SOFT COMPUTING: INVESTIGATION OF TWO-STEP APPROACH TO IDENTITY THEFT DETECTION IN ELECTRONIC PAYMENTS

[1]ISMAILA W. OLADIMEJI, [1]FALOHUN A. S., [2]ISMAILA FOLASADE. M., [1]OGUNJINMI TEMITOPE O., [1]BABALOLA O. RICHARD

[1]Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria, [2]Department of Computer Science, Osun State Polytechnic, Iree, Nigeria.

Abstract

Identity theft and identity fraud are terms used to refer to all types of crime in which someone wrongfully obtains and uses another person's personal data in some way that involves fraud or deception, typically for economic gain. In e-commerce, the growing number of credit card transactions provides more opportunity for thieves to steal credit card numbers and subsequently commit fraud. Despite significant efforts by merchants, card issuers and law enforcement to curb fraud, online fraud continues to plague electronic commerce web sites. In this research work, Communal detection (CD) and Counter Propagation Neural Network (CPNN) were employed for fraud detection in identity theft in credit cards online transactions. Three thousand credit cards transactions were simulated. The selected simulated applicants attributes were worked on by CD to produce whitelists and blacklists. The result-cum-applicants data were preprocessed and passed to the CPNN to performed classifications. The results showed that CD-CPNN system produced average false positive rate, average false alarm rate, average detection rate and average prediction accuracy of 26.3, 2.7, 93.6, and 92.5%; CD system generated 67.2, 6.8, 90.4 and 85.3%; while CPNN system gave 64.8, 7.4, 90.5 and 88.7% respectively.

Keywords: *Identity theft, Whitelists, Blacklists, CD, CPNN*

Introduction

Fraud is a type of criminal activity, defined as *abuse of position, or false representation, or prejudicing someone's rights for personal gain.* Put simply, fraud is an act of deception intended for personal gain or to cause a loss to another party. The general criminal offence of fraud can include deception whereby someone knowingly makes false representation, failure to disclose information or abuse a position. Fraud in e-commerce includes counterfeit fraud, identity fraud, internet

fraud, credit card fraud, etc. Identity fraud are used to consult all kinds of crime in which a person wrongfully obtains and makes use of some other individual's personal statistics in a few manner that includes fraud or deception, commonly for financial advantage. Identity fraud normally are available in photograph while people meaningfully and without criminal authority produces authentication characteristic and identification document as well as a false identification file with the intent to defraud others, which includes the us. (Ramkumar and Kavitha, 2013). Around the world, the theft and subsequent misuse of personal details are on the increase. In 2016, 15.4 million Americans were hit by identity theft. They suffered losses to the extent of 16 billion dollars. In 2017, 5.7% of adult Dutch people were a victim of identity fraud one way or the other. In England, about 90,000 cases of identity theft were registered in the first half of 2017. Most of them took place on the Internet. Criminals pose as someone else in order to steal money, buy goods or to take out a car insurance. Fraud detection is a continuously evolving discipline and ever changing tactics to commit fraud. So, it needs special methods of intelligent data analysis to detect and prevent it (Razak and Ahmed, 2014). Existing fraud detection systems may not be so much capable to reduce fraud transaction rate. Improvement in fraud detection practices has become essential to maintain existence of payment system **(**Khan et. al., 2014). To address this problem, Communal-cum-neural network detection technique for fraud detection in identity theft in credit cards online transaction is proposed. The Communal Detection is used to find the suspicious data of the fraudulent people. It also used to find the communal relationship that are near to reflect the family bond.(i.e. parent – child). It is whitelist oriented. (Sandeep and Prashant, 2016)

Identity Theft

Identity theft is the assumption of a person's identity in order, for instance, to obtain credit; to obtain credit cards from banks and retailers; to steal money from existing accounts; to rent apartments or storage units; to apply for loans; or to establish accounts using another's name. An identity thief can steal thousands of dollars in a victim's name without the victim even knowing about it for months or years. Identity thieves are able to accomplish their crimes by doing things such as opening a new credit card account with a false address, or using the victims' name, date of birth, and Social Security number. When the thief uses the credit card and does not pay the resulting bills, the delinquent account is reported on the victim's credit report. The types of identity theft present in our society today are as numerous as they are dangerous. This crime is possible wherever there are opportunities for criminals to access and use your personal information for their own benefit. In other words, anyone can become a victim.

