International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 2, Issue 3, pp 27-30, May-2018

Multilayered Approach for Identity Crime Detection system

[1]Ashwini Mote, [2]Nitu Pariyal

[1]*Student, MGM College of Engineering, Nanded, Maharashtra, India*

[2]*Professor, CSE Department, MGM College of Engineering, Nanded, Maharashtra, India*

Abstract – Identity crime is well known, prevalent, and costly, and credit application scam is a specific case of identity crime. The methods communal detection (CD) and spike detection (SD) are unsupervised algorithms. CD finds real social relationships to reduce the suspicion score, and is tamper resistant to synthetic social relationships. It is the white list-oriented approach on a fixed set of attributes. SD finds spikes in duplicates to increase the suspicion score, and is probe-resistant for attributes. It is the attribute-oriented approach on a variable-size set of attributes. Together, CD and SD can detect more types of attacks, better account for changing legal behaviour, and remove the redundant attributesThe work here is motivated by identity crime detection or more specifically, credit application fraud detection (Phua et al. 2005), also known as white-collar crime. Data stream mining involves detecting real-time patterns to produce accurate suspicion scores which are indicative of anomalies. At the same time, the detection system has to handle continuous and rapid examples also known as records tuples, and instances where the recent examples have no class-labels.

Keywords: Data Mining Based Fraud Detection, Security, Data Stream Mining, Anomaly Detection, Case Based Reasoning

I. INTRODUCTION

Resilient data stream mining is necessary to prevent failure of detection systems. It is the security systems' ability to degrade gracefully, or to adjust to changing circumstances when under attack. Resilient data stream mining requires a series of multiple, independent, and sequential layers in a system. This is termed as defense in depth. These layers are interacting with each other to deal with the new and deliberate attacks, and make it much harder for persistent adversaries to circumvent the security system. For example, consider personal identity databases of financial institutions. They contain individual applicant's details from real identity theft and synthetic identity fraud. The former refers to innocent peoples identity details being used illegally, without their permission. Identity crime detection procedures consist of:

a) Known fraud matching

This is the first-layer defense - it is effective for repetitive frauds and real identity theft. However, there is a long delay between time that the identity is stolen and time the identity is actually reported stolen.

b) Communal detection

This is the second-layer defense - It utilizes an example based approach similar to graph theory and record linkage by working on a fixed set of attributes. It reduces the significant time delay and false alarms by filtering normal human relationships with white lists.

c) Spike detection

This is third layer of defense. it uses an attribute-oriented approach (similar to time series analysis) by working on a variable-size set of attributes. It reduces significant time delay by searching for recent duplicates.

II. CD AND SD ALGORITHM

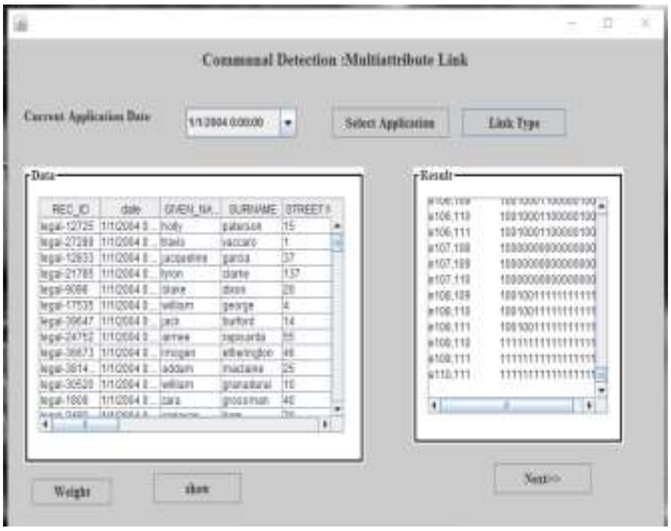
The CD algorithm matches the current application against a moving window of previous applications. The CD algorithm matches all links against the white list to find communal relationships and reduce their link score, and then the CD algorithm calculates the current application's score using every link score and previous application score. At the end of the current micro discrete data stream, the CD algorithm determines the So A 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 whitelist. When there are two credit applications where in similar kind of records exist with very minor changes, there are three possibilities. The first possibility is that there are twin brothers whose data is similar but slight change in the name. The second possibility is that a fraudster is attempting to get several credit cards from financial institution. Other possibility is that a person is applying double in order to get financial benefits. Communal Detection is an approach which can identify such scenarios. This algorithm compares data a variety of credit applications. It works on fixed set of attributes and it uses a white-list oriented approach. It finds self-relationships as well as communal relationships between the applications. The communal relationships are nothing but records with near duplicate values on the chosen attributes. A white-list is constructed with entities that display more probabilities of communal relationships.

