International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

General Article

Identity Crime Detection for Multiple Applications

A.S Ghorpade[Á*], Y.B.Sumbe[Á] and J.E.Nalavade[Á]

[Á]Department of CSE, Sinhagad Institute of Technology, Lonavala, India

Accepted 10 May 2014, Available online 01 June 2014, **Vol.4, No.3 (June 2014)**

Abstract

Credit card fraud is an element of identity fraud. Information on the card can be used and other identity crime can arise. Protect from miss using signature on the back of a credit card which is stolen, giving a credit card to a friend or family member can cause someone to obtain details which they need to open other credit card accounts or bank accounts with the victim's name. Credit applications are costumer request f o r smart c1ards, mortgage loans, and personal loans. This can be of the paper work or internet base form. Credit application fraud is a type of fraud which includes synthetic identity fraud and real identity theft. This paper contents addition of two layers: communal detection (CD) and spike detection (SD) [4]. CD checks for social relationships to reduce the suspicion score, and it is not effected by synthetic identity. Whitelist-oriented approach is used on fixed set of attributes. SD increases the suspicion by duplicating on certain attributes. Attribute-oriented approach is used on variable-size set of attributes,

Keywords: *Communal Detection, Spike Detection, fraud detection*

1. Introduction

The data mining consists of multiple algorithms for detection. Data mining algorithms are used in the online credit card application for counterfeit detection. The algorithms are used in this system are spike and communal detection. These algorithms are used to detect the fraud and save the data in the database as original data or blacklist database. Database is manually updated by system. This system does not give a chance to defaulters in credit card application. Identity crime is defined as largely as feasib le in this method. At one extreme, real identity theft refers to illegal use of innocent people's complete characteristics. These can be harder to obtain (although large volumes of some identity data are widely available) but easier to successfully apply. In actuality, identity crime can be committed with a mix of both synthetic and real identity details. Credit applications are costumer request for smart cards, mortgage loans, and personal loans. This can be of the paper work or internet base form. Credit application deceit is a specific case of identity crime, involving synthetic identity fraud and real identity theft 0. As in identity crime, the credit application fraud has reached a critical mass of defaulters who are highly experienced, organized, and sophisticated. There are two types of duplicates: exact (or identical) Duplicates have the all same values; near (or approximate) duplicates have some same values (or characters), some similar values with a little changed spelling, or both. IN short, the new methods are based on White-listing and Detecting spikes of similar applications. White-listing uses existent common relationships on a fixed set of attributes. It lowers suspicion score reducing chance of false match. Detecting spikes in duplicate, on a variable set of attributes. This makes probability of positives by adjusting score.

2. Related Work

Identity Crime is defined as broadly as possible. At the other extreme, identity theft means using innocent people's complete identity details without their permission. These can be harder to obtain (although large volumes of some identity data are widely available) but easier to successfully apply. In reality, identity crime can be committed with a mix of both synthetic and real identity details 0.

Identity crime has increased so much because of the amount of data available on the internet. It has also made easier for fraud person to hide real identity. This can happen in a myriad of insurance, credit, and telecommunications fraud, as well as other more serious crimes. In addition to this, identity crime costly in developed countries that do not have nationally registered identity.

Data breaches which involve lost or stolen consumers' identity information can lead to other frauds such as tax returns, home equity, and payment card fraud. Consumers can incur thousands of dollars in out-of-pocket expenses. The US law requires offending organizations to notify consumers, so that consumers can mitigate the harm. As a result, these organizations incur economic damage, such as notification costs, fines, and lost business.

*Corresponding author: **A.S Ghorpade**

Their visible patterns can be different to each other and constantly change. They are persistent, due to the high financial rewards, and the risk and effort involved are minimal. Based on an observation of experienced credit application investigators, fraudsters can use software automation to manipulate particular values within an application and increase frequency of successful values.

Duplicates refer to applications which share common values. There are two types of duplicates: exact (or identical) duplicates have the all same values; near (or approximate) duplicates have some same values (or characters), some similar values with slightly altered spellings, or both. This paper argues that each successful credit application fraud pattern is represented by a sudden and sharp spike in duplicates within a short time, relative to the established baseline level.