Some of the most common types according to identity theft statistics are: Social Security Identity Theft (once your social security number is stolen, perpetrator can either sell it to undocumented workers or use it to steal property and money, or access opportunities and services available only to social security holders); Financial identity theft (occurs when a person's credit card and bank account information is stolen and used to purchase goods and services); Driver's License Identity Theft (All the perpetrator needs is access to your driver's license which can easily occur if you lose it and it falls into the wrong hands) Criminal identity theft (occurs when an individual commits a crime under another person's name) Medical identity theft (occurs when criminals use an individual's personal information such as medical identification numbers to access medical products and services); Insurance Identity Theft (It occurs when perpetrators steal medical identity information to access your insurance in order to receive medical treatment. Synthetic identity theft(occurs when criminals use your Social Security Number in combination with fake information); Tax identity theft (manifests itself through fraudulent tax refund claims) and; Child Identity Theft (This occurs when a child is the victim of the crime) **(**Alka and Susmita, 2013).

Related Works

In many years ago, the credit card industry has studied computing models for automated detection systems; recently, these models have been the subjects of academic research, especially with respect to e-commerce. Bentley (2000) proposed a system based on genetic programming. A Genetic algorithm was used to establish logic rules capable of classifying credit card transactions into suspicious and non-suspicious classes. Basically, this method followed the scoring process in which overdue payment was checked against last three months payment. If it is greater than that of last three months, then it was considered as suspicious or else it were non-suspicious. Vladimir and Strizhak, (2008) applied self-organizing map algorithm to create a fraud detection model. The pattern of legal and fraudulent transactions is observed from the earlier transactions and it is created based on the neural network training. If a new transaction does not match to the pattern of legal cardholder or is similar to the fraudulent pattern it is classified as suspicious for fraud. Tao and Gui-Yang, 2008 applied the neural data mining method on online transactions. This model was based on customer's behaviour pattern. Deviation from the usual behaviour pattern was taken as an important task to create this model. The neural network was trained with the data and the confidence value was calculated. The credit card transaction with low confidence value was not accepted by the trained neural network and it was considered as fraudulent. If the confidence value was abnormal, then again it was checked for additional confirmation. The detection performance was based on the setting of threshold. Zhang *et al.,* 2009 suggested a behaviour based credit card

fraud detection model. Here they used the historical behaviour pattern of the customer to detect the fraud. The transaction record of a single credit card was used to build the model. In this model, unsupervised Self organizing map method was used to detect the outliers from the normal ones.

Falaki *et.al*., 2012 developed a probabilistic credit card fraud detection system in online transactions using hidden markov model variant. The proposed probabilistic based model serves as a basis for mathematical derivation for adaptive threshold algorithm for detecting anomaly transactions. The model was optimized with Baum-Welsh and hybrid posterior-Viterbi algorithms. The results showed that the proposed model performed better than Viterbi and old detection model. The results obtained from the evaluation showed the overall average of accuracy and precision are about 84% and about 86% respectively. Also, Sandeep and Prashant (2016) authors proposed adaptively Communal detection and Spike detection for identity fraud detection. It was multilayered based detection algorithms which includes layer Communal Detection (CD) through operating on a set of attributes and Spike Detection (SD): by way of working on a variable size set of attributes, to defines identity crime as widely as viable. Fraud prevention gadget makes use of an expansion of policy guidelines to decide the probability of a fraudulent utility. Ramkumar. and Kavitha (2013) proposed a three layered fraud detection system which contained CD and SD that can detect more types of attacks; better account for changing legal behavior, remove the redundant attributes and to store the fraudulent datum in blacklist using CBR algorithm. CBR algorithm analysis employed retrieval, diagnosis and resolution schemes to make the data more secure and to find the malicious data. The suspicious data was thrown into the blacklist. Together CD, SD and CBR ensure the data provided by the customer is original. However, evaluation of some of these works were not explicit enough with little or no comparison with existing systems.

