JES
Journal of Engineering Sciences
Vol 10, Issue 11, Nov /2019
ISSN NO:0377-9254

CoDetect Financial Fraud Detection using Anomaly Feature Detection

[1]SHAIK HEENA FARHEEN, [2]K.AMARENDRANATH, [3]A.D.SIVARAMA KUMAR

[1]M.Tech Student, [2,3]Assistant Professor
DEPT OF CSE
SVR Engineering College, Nandyal

ABSTRACT:

Financial fraud, such as money laundering, is known to be a serious process of crime that makes illegitimately obtained funds go to terrorism or other criminal activity. This kind of illegal activities involve complex networks of trade and financial transactions, which makes it difficult to detect the fraud entities and discover the features of fraud. Fortunately, trading/transaction network and features of entities in the network can be constructed from the complex networks of the trade and financial transactions. The trading/transaction network reveals the interaction between entities, and thus anomaly detection on trading networks can reveal the entities involved in the fraud activity; while features of entities are the description of entities, and anomaly detection on features can reflect details of the fraud activities. Thus, network and features provide complementary information for fraud detection, which has potential to improve fraud detection performance. However, the majority of existing methods focus on networks or features information separately, which does not utilize both information. In this paper, we propose a novel fraud detection framework, CoDetect, which can leverage both network information and feature information for financial fraud detection. In addition, the CoDetect can simultaneously detecting financial fraud activities and the feature patterns associated with the fraud activities. Extensive experiments on both synthetic data and real-world data demonstrate the efficiency and the effectiveness of the proposed framework in combating financial fraud, especially for money laundering.

I.INTRODUCTION

Aggregation methods are also used to enrich the information of data. After generating feature points In recent years,financial fraud activities such as credit card fraud, money laundering, increase gradually. These activities cause the loss of personal and/or enterprises' properties. Even worse, they endanger the security of nation because the fraud may go to terrorism. Thus, accurately detecting financial fraud and tracing fraud are necessary and urgent. However, financial fraud detection is not an easy task due to the complex trading networks and transactions involved. Taking money laundering as an example, money laundering is denied as the process of using trades to move money/goods with the intent of obscuring the true origin of funds. Usually, the prices, quantity or quality of goods on an invoice of money laundering are fake purposely. The misrepresentation of prices, quantity or quality of goods on an invoice merely exposes slight difference from regular basis if we use these numbers as features to generate detection policy.Under certain circumstances, this kind of detector may work well with relatively stable trading entities. Unfortunately,the real world situation is more complicated, especially within Free Trade Zones (FTZs) where international trade involves complex procedures and exchange of information between trading entities. The fraud activities, especial money laundering, are deeper stealth. Money laundering activities may take different forms such as the concealing transportation of cash using trading operations; the acquisition and sale of intangibles; and related party transactions. Not only the trading of goods shows on much more diversity, but also different type of companies, shell and front companies involve into facilitate money laundering. In contrast with other fraud activities, money laundering demonstrates special characteristic which presents high risk to financial system with obscuring the money trail, collectivization behaviour and wild trading regions in FTZs.

Many fraud detection models work with attribute value data points that are generated from transactions data.Some from transactions, supervised and unsupervised methods can be used to perform detection. Usually, these data points are

assumed to be independent and identically distributed. However, the characteristic of money laundering is different from attribute-value data. The collectivization behaviour means the data is inherently linked or partly linked.Obviously, trading activity involves at least two business entities.Linked data is patently not independent and identically distributed, which contradicts the assumptions of traditional supervised and unsupervised methods. On the other side,some linked data is auto correlated. For example, trading between business entity A and B implies that feature points A and B are correlated. Some features used to describe the properties of trading goods can be identical between A and B.

