of data and alarm fraud. Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Alternatively, it can be defined as the process of data selection and exploration and building models using vast data stores to uncover previously unknown patterns [3]. Data mining is not new; it has been used intensively and extensively by financial institutions, for credit scoring and fraud detection. The data mining consists of multiple algorithms for detection [5]. Data mining algorithms are used in the online credit card application for counterfeit detection.

The main objective of this paper is to highlight the use of efficient data mining algorithms in credit application fraud detection system. The algorithms are used in this system is the spike detection and communal detection together.

Case-based reasoning (CBR) is now making a significant contribution to the task of crime detection. CBR systems are able to learn from sample patterns of smart card use to categorize new cases, and this method also has the promise of being able to adapt new patterns of crime as they come forward. The CBR system is the application of adaptive and hybrid culture systems. The CBR problems are previously considered too dynamic, chaotic, or complex to precisely model.

II. LITERATURE SURVEY

2005, Efstathios Kirkos et al. explores the effectiveness of Data Mining (DM) classification techniques in detecting firms that issue fraudulent financial statements (FFS) and deals with the identification of factors associated to FFS.

In 2009, G. Apparao et al. analyzes that the prevention is the best way to reduce frauds, fraudsters are adaptive and will usually find ways to circumvent such measures.

In 2011, Tatsuya Minegishi et al. focus on classification learning, which is an analytical method of stream mining.

In 2012, Sherly K.K et al. evaluates three classification methods to solve the fraud detection problems for data mining and shows how advanced techniques can be combined successfully to obtain high fraud coverage with maximum confidence and minimum false alarm rate. In 2012, Clifton Phua et al. observe that the credit application fraud is a specific case of identity crime. The existing non-data mining detection system of business rules and scorecards, and known fraud matching have limitations. To address these limitations and combat identity crime in real time, they propose a new multilayered detection system complemented with two additional layers: communal detection (CD) and spike detection (SD).

III. IMPLEMENTATION DETAILS

The system utilizes two layers called Communal Detection (CD) and Spike Detection (SD). The main contribution of this paper is to enhance secure transaction in credit card applications by using two new data-mining layers. These new layers improve detection of fraudulent applications because the detection system can detect various kinds of attacks, better account for changing legal behavior, and eliminate the redundant attributes.

3.1. Communal Detection

Communal Detection layer is based on white list-oriented approach. It utilizes fixed set of attributes. Communal Detection (CD) finds real social relationships to reduce the suspicion score, and is tamper- resistant to synthetic social relationships [1]. The CD algorithm matches all links against the white list to find communal relationships and reduce their link score. CD has a fundamental weakness in its attribute threshold. Specifically, CD must match at least three values for our data set. With less than three matched values, our white list does not contain real social relationships because

some values, such as given name and unit number, are not unique identifiers. The fraudster can duplicate one or two important values which CD cannot detect.
With this data stream perspective in mind, the CD algorithm matches the current application against a moving window of previous applications. It accounts for attribute weights which reflect the degree of importance in attributes. The CD algorithm matches all links against the white list to find communal relationships and reduce their link score. It then calculates the current application's score using every link score and previous application score. At the end of the current micro discrete data stream, the CD algorithm determines the SoA and updates one random parameter's value such that it trades off effectiveness with efficiency, or vice versa. At the end of the current Mini discrete data stream, it constructs the new white list.

CD Algorithm–

1. Every application value is compared against a list of previous application values to find the links.
2. Every application's link is matched against the white list to find communal relationships among applications and reduce their Link score.
3. Every previous application's score is to be included into the current application's score. Previous score acts as a baseline level.
4. Calculate every current application score using link and previous application's score.
5. The algorithm updates one random parameter's value such that there is a tradeoff between effectiveness with efficiency, or vice versa.
6. A new white list is constructed on the current Mini-discrete stream links.

3.2 Spike Detection
In contrast to CD, SD finds spikes to increase the suspicion score, and is probe resistant for attributes. Probe resistance reduces the chances a fraudster will discover attributes used in the SD score calculation. It is the attribute- oriented approach on a variable-size set of attributes. SD complements CD. The redundant attributes are either too sparse where no patterns can be detected, or too dense where no denser values can be found. The redundant attributes are continually filtered; only selected attributes in the form of not-too-sparse and not-too-dense attributes are used for the SD suspicion score. In this way, the exposure of the detection system to probing of attributes is reduced because only one or two attributes are adaptively selected [1].
From the data stream point-of-view, using a series of window steps, the SD algorithm matches the current application's value against a moving window of previous application's values. It calculates the current value's score by integrating all steps to find spikes. Then, it calculates the current application's score using all values scores and attribute weights. Also, at the end of the current Mini-discrete data stream, the SD algorithm selects the attributes for the SD suspicion score, and updates the attribute weights for CD.

