Review Article

Identity Crime Detection: A Review

K.R.Gavhane[À*], R.S.Chennaram[À] and J.E.Nalavade[À]

[À]Sinhagad Institute of Technology, Lonavala, India

Abstract

In this paper, we detects identity crime, verify them and if this is identity crime we ban there application or check there documents. If after checking documents we found this is real user then we accept them. If we found crime then we ban them. Data mining is best solution for Identity crime detection. Another way, we also used CD algorithm and SD algorithm. In CD algorithm i.e. Communal Detection we found actual social relationship. Using CD algorithm we search for crime using linking. This is the best way for Identity Crime Detection. Fraud detection using mining is time saving, costless and efficient in nature. SD is spike detection finds spikes in duplicates to increase the suspicion score. Instead of SD algorithm we used to upload documents. Documents upload is another best way for verifying identity crime. We make use of this for Credit application as well as Home loan application. With refer of another paper we used this for many useful purposes.

Keywords: Data mining-based fraud detection, security, Documents.

1. Introduction

Our Identity crime detection system is web based and data mining based as well. We use CD and SD algorithm. We used data mining, because of it gives multiple detection algorithms. This is used for online credit card application and Home Loan application. We use Spike detection, Communal detection and CBR algorithm. CBR is Case-Based Reasoning is take uses of spike detection and Communal detection (Clifton Phua *et al*, 2012). Identity crime is mixture of real and synthetic identity detection. If we found similar attribute then we make linking to them. But many time we required match spelling in capital and small letter. Dictionary based approach is another way for checking and matching spelling (O.Kursun et al, 2006). But dictionary based approach is failed to match small and capital letter. So we used levenshtein algorithm (Li Yujian et al, 2007). Levenshtein algorithm can matched every other letter and ignore cases. Identity crime detection is method of determine identity crime and detect fraud person. If person apply at primary level then it is very much difficult to determine crime. So we used to upload documents and because of that detection of identity crime is make simple. We found 80% real data and 15% fraud data and remaining 5% is not give result so in this case documents verification make good role in nature. SD algorithm is spike detection is somehow difficult but efficient algorithm after CD algorithm (Ramkumar E et al, 2013). Whitelisting used for storing real social relationships and fixed the attribute. False positive can be

reduced using this. Variable set of attribute can be detecting in spike duplicates.

2. Data mining based detection

The main challenge is detection of identity crime in data mining based. Previously system uses manual crime detection system. But those systems are time costing and more costly. Efficiency of that system was not better. So we use of data mining based system. It is giving better efficiency of detecting real crime. In real world identity crime happens many times. So detection of them is most important concerned in today world. So we make identity crime detection system on data mining based. The present detecting systems are based on non-data mining concepts, due to which defence becomes weak. The System needs 'defence in depth' with multi-layered defence to cover different attacks (O.Kursun et al, 2006). Data mining is layer based system and in every layer it checks for its actual detection method so we can manage all of our data in a many variant way. Blacklist and whitelist are the concept mainly coming from identity crime dictionary. Identity theft are stole someone's identity and used them for our work. In credit application if fraud person want any free ship then they are doing this type of work. Quality data are highly desirable for data mining and data quality can be improved through the real time removal of data errors.

3. Identity crime detection

In identity crime detection difference algorithm are used and they can be differs from each other using their

---

*Corresponding author: **K.R.Gavhane**; **J.E.Nalavade** is working as Associate Professor

Literature survey

Sr.No	Algorithm/ Methodology	Author	Advantages	Disadvantages
1.	Data Stream Mining	Clifton Phua et al, 2012	1.Highly Effective 2. Data Mining Based	Human Resource Intensive
2.	Fraud Matching	Divya R. et al, 2012	Generates Blacklist With Benefit And Clarity	Non Data mining Concepts Causing Less Percentage Of Fraud Detection
3.	Cd-Algorithm	Clifton Phua et al, 2012	Communal Fraud Detection And Generates Suspicious Score	Less Effective To Achieve Adaptively And Quality
4.	Synthetic Identity	T.P.Latchoumi et al, 2013)	Detection Of Credit Card Application On Data Mining Layer	This System Works On Only Application
5.	Data Stream Mining	Clifton Phua et al, 2012	Eliminate This Error.	If The System Is Data Is Fraud The Process Do Not Proceed To The Next Level
6.	Approach To Fast And Accurate Name	O.Kursun et al, 2006	Easy Name Sortable Method.	Unable To Sort Out Name Like Allan" And "Allen
7.	Approach To Fast And Accurate Name	Li Yujian et al, 2007	Eliminate Above Error	Sortable Item Should Be In Uppercase

property. All the time this data mining based system must be verified in online nature. So many algorithms should be used in data mining based layer. For online credit card or home loan fraud detection can be handled in easier manner. Previously system can be specifying with its actual property. So now a day we have to used new algorithm with this system.