Duplicates are hard to avoid from fraudster's point of-view because duplicates increase their success rate. The synthetic identity fraudster has low success rate, and is likely to reuse fictitious identities which have been successful before. The identity thief has limited time because innocent people can discover the fraud early and take action, and will quickly use the same real identities at different places.

It will be shown later in this paper that many fraudsters operate this way with these applications and that their characteristic pattern of behavior can be detected by the methods reported. In short, the new methods are based on white-listing and detecting spikes of similar applications. White-listing uses real social relationships on a fixed set of attributes. It lowers suspicion score reducing chance of false match., On a variable set of attributes, detecting spikes duplicates, increases true positives by adjusting suspicion scores appropriately. Throughout this paper, data mining is defined as the real-time search for patterns in a principled (or systematic) fashion. These patterns can be highly indicative of early symptoms in identity crime, especially synthetic identity fraud.

3. Identiy Crime Detection

A. Data Mining

Data mining, the extraction of hidden predictive information from huge number of records, is a powerful innovative technology with great likely to help companies focus on the most important processed data in the data warehouses. Data mining tools forecast future trend and behavior, allowing businesses to make proactive, decision support systems. The administrator verifies the provided data with the existing datum to find whether it is defaulter or original. If the data is original, it will be added to the database otherwise it will be placed into the blacklist.

B. Existing System

There are non-data mining layers of defense to protect against credit application fraud, each with its unique strengths and weaknesses. The first existing defense is made up of business rules and scorecards. One business rule is the hundred-point physical identity check test which requires the applicant to provide sufficient point-weighted identity documents face-to-face. Another business rule is to contact (or investigate) the applicant over the telephone or Internet. The above two business rules are highly effective, but human resource intensive. To rely less on human resources, a common business rule is to match an application's identity number, address, or phone number against external databases. This is convenient, but the public telephone and address directories, semi-public voters' register, and credit history data can have data quality issues of accuracy, completeness, and timeliness.

The second existing defense is known fraud matching. Here, known frauds are complete applications which were confirmed to have the intent to defraud and usually periodically recorded into a blacklist. Subsequently, the current applications are matched against the blacklist. This has the benefit and clarity of hindsight because patterns often repeat themselves. However, there are two main problems in using known frauds.

In the real-time credit application fraud detection domain, it argues against the use of classification algorithms which use class labels. In addition to the problems of using known frauds, these algorithms, such as logistic regression, neural networks, or Support Vector Machines (SVM), cannot achieve scalability or handle the extreme imbalanced class in credit application data streams. As fraud and legal behavior changes frequently, the classifiers will deteriorate rapidly and the supervised classification algorithms will need to be trained on the new data. But the training time is too long for real-time credit application fraud detection because the new training data has too many derived numerical attributes and too few known frauds. For that extremely valuable classifiers were used to make algorithms for communal detections and spike detections.

C. Demerits of existing system

The system detects the whether the data is fraud or original. If the system is data is fraud the processes do not proceed to the next level. The system is attribute oriented that the data is updated in the communal detection manually. The system does not verify from the blacklist database. Through the spike detection the system updates the Attributes regularly. The system is not secure and it detects the original data also as fraud. (for e.g.-twins applying the card is also detects as the fraudulent data). There are non-data mining layers of defense to protect against credit application fraud, each with its unique strengths and weaknesses. There are no rules and layer for fraud detection in online commerce, which is giving scope to intruders and hackers to perform crime in online shopping. It has also become easy for perpetrators to hide their true identities. This can happen in a myriad of insurance, credit, and telecommunications fraud, as well as other more serious crimes.

4. System Architecture

The architecture represents the overall structure of the system. The data is detected for the crime detection using the data mining algorithm communal detection and spike detection algorithm. These two algorithms combine together to remove the negative false and then proceeded to the proposed system algorithm. This algorithm retrieved and diagnosis the datum. If the data is fraud it is thrown into the black list database. If the data is original the data is stored in the database. The communal detection focused on attacks in the white list by fraudsters when they submit applications with synthetic relationship. The volume and ranks of the white list's real communal relationships change over time to make the white list exercise caution with (more adaptive) changing legal behavior, the white list is continually being reconstructed. The spike detection is attributing oriented.