SD Algorithm –

1. Every application value is compared against a list of previous application values step by step.
2. Calculate application's current value score by integrating the steps to find the spikes.
3. Calculate application's score using attribute weights.
4. Identify the key attributes to calculate the SD suspicion score. The final step updates the weights of the attributes.

3.3 CBR Algorithm—

Nearest neighbor matching is common to many CBR systems. Again using the basic exploratory facilities of CBR test bed, a set of cases which were considered to be very similar, above a certain percentage of similarity, were recovered. Applying the general principle of threshold retrieval, a multi-algorithmic approach to final match analysis was developed as a result of the design and testing of a variety of single discrimination algorithms. It has been suggested that no single algorithm may perform equally well on all search and classification tasks, and that an algorithm's improved performance in one learning situation may come at the expense of accuracy in another.

If a set of algorithms is asked to diagnose the set of cases retrieved for an unknown credit request, it is possible that the algorithms may disagree on the result, and resolution strategies were implemented to resolve the varying diagnoses into a single result.

IV. SYSTEM ARCHITECTURE

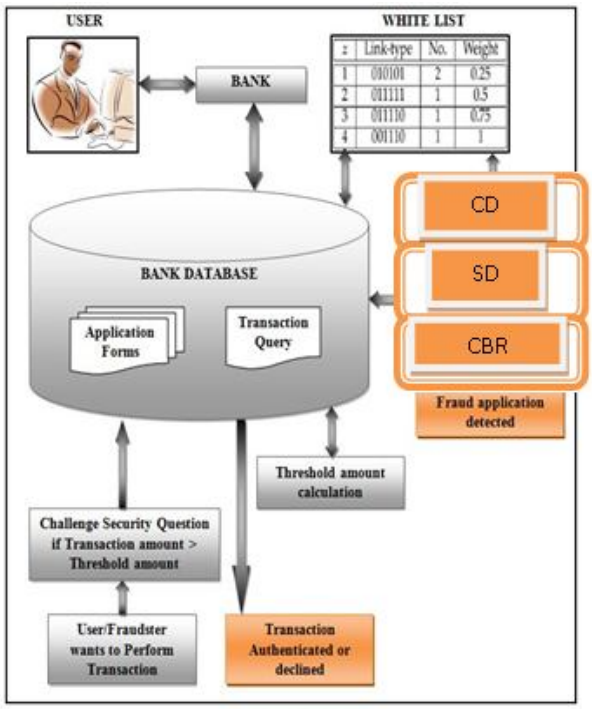

Fig 1.System architecture

The figure illustrates the system architecture-

The Description of System Architecture is given below [2] -Display GUI for entering credit card application details. Take input from user.

Compare new application with each other application in the bank database and assign a link type. The link type is nothing but a binary string (e.g. 01000101) in which "1" represents matched fields and "0" represents unmatched fields. Initial white list is created. The White list has list of verified applications, link type, number of applications corresponding to a particular link type and weight.

Communal Detection -

New Application is compared with windows of applications in the white list.CD layer finds communal relationships between the applications. If four or more fields are matched in the new application against application in the white list, then CD assigns less suspicious score. Otherwise the new application form is added into the white list and the list is updated. Suspicious score assigned to new application form is given as input to the SD layer.

Spike Detection –

Spike Detection (SD) layer verifies the matched fields for their priority. The unique ID fields are given higher priority.If unique IDs are matched then the suspicious score gets increased and the application form is declared as fraud and hence finally rejected.

If none of the unique IDs are matched then the application form is added into the white list and the list is updated.

Threshold Transaction Amount Calculation –

Based on the previous transactions made by the user the system calculates a threshold value of the transaction amount. The threshold value is nothing but average of all the previous transactions.

The CBR is used is the fraud detection system that the data is original or not that the data is original or not by retrieving the data from the blacklist verification. This method finds the fraudulent data by the artificial intelligence. The CBR algorithm involves with the data mining concept with match analysis.