Materials

A. Communal Detection (CD)

CD is set of rules that discovers communal courting among the two identical records. If there are two credit card programs that were furnished with the identical postal code, telephone or cell numbers and date of start, but within the first application the applicant's call to be Sam Peter, and in the different application the applicant's call to be Saam Peter. Either it is a defaulter attempting to reap more than one credit score cards using near replica facts. Or possibly there are twins residing inside the equal residence who both are making use of a credit card. Or it can be the equal person making use of two times and there may be a typographical mistake of one man or woman within the first call. There are troubles with the white listing. to begin with there may be centered attacks at the white listing by using defaulters when they publish

programs with synthetic communal relationships second, the quantity and ranks of the white listing's real communal relationships requires modifications from time to time. To make the White list exercise caution with (or greater adaptive to) changing prison conduct, the white list is constantly being reconstructed (Swathi and Kalpana, 2013). CD layer is based on whitelist-oriented approach. It utilizes fixed set of attributes. White-listing makes use of real social relationships. This reduces false positives by lowering the suspicion scores. A threshold transaction amount is calculated based on the previous transactions made by the user. If the credit transaction amount is higher than the threshold, the user performing the transaction has to answer a security question. If the answer results to success, the transaction is authenticated or else it will be declined. Inthis manner a secure transaction will be processed.

The algorithm of communal detection is expressed in figure 1.

Inputs
current application
number of (moving window)
(link-types in current whitelist)
(string similarity threshold)
(attribute threshold)
(exact duplicate filter)
(exponential smoothing factor)
(input size threshold)
(state-of-alert)

Outputs
(suspicion score)
Same or new parameter value
New whitelist

Figure 1: CD algorithm

B. Counter Propagation Neural Network

The counter-propagation network in figure 1. is a variant of artificial neural network which is a combination of a portion of the Kohonen (1995) self-organizing map and Grossberg (1982) outstar structure. During learning, pairs of the input vector ***X*** and output vector ***Y*** were presented to the input and interpolation layers, respectively. These vectors propagate through the network in a counter flow manner to yield the competition weight vectors and interpolation weight vectors. Once these weight vectors become stable, the learning process is completed. The output vector ***Y***' of the

network corresponding to the input vector $\boldsymbol{X}$ is then computed. The vector $\boldsymbol{Y}$' is intended to be an approximation of the output vector $\boldsymbol{Y}$' i.e. Y ≈ Y = $f(X)$ (1)
The Euclidean distance between the input vector $\boldsymbol{X}$ and the competition weight vector $\boldsymbol{U}_j$ of the j-th neuron is calculated, i.e.:

$$d_j = \|X - U_j\| \sqrt{\sum_{t=1}^{m} (x_t - u_{jt})^2} \quad (3)$$

The output of the j-th neuron in the competition layer can be expressed as:

$$z_j = \begin{cases} 1.0 & \text{if } d_j < d_t \text{ for all } i \\ 0.0 & \text{otherwise} \end{cases} \quad (4)$$

The weight $\boldsymbol{u_{ji}}$ connecting the j-th neuron in the competition layer to the i-th neuron in the input layer is adjusted based on the Kohonen learning rule, i.e.:

$$u_{ij}(p+1) = u_{ji} + \beta \left(x_i - u_{ji}(p)\right) z_i \quad (5)$$

Where β is the learning coefficient and p is the iteration number. The weight $\mathbf{v_{ji}}$ is adjusted based on the Grossberg learning rule, i.e.

$$v_{ij}(p+1) = v_{ji} + \gamma \left(y_i - v_{ji}(p)\right) z_i \quad (6)$$

The j-th component $\boldsymbol{y'_j}$ of the output vector $\boldsymbol{Y}'$ can be expressed as:

$$y'_j = \sum V_{ji} Z_t \quad \ldots(7)$$

Method

This section explains the work flow (as shown in figure 2) of CD-CPNN based fraud detection system in credit cards online transactions.