It cannot be detected by fraud attribute will be updated regularly. The attributes used in spike detection will not be communal detection. By using the spike detection and communal detection detects the fraudsters in credit card application In addition to communal detection and spike detection we use case based reasoning algorithm to make this approach more efficient. CBR implements retrieval, diagnosis and resolution to make the data more secure. The CBR used to analyze and retrieval of data from the existing blacklist. The fraudulent datum is moved to the blacklist and the original datum is stored in the database.

A. Communal Detection

If there are two credit card applications that provided the same postal address, phone number, cell number and date of birth(DOB), but in the first application the applicant's name to be John Smith, and in the other application the applicant's name to be Joan Smith. Either it is a defaulter attempting to obtain multiple credit cards using near duplicate data. Possibly there are twins living in the same house who both are applying for a credit card. Or it can be the same person applying two times and there is a typographical error of one character in the first name. It is crucial because it reduces the scores of these legal behaviors and false positives. There are two problems with the white list. Initially there can be targeted attacks on the white list by defaulters when they submit applications with synthetic communal relationships Second, the volume and ranks of the white list's real communal relationships changes from time to time. To make the White list exercise caution with (or more adaptive to) changing legal behavior, the white list is continually being reconstructed.

B. Spike Detection

Spike Detection (SD) finds spikes to increase the mistrust value, and is query resistant for attributes. Probe resistance reduces the chances a Defaulter will discover attributes used in the Spike Detection (SD) score calculation . It is the attribute-oriented approach on a variable-size set of attributes. 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 Spike Detection (SD) suspicion score.

C. Crime Detection

The crime detection consists of the two algorithms, communal detection and spike detection. The communal detection detects the fraudsters. This detection is the relationship oriented. This detection is attribute oriented. The spike detection detects the system fraudsters by updating the system attributes. These system finds the data whether the data is original or not. These two detections are mainly involved in existing crime detection.

D. Finding Legitimate User

The process 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 algorithm involves with the data mining concept with match analysis.

E. Blacklist Verification

With the provided sets of details are taken into consideration to avoid the identity crime. The data is verified using the above algorithms to make the credit card application enormously efficient. If the data is original further processes will be enforced or otherwise the data will be found as fraud and it will be enrolled in the black list.

F. Crime Detection

The crime detection consists of the two algorithms, Communal detection and spike detection. The communal detection detects the fraudsters. This detection is the relationship oriented. This detection is attribute oriented. The spike detection detects the system

fraudsters by updating the system attributes. These system finds the data whether the data is original or not. These two detections are mainly involved in existing crime detection.

G. Finding Legitimate User

The process 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 algorithm involves with the data mining concept with match analysis.

H. Blacklist Verification

With the provided sets of details are taken into consideration to avoid the identity crime. The data is

verified using the above algorithms to make the credit card application enormously efficient. If the data is original further processes will be enforced or otherwise the data will be found as fraud and it will be enrolled in the black list.

5. Implementation Description

A. *Credit module*

Credit applications are costumer request for smart cards, mortgage loans, and personal loans. This can be of the paper work or internet base form. Users should submit identity details and it must be original. To reduce identity crime, the most important textual identity attributes such as personal name, father name, date of birth ,mobile number ,Address , Email id must be used in credit application.

B. *Making White List:*

CD checks for social relationships to reduce the suspicion score, and it is not affected by synthetic identity. Whitelist- oriented approach is used on fixed set of attributes. With the CD layer, any two similar applications could be easily interpreted as because this paper's detection methods use the similarity of the current application to all prior applications (not just known frauds) as the suspicion score. However, for this particular scenario, CD would also recognize these two applications as either or by lowering the suspicion score due to the higher possibility that they are legitimate. To account for legal behavior and data errors, Communal Detection is the white list-oriented approach on a fixed set of attributes. The white list, a list of communal and self-relationships between applications, is crucial because it reduces the scores of these legal behaviors and false positives.