SD is attributing oriented approach on fixed set of attributes. SD cannot use whitelist approach. SD can detect more types of attacks, better account for changing the legal behavior, and remove the redundant attribute. SD method based on detecting spikes of similar application. Spike detection process is essential in order to develop adaptively as well as resilience of the proposed solution for Credit crime detection. Communal detection has a restriction in the form of attribute threshold.

The spike detection complements communal detection which providing attribute weights. Entry of new applications can also be modified using spike detection. Detecting spikes in duplicates, on a variable set of attributes, increases true positives by adjusting suspicion scores appropriately.

III. RESULTS AND DISCUSSIONS

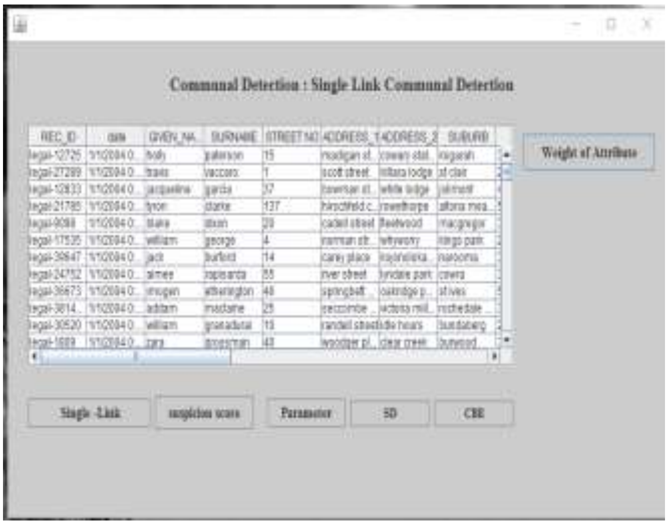
Identity crime can be found in private and commercial databases containing information collected about customers, employees, suppliers, and rule violators. The same situation occurs in public and government regulated databases such as birth, death, patient and disease registries; taxpayers, residents' address, bankruptcy, and criminals lists. To reduce identity crime, the most important textual identity attributes such as personal name, Social Security Number (SSN), Date-of-Birth (DoB), and address must be used. We choose real data set because, at experimentation time, it had the most recent fraud behavior. Although this real data set cannot be made available, there is a synthetic data set of 50,000 credit applications which is dataset is downloaded and is available at https://sites.google.com/site/cliftonphua/communalfraudscori-data.zip.



REC_ID	date	GIVEN_NA...	SURNAME	STREET N
legal-12725	1/1/2004 0...	holly	paterson	15
legal-27289	1/1/2004 0...	travis	vaccaro	1
legal-12833	1/1/2004 0...	jacqueline	garcia	37
legal-21785	1/1/2004 0...	tyron	clarke	137
legal-9098	1/1/2004 0...	blake	dixon	20
legal-17535	1/1/2004 0...	william	george	4
legal-39647	1/1/2004 0...	jack	burford	14
legal-24752	1/1/2004 0...	aimee	rapisarda	55
legal-36673	1/1/2004 0...	imogen	etherington	48
legal-3814...	1/1/2004 0...	addam	madaine	25
legal-30520	1/1/2004 0...	william	granadural	10
legal-1808	1/1/2004 0...	zara	grossman	40

e106,110	1001000110000010
e106,111	1001000110000010
e107,108	1000000000000000
e107,109	1000000000000000
e107,110	1000000000000000
e108,109	1001001111111111
e108,110	1001001111111111
e108,111	1001001111111111
e109,110	1111111111111111
e109,111	1111111111111111
e110,111	1111111111111111