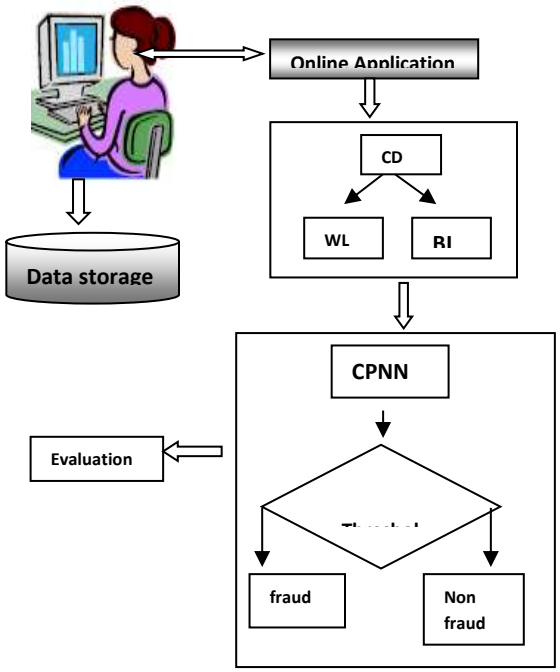

Figure 2. Workflow of the Detection system

Data Accumulation

The applicant/cardholder data contains many attributes which include those that identity thieves need to unlawfully obtain are: the victim's call, date of beginning, Social safety number, nonpublic names, addresses, phone numbers, Date of Birth, and mother's maiden name.

CD layer

The CD system compared the application submitted by the user with applications in the synthetic data set and validate the application. The whitelist was constructed from the input data set and a CD suspicious score is assigned to each application as a result of communal detection algorithm. If none of the link matched with previous accepted application in that case the application is genuine and accepted at CD layer.

CPNN layer

The results in form of blacklist and whitelist from CD were preprocessed and passed into neural network. The CPNN network was trained with some applicant dataset for optimum parameter settings and tested with the remaining. The CPNN network classified the dataset into malicious and genuine.

Evaluation

Thereafter, the results were evaluated using performance metrics viz; False Positive Rate (FPR), False Alarm Rate (FAR), Detection Rate (DR) and Accuracy.

Results and Discussion

An interactive Graphic User Interface (GUI) was developed with C-sharp programming language tool on Windows 7 Ultimate 32-bit operating system, Intel ®Pentium® CPU B960@2.20GHZ Central Processing Unit, 4GB Random Access Memory and 500GB hard disk drive.

Figure 4.1: GUI of CPNN Detection System

Figure 4.2: GUI of Communal Detection System

A simulator was used to generate cardholders transactions consisting of legal transactions intertwined with malicious types. The number of transactions used for training is 70% of total transactions and 30% are used for testing. The experimental results classified in terms of true positive (TP), true negative (TN), false positive (FP) and true negative (TN), and evaluated FPR, FAR, DR and ACC produced by CD, CPNN and CD-CPNN were presented in table 1, 2 and 3.

Table 1: Table showing results generated with CD

Cards	TP	FP	FN	TN	FPR (%)	FAR (%)	EER (%)	ACC (%)
1	36	0	3	1	0.0	0.0	92.3	96.3
2	36	1	2	1	0.0	0.00	94.7	97.5
3	30	5	2	2	71.4	14.3	93.8	80.0
4	35	1	3	1	50.0	2.2	92.1	90.0

Table2 : Table showing results generated with CPNN

Cards	TP	FP	FN	TN	FPR (%)	FAR (%)	DR (%)	ACC (%)
1	35	4	0	1	80.0	10.3	100.0	90.0
2	34	3	2	1	75.0	8.1	94.4	87.5

3	35	3	1	1	70.0	7.9	97.2	90.0
4	36	3	1	0	100.0	5.4	97.3	90.0

Table3 : Table showing results generated with CD-CPNN

Cards	**TP**	**FP**	**FN**	**TN**	**FPR (%)**	**FAR (%)**	**DR (%)**	**ACC (%)**
1	39	0	1	0	0.0	0.0	97.5	97.5
2	37	1	1	1	50.0	2.6	97.4	92.5
3	35	1	2	2	33.3	2.8	94.6	92.5
4	36	0	2	2	0.0	0.0	94.7	95.0

The CD-CPNN system produced average FPR, average FAR, average DR and average ACC of 26.3%, 2.7%, 93.6%, and 92.5% respectively, CD system generated 67.2%, 6.8%, 90.4% and 85.3%; while CPNN system gave 64.8%, 7.4%, 90.5% and 88.7% respectively.