Communal relationships are near duplicates which reflect the social relationships from tight familial bonds to casual acquaintances: family members, housemates, colleagues, neighbors, or friends The family member relationship can be further broken down into more detailed relationships such ashusband-wife, parent-child, brother-sister, male-female cousin (or both male, or both female), as well as uncle niece (or uncle- nephew, auntie-niece, auntie-nephew).

C. *Verification module*

In this module loan provider verify the user identity details. Duplicates (or matches) refer to applications which share common values. There are two types of duplicates: exact (or identical) duplicates have the all same values; near (or approximate) duplicates have some same values (or characters), some similar values with slightly altered spellings, or both.

D. *Mobile Layer Interface*

Mobile Layer Interface is used to increase protection for our crime detection project. On every shopping a random number is generated and a sms is forwarded to client mobile, on finishing the shopping SMS code to be integrated with shopping site, if SMS code matches with generated code then shopping will be successful.

6. Results

A methodology is the process of acquiring communication traces in large scale parallel application.

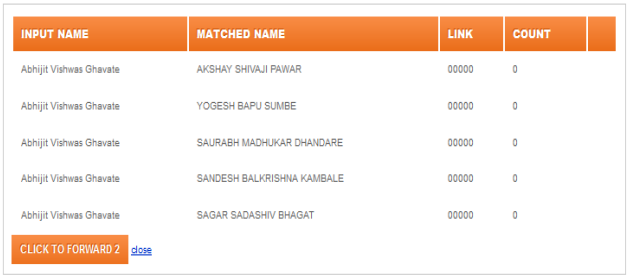

Fig. 1 List of user's with count and link

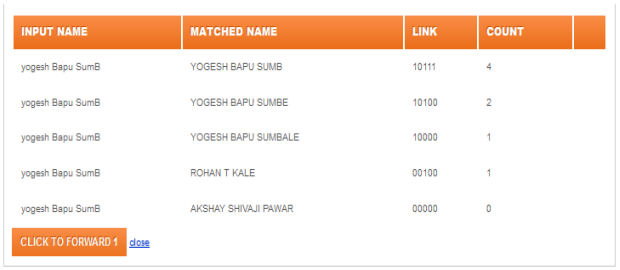

Fig. 2 List of user's with count and link

Fig. 3 User identified as real

Fig.4 User identified as fraud

7. Future Enhancement

The detection of credit card application is used with the data mining layers. This system is used only on the application, in future the fraud detection in credit process (i.e) the card is used by the unauthorized user. This system can be developed with the data mining system.

Conclusion

The system detects fraud detection online credit card

application. T his system is used avoid the duplicates from the fraudsters while applying the credit card. Data mining algorithms are used this system. The existing algorithm communal detection and spike detection used to detect the multiple applicants. In proposed system combing with the existing algorithm spike detection and communal detection the algorithm is used to make the system more efficient and secure. The Algorithm is used to throw the fraudulent data in the blacklist and retrieve the datum from the blacklist database. The identity thief has limited time because innocent people can discover the fraud early and do the necessary tasks, and will rapidly use the same real identities at different places.

References

Bifet, A. and Kirk by, R. 2009. Massive Online Analys is, Technical Manual, University of Waikato.

Bolton, R. and Hand, D. 2001. Unsupervised Profiling Methods for Fraud Detection, Proc. of CSCC01.

Brockett, P., Derrig, R., Golden, L., Levine, A. and Alpert, M. 2002. Fraud Classification using Principal Component Analysis of RIDITs, The Journal of Risk and Insurance 69(3): pp. 341-371. DOI:10.1111/1539-6975.0002.

Clifton Phua, Kate Smith-Miles, Ross Gayler, Vincent Cheng-Siong Lee, Resilient Identity Crime Detection, IEEE transactions on knowledge and data engineering, vol. 24, no. 3, march 2012

Shyam Varan Nath, Crime Pattern Detection Using Data Mining, T.P.Latchoumi and V.M.Vijay Kannan, Synthetic Identity of Crime Detection, *IJARCSSE* Volume 3, Issue 7, July 2013.

Witten and E. Frank, Data Mining: Practical Machine LearningTools and Techniques wit Java Morgan Kauffman, 2000.