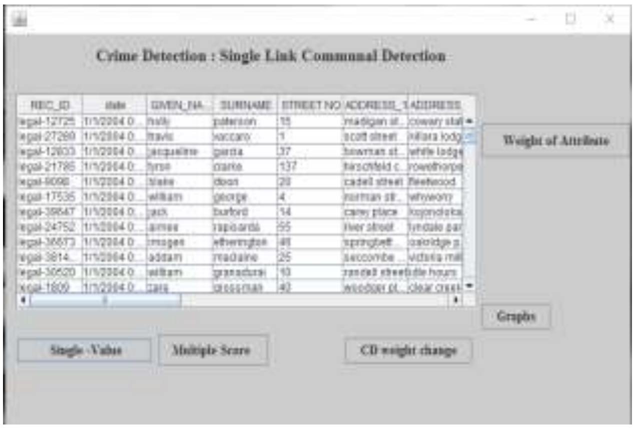
Figure-1: Link type generation



REC_ID	date	GIVEN_NA...	SURNAME	STREET NO	ADDRESS_1	ADDRESS_2	SUBURB
legal-12725	1/1/2004 0...	holly	paterson	15	madigan st.	cowan stat...	kogarah
legal-27289	1/1/2004 0...	travis	vaccaro	1	scott street	hillara lodge	st clair
legal-12833	1/1/2004 0...	jacqueline	garcia	37	bowman st.	white lodge	gilmore
legal-21785	1/1/2004 0...	tyron	clarke	137	hirschfeld c...	rowethorpe	altona mea...
legal-9098	1/1/2004 0...	blake	dixon	20	cadell street	fleetwood	macgregor
legal-17535	1/1/2004 0...	william	george	4	norman st.	whyworry	inala park
legal-39647	1/1/2004 0...	jack	burford	14	carey place	koorindaka...	naroomia
legal-24752	1/1/2004 0...	aimee	rapisarda	55	river street	lyndale park	cowra
legal-36673	1/1/2004 0...	imogen	etherington	48	springbett...	oakridge p...	st ives
legal-3814...	1/1/2004 0...	addam	madaine	25	seccombe ...	victoria mill...	rochedale
legal-30520	1/1/2004 0...	william	granadural	10	randell street	idle hours	bundaberg
legal-1689	1/1/2004 0...	zara	grossman	48	woodger pl...	clear creek	burwood

Figure-2: Calculating single link score, suspicions score and parameter

ISSN (online): 2581-3048
Volume 2, Issue 3, pp 27-30, May-2018



Crime Detection : Single Link Communal Detection

REC_ID	date	GIVEN_NA...	SURNAME	STREET NO	ADDRESS_1	ADDRESS
legal-12725	1/1/2004 0...	holly	paterson	15	madigan st...	cowary stal
legal-27289	1/1/2004 0...	travis	saccaro	1	scott street	hillara lodg
legal-12833	1/1/2004 0...	jacqueline	garcia	37	bowman st...	white lodge
legal-21786	1/1/2004 0...	tyrse	clarke	137	hirschfeld c...	rowethorpe
legal-9098	1/1/2004 0...	blake	dixon	20	cadell street	fleetwood
legal-17536	1/1/2004 0...	william	george	4	norman st...	whyworry
legal-39647	1/1/2004 0...	jack	burford	14	carey place	kojondoka
legal-24752	1/1/2004 0...	aimee	rapisarda	55	river street	lyndale par
legal-36873	1/1/2004 0...	imogen	etherington	48	springbett...	oakridge p...
legal-3814...	1/1/2004 0...	addam	madaine	25	seccombe ...	victoria mil
legal-30520	1/1/2004 0...	william	granadurai	10	randell street	idle hours
legal-1809	1/1/2004 0...	cara	grossman	40	woodger pl...	clear creek