Secure Transaction -

Now the fraudster or the legal user performs credit transaction. If the credit transaction amount is higher than the threshold, the fraudster or legal user is asked to challenge the security question. If the challenge is success i.e. in case of legal user the transaction is authenticated otherwise it is declined in case of fraudster. Hence the secure transaction is performed.

V. RESULT

The input data set for the system is a synthetic data set of 50,000 credit applications which is available at https://sites.google.com/site/cliftonphua/communal-fraud-scoring-data.zip.There are about 30 attributes are present in synthetic data set out of which only 19 most important identity attributes are selected for the processing. The Table 1 shows a sample data set of six credit applications with six attributes. The system will take input credit application for credit card from the user through online web portal and the application submitted by user act as a real-time entry to the system. The system will compare the application submitted by the user with applications in the synthetic data set and validate the application.

***Table 1:** SAMPLE DATA SET OF SIX CREDIT APPLICATIONS WITH SIX ATTRIBUTES*

I Or j	First Name	Last Name	Unit No.	Street Name	Home Phone No.	DOB
1	John	Smith	1	Circular Road	91234567	1/1/1982
2	Joan	Smith	1	Circular Road	91234567	1/1/1982
3	Jack	Jones	3	Square Drive	93535353	3/2/1955
4	Ella	Jones	3	Square Drive	93535353	6/8/1957
5	Riley	Lee	2	Circular Road	91235678	5/3/1983
6	Liam	Smyth	2	Circular Road	91235678	1/1/1982

Table ii :sample white list

z	Link type	No.	Weight
1	010101	2	0.25
2	011111	1	0.5
3	011110	1	0.75
4	001110	1	1

5.1 Result Set

The white-list is constructed form the input data set and a CD suspicious score is assigned to each application as a result of communal detection algorithm. The Table 2 shows the sample white-list constructed from credit applications in table 1. The Spike detection algorithm outputs the SD suspicious score. CD and SD scores are combined together to give a single score. SD updates the CD attribute weights. At last after calculating link type, CD Suspicious score, Multi-attribute score, the system gives the results as application is accept or reject after applying CD & SD algorithms on record.

VI. ACKNOWLEDGEMENT

Sincerely thank to Mr. Anurag Sangale ,my friends, researchers for providing us such helpful opinion, findings, conclusions and recommendations.

CONCLUSION

Main focus of this project is the detection of fraudsters in credit applications by implementing a new multilayered detection system complemented with the new data mining layers Communal Detection(CD) and Spike Detection(SD) together which helps in performing a secure transaction in real-time. The implementation of CD and SD layers is done to detect fraudulent activities in duplicates as well as the real social relationships. Communal Detection and Spike Detection layers are continuously updated so that the fraudster should never get a chance of attacking again. It has documented the development and evaluation in the data mining layers of defense for a real-time credit application fraud detection system. The main focus of this project is the real-time search for patterns in a principled fashion, to safeguard credit applications at the transaction. The system implements the concepts of resilience (multilayer defense), adaptivity (accounts for changing fraud and legal behavior), and quality data (real-time removal of data errors).

REFERENCES

[1] Clifton Phua, Kate Smith-Miles, Vincent Cheng-Siong Lee and Ross Gayler, "Resilient Identity Crime Detection", IEEE Transactions on Knowledge and Data Engineering, vol.2, no. 3,pp.533-546, 2012.

[2] Alka Herenj, Susmita Mishra, "Secure Mechanism for Credit Card Transaction Fraud Detection System", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 2, February 2013.

[3] Namrata Shukla, Shweta Pandey, "Document Fraud Detection with the help of Data Mining and Secure Substitution Method with Frequency Analysis", International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume 2, Number 2, June 2012.

[4] K. Vidhya, P. Dinesh Kumar, "Multi-Secure Approach for Credit Card Application Validation", International Journal of Computer Trends and Technology, volume 4,Issue 2, 2013.

[5] M. Swathi, K. Kalpana, "Spirit of Identity Fraud And Counterfeit Detection", International Journal of Computer Trends and Technology (IJCTT) , volume 4, Issue 6, June 2013.

[6] Clifton Phua, Kate Smith-Miles, Vincent Lee and Ross Gayler- Adaptive Spike Detection for Resilient Data Stream Mining, 2010.

T. P. Latchoumi, V. M. Vijay Kannan, "Synthetic Identity of Crime Detection", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013.