This characteristic of auto correlation reduce the effective size of data for learning. Furthermore, feature points don't use the interaction information in data. The relations between any business entities indicate the potential causality that means,if businesses on going, fraud entity can be located by other is identified fraud entity. This means the entity,which have connection with fraud entity, are suspicious. Consequently, feature based detection models with supervised or unsupervised methods have inherent limitation of incapacity of identifying what the fraud relations are.what the fraud relations are.Graph-based mining methods are one of the most important theories that attempt to identify relations between data points. Financial activities can be modelled as a directed graph, then a sparse adjacent matrix can represent this graph. With graph-mining method,the sparse matrix can be approximated as summation of low-rank matrix and outlier matrix. The outlier matrix is a sign of suspicious fraud activities. Exploiting the graph based mining provides a new perspective for fraud detection and enables us to do advanced research on fraud detection.With the fraud activities detected by graph-based detection technique we are able to draw the conclusion that several business entities involved in fraud, however, we still don't know how these fraud activities are operated and why these activities labelled as fraud, i.e., the detailed features of the fraud activities. The majority of this how-and-why information is used in features points, which have essential meaning for financial fraud because of the tracing necessity.For example, doing business with misrepresentation of the price may transfer additional value to exporter. The value in this example reveals how did the fraud happen. This simple example requires the detection system to mark value as fraud property. Another example, fraud activities might go deeper stealth with multi-entities involved. If the same good or service invoices a number of different business entities to make the payments, then there are several properties should be consider as suspicious: business location,name,direction,good and service etc.With the knowledge of these suspicious properties, tracing fraud can be much easier for executives.

II.EXISTING SYSTEM:

Bahnsen et al. [38] improve the detection performance by calibrating probabilities before establishing Bayes model. HMM model is used to model the customers' credit card shopping patterns for detection of credit card fraud. The shopping items indicate the hidden state and the corresponding prices from certain ranges are the observation. LR(Logistic Regression), Support Vector Machines(SVMs) and Random Forest(RF) are evaluated for credit card detection. The detection models are built on primary features and derived features from transaction.

Whitrow et al. [28] proposed a new preprocessing strategy for better fraud detection with SVMs and KNN classification. Transactions aggregated in term of time window, then data with new features is used to model the pattern.

Wei et al. [29] addressed the problem of unbalanced financial data and employed cost-sensitive neural network to punish the misclassification of fraud transaction. Sahin et al. [33] incorporate cost function into decision tree to boost performance on unbalanced data. Following the general procedure of classification, feature selection is proceed to boost the detection performance of credit card fraud.

Perols [35] performed a systematic analysis of financial fraud detection with popular statistical and machine learning models. The evaluation is under the supervised manner. All these methods rely on accurate identification of fraud patterns from data set and these methods also suffer from the problem of unbalanced data. Bolton and David perform fraud detection with clustering methods. This unsupervised manner is under the assumption that small cluster indicates the anomaly in data.

CoDetect is an unsupervised model which is based on matrices co factorization. The matrices from graph represent the genuine proprieties (features and

connections) of financial data. The detection results give a better understanding of fraud patterns and furthermore, help to trace the originate of fraud groups.

Disadvantages:

There is no Evaluation with Subspace Clustering Methods.

There is no SVM Classification in Credit Card Fraud Detections.

III.PROPOSED SYSTEM:

In the proposed system, the system would like to develop a novel framework for fraud detection by considering the special detecting and tracing demanding of fraud entities and behaviors. Specifically, we investigate: (1) how to utilize both graph matrix and feature matrix for fraud detection and fraud tracing; (2) how to mathematically model both graph matrix and feature matrix so as to simultaneously achieve the tasks of fraud detection and tracing. In an attempt to solve these challenges.

The system proposed a novel detection framework CoDetect for financial data, especially for money laundering data. The system incorporates fraud entities detection and anomaly feature detection in the same framework to find fraud patterns and corresponding features simultaneously. Combining entities detection and feature detection enables us to build a novel fraud detection framework for noisy and sparse financial data: relevant fraud patterns help the identification of fraud identities, and relevant features in turn help revealing of the nature of fraud activities.

Advantages:

- Provide an approach to establish weighted graph from financial network, incorporating properties of nodes and links.
- Demonstrate different scenarios of financial fraud and formulate the patterns of fraud in term of graph and sparse matrix.
- Propose a novel unsupervised framework, CoDetect, for the problem of complex patterns discovery and anomaly features identification, employing two matrices residual analysis on graph-based financial network.
- Evaluate framework using synthetic and real world data to demonstrate both effectiveness and efficiency of the proposed framework.