CONCLUSION

This paper presented a Communal detection –cum-counter propagation neural network system to detect identity fraud in credit card online transactions. Identity theft and identity fraud are terms used to refer to all types of crime in which someone wrongfully obtains and uses another person's personal data in some way that involves fraud or deception, typically for economic gain. Three thousand credit cards transactions were simulated. The selected simulated applicants attributes were worked on by CD to produce whitelists and blacklists. The result-cum-applicants data were preprocessed and passed to the CPNN to performed classifications. The results showed that CD-CPNN system produced average FPR,, average FAR, average DR and average ACC of 26.3%, 2.7%, 93.6%, and 92.5%; CD system generated 67.2%, 6.8%, 90.4% and 85.3%; while CPNN system gave 64.8%, 7.4%, 90.5% and 88.7% respectively.

Acknowledgement

I acknowledge the Tertiary Trust Fund (TETFUND) Nigeria for sponsoring this research work and publication.

References

Sandeep S. W., Prashant P R. (2016). Multilayer Approach for Identity Fraud Detection System Using Communal Detection and Spike Detection, International Journal of Innovations in Engineering Sciences and Technology: CS , volume. 2, issue. 2, IJIEST: CS-2016

Ramkumar. E & Kavitha.P (2013).Online Credit Card Application and Identity Crime Detection, International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 2, ISSN: 2278-0181.

Bentley P. J., Kim J. J., Gil-Ho and Choi, J., (2000) "Fuzzy Darwinian Detection of Credit Card Fraud" , *Proc of 14th Annual Fall Symposium of the Korean Information Processing Society.*

Valdimir Z. and Strizhak A., (2006) "Credit Card Fraud Detection using Self Organizing Maps", *Information and Security: An International Journal*, 18: 48-63.

Falaki S. O., Alese B. K., Adewale O. S., Ayeni J. O., Aderounmu G. A. and Ismaila W. O. (2012) Probabilistic Credit Card Fraud Detection System in Online Transactions" In: *International Journal of Software Engineering and Its Applications,* Vol. 6, No. 4.

Tao G. and Gui-Yang L. (2008), "Neural Data Mining for Credit Card Fraud Detection", *International Conference on Machine Learning and Cybernetics*, 7; 3630-3634.

Zhang Y, Fucheng Y. and Liu H.,(2009) "Behavior–Based Credit Card Fraud Detection Model", *Fifth International Joint Conference on INC, IMS and* IDC, pp. 855-858.

Asari, V.K., (2001), "Training of a Feed forward Multiple-Valued Neural Network by Error Back Propagation with a Multilevel Threshold Function", *IEEE Trans. on Neural Networks*, 12(6): 1519-1520.

Khan M. Z., Pathan J. D., Ahmed A. H. E. (2014) "Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering" In: *International Journal of Advanced Research in Computer and Communication Engineering*, 3(2), 5458-5461.

Kohonen T., (1995). Self-Organizing Map. 2nd edition, Berlin: Springer-Verlag. pp. 1-12.

Grossberg S. (1982). Studies of Mind and Brain. Boston: Reidel Publishing. Hans-Ulrich Bauer and Klaus R. Pawelzik. (1992). Quantifying the neighborhood preservation of Self-Organizing Feature Maps. IEEE Transactions on Neural Networks, 3(4), pp.570–579

Razak T. A. and Ahmed G. N. (2014). A Comparative Analysis on Credit Card Fraud Techniques Using Data Mining. *International Journal of Data Mining Techniques and Applications, Integrated Intelligent Research (IIR),* 3(2), pp. 398-400.

https://www.mountainalarm.com/blog/9-most-common-types-of-identity-theft/

Alka H. and Susmita M. (2013). 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.