Figure-3: Calculating single- value, multiple score & updated CD weight

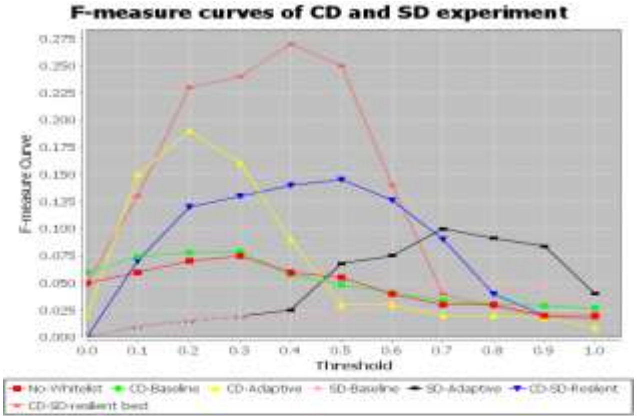


Figure-4: F-measure curve for CD and SD experiments

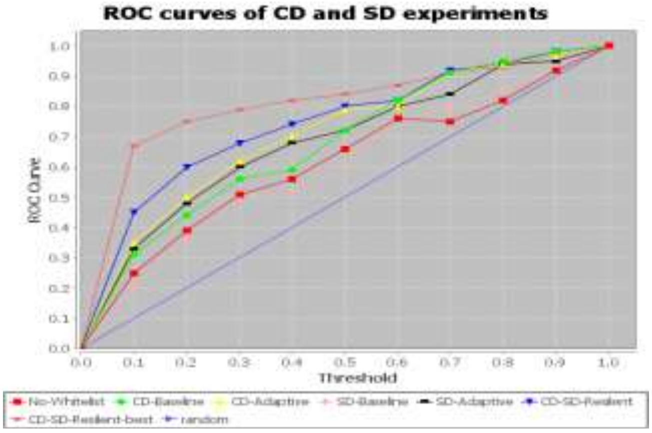


Figure-5: ROC curve for CD and SD experiments

IV. CONCLUSION

We focused on robust Credit crime detection. It has implemented algorithms to protect applications that occupy credit card. It proposed prototype application has numerous layers of defense using data mining which can be used in the real world credit applications or else for credit card fraud detection. The implementation of Communal Detection algorithm is practical because this algorithm designed for use to complement the existing detection system. The algorithm can search with a larger moving window number of link types in the white list, and a number of attributes.

V. FUTURE SCOPE

In the future work, we comparing the incoming application with the applications in the moving window of the data set for generating links for attributes consume lot of time when the window size is large. Parallelizing the code to run on multicore architectures would drastically reduce the execution time of window.

REFERENCES

[1] Clifton Phau, Member IEEE, Kate Smith-Miles, Senior Member IEEE, Vincent Cheng-Siong Lee, and Ross Gayler “Resilient Identity Crime Detection”, *IEEE Transactions* on, vol.24, no.3, 2012.

[2] A.K. Racheal Praveena, Dr.G.Venkata RamiReddi C.K Suresh Babu, “A Secure Mechanism for Resilient Of Data Mining Based Fraud Detection”, *International Journal on Computer Science and Network Solution, ISSN: 2345-3397,* Volume: 1, No3, Nov 2013.

[3] R. Bolton and D. Hand, “Unsupervised Profiling Methods for Fraud Detection”, *Statistical Science,* vol. 17, no. 3 pp.235-255, 2001.

[4] Zakia Ferdaousi and Akira Maeda A.J., “Anomaly Detection Using Unsupervised Profiling Methods in Time Series Data”, 525-8577, Japan.

[5] P.Christen and K. Goiser, “Quality and Complexity Measures for Data Linkage and Deduplication”, Quality Mearures in Data Mining F.Guillet and H.Hamilton, eds.,vol 43, *Springer,* doi 10.1007/978-3-540-44918-8, 2007.

[6] Bifet and R. Kirkby.Massive Online Analysis, Technical.Univ. of Waikato, 2009.

[7] Cortes, D. Pregibon, and C. Volinsky, “Computational Methods for Dynamic Graphs”, *J. Computational and Graphical Statistics,* vol.12, no.4, pp.950-970, 2003.

[8] T. Fawcett, “An Introduction to ROC Analysis,” Pattern Recognition Letters, vol. 27, pp. 861-874, 2006, doi:10.1016/j.patrec.2005.10.010.

[9] O. Kursun, A. Koufakou, B. Chen, M.Georgiopoulos, K.Reynolds, and R. Eaglin, “A Dictionary Based Approach to Fast and Accurate Name Matching in Large Law Enforcement Databases,” *Proc. IEEE Int’l Conf. (ISI ‘06),*pp.72-82, doi:10.1007/11760146, 2006.

[10] J. Jonas, “Non-Obvious Relationship Awareness,” *Proc. Identity mashup*, 2006.

[11] M. Kantarcioglu, W. Jiang, and B. Malin, “A Privacy – Preserving Framework for Integrating Person-Specific databases,” *proc. UNESCO Chair in Data privacy Int’l Conf. Privacy in Statistical Databases(PSD’08),* pp.298- 314, doi:10.1007/978-3-540- 87471-3_25, 2006.

[12] J. Kleinberg, Data Stream Management: Processing High-Speed Data Streams, *Springer*, 2005.

How to cite this article:

Ashwini Mote, Nitu Pariyal, “Multilayered Approach for Identity Crime Detection system”, in *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, Volume 2, Issue 3, pp 27-30, May 2018.

****