The Existing systems are based on non-data mining layers of defence to protect against credit application, banking application fraud.

(Clifton Phua et al, 2012) Have developed an important domain that has many problems with many other data mining based system. They give data layer without human interface. It was very much easy and chip in database operation. This is first framework which works on data mining based identity crime detection.

(Ramkumar E et al, 2013) had determined the CD focused on attacks in the white list by fraudster when they submit applications with synthetic relationship. The volume and ranks of the white list's real communal relationships change overtime. CBR implements retrieval, diagnosis and Resolution to make the data more secure.

(Li Yujian et al, 2007) Develop matching algorithm using levenshtein algorithm. This algorithm is more efficient than every another data matching algorithm. It is very much faster in various way and nature. This is very efficient and cost of this algorithm is very low over another algorithms.

(O.Kursun et al, 2006) Useful as dictionary based approach i.e. string matching algorithm. Over much other application this is very important and can be normalizing by another application. This is deterministic in nature. It uses prefix algorithm which is used in DBMS algorithm.
(Divya R. et al, 2012) They work on user pin number. If user enters pin number. After checking pin it allows check with particular account. They check many possibilities that user match with another user.

(T.P.Latchoumi. et al, 2013) with the existing algorithms

proposed the Case based reasoning algorithm to make the method secure. This method verifies data with blacklist data. Diagnosis of data is more secure in their system.

4. Crime detection algorithm

4.1 Communal Detection - If there are two credit card applications that provided the same postal address, home phone number, and date of birth, but one stated the applicant's name to be John Smith, and the other stated the applicant's name to be Joan Smith. Either it is a fraudster attempting to obtain multiple credit cards using near duplicated 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 twice, and there is a typographical error of one character in the first name. This research focuses on one rapid and continuous data Stream (Ramkumar E et al, 2013) of applications. It is crucial because it reduces the scores of these legal behaviours and false positives. There are two problems with the white list. First there can be focused attacks on the white list by

fraudsters when they submit applications with synthetic communal relationships Second, the volume and ranks of the white list's real communal relationships change over time. To make the White list exercise caution with (or more adaptive to) changing legal behaviour, the white list is continually being reconstructed.

4.2 Spike Detection - 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 (Clifton Phua et al, 2012) 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 SD Suspicion-score.

4.3 Case Based Reasoning (CBR)

4.3.1 Retrieval **-** Nearest neighbour 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, i.e. above a certain percentage of similarity, were retrieved.

4.3.2 Diagnosis - 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.

4.3.3 Resolution - If more than one algorithm 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.

5. Conclusion

The system detects the fraud detection online credit card application and home loan application as well. Many time this system used find duplicate entry of user data. It is very much efficient than any other system. For recognizing any other identity crime system this system is more than efficient. Duplicate entry from users can be handled easily using this system.

6. Future scope

The detection of credit card application is on mining based. This system is only for detection of credit card application, But not for card purchasing. Card used by unauthorized user can be easily handled.

References

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

O. Kursun, A. Koufakou, b. Chen, M. Georgiopoulos, K. Reynolds, and R. Eaglin (2006), A Dictionary-based approach to fast and accurate name matching in large law enforcement databases, *proc. IEEE int'l conf. Intelligence and security informatics (isi '06),* pp. 72-82.

Li Yujian, Liu Bo (2007), A Normalized Levenshtein Distance Metric, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 6, pp. 1091-1095.

Ramkumar. E., Mrs. Kavitha P (2013), online credit card application and identity crime detection, *International Journal of Engineering Research & Technology,* Vol.2 - Issue 2.

Latha venkateshwari, Divya R. saravana Deepak M. (2013), Identity Crime detection on credits cards, *International Journal of Engineering Associates,* Vol. 1 Issue 5.

T.P.Latchoumi and V.M.Vijay Kannan (2013), Synthetic Identity of Crime Detection, *IJARCSSE,* Volume 3, Issue 7.