IV.SYSTEM DESIGN

4.1 System Architecture

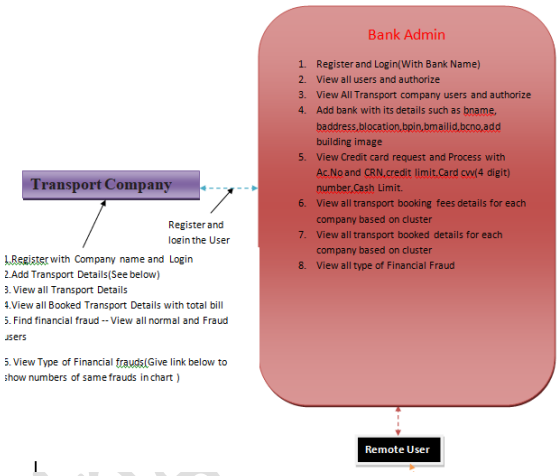

Fig 1 System Architecture

The architecture describes the activities performed by bank admin such as

1. Register and Login, View your profile, Manage Bank Account
2. Request Credit card with Details and view the same
3. View Card Transactions based on transport booked details
4. View your payments and transfer to your cc account (if user doesn't have enough amount to transfer then he is a fraud user or abnormal user)
5. View all transport company and select corresponding company and book, give reviews, increment rank ,enter card cvv number(Find fraud if no balance in cc,if cvv number is wrong)

6. View all Booked transport

The operations performed by the Transport company are

1.Register with Company name and Login
2.Add Transport Details(See below)
3. View all Transport Details
4.View all Booked Transport Details with total bill
5. Find financial fraud -- View all normal and Fraud users
6. View Type of Financial frauds(Give link below to show numbers of same frauds in chart)

Data flow diagram can be explained as follows,

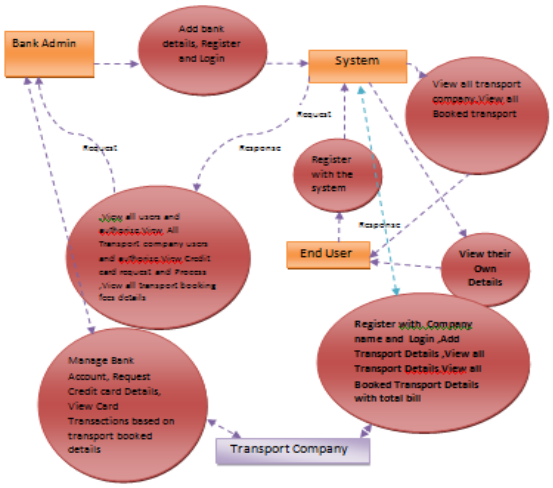

Fig 2 Data flow diagram

V.IMPLEMENTATION

5.1MODULE DESCRIPTION:

Bank Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as View all users and authorize, View all Transport Users and authorize, Register and Login(With Bank Name) ,View all users and authorize ,View All Transport company users and authorize,Add bank with its details such as bname, baddress,blocation,bpin,bmailid,bcno,add building image,View Credit card request and Process with Ac.No and CRN,credit limit,Card cvv(4 digit) number,Cash Limit,View all transport booking fees details for each company based on cluster ,View all transport booked details for each company based on cluster,View all type of Financial Fraud based on cluster,View all users with Financial Fraud and give link to show number of same user is fraud in chart

User

In this module, there are n numbers of users are present. User should register with group option before doing some operations. After registration successful he has to wait for admin to authorize him and after admin authorized him. He can login by using authorized user name and password. Login successful he will do some operations like Register and Login, View your profile, Manage Bank Account ,Request Credit card with * Details and view the same ,View Card Transactions based on transport booked details ,View your payments and transfer to your cc account (if user doesn't have enough amount to transfer then he is a fraud user or abnormal user) ,View all transport company and select corresponding company and book, give reviews, increment rank ,enter card cvv number(Find fraud if no balance in cc,if cvv number is wrong) ,View all Booked transport

Transport Company

In this module, there are n numbers of users are present. Transport Company user should register with group option before doing some operations. After registration successful he has to wait for admin to authorize him and after admin authorized him. He can login by using authorized user name and password. Login successful he will do some operations like Register with Company name and Login ,Add Transport Details(See below) ,View all Transport Details ,View all Booked Transport Details with total bill ,Find financial fraud -- View all normal and Fraud users ,View Type of Financial frauds(Give link below to show numbers of same frauds in chart

VI.CONCLUSION

In this project a new framework, CoDetect, which can perform fraud detection on graph-based similarity matrix and feature matrix simultaneously. It introduces a new way to reveal the nature of financial activities from fraud patterns to suspicious property. Furthermore, the framework provides a more interpretable way to identify the fraud on sparse matrix. Experimental results on synthetic and real world data sets show that the proposed framework (CoDetect) can effectively detect the fraud patterns as well as suspicious features. With this codetect framework, executives in financial supervision cannot only detect the fraud patterns but also trace the original of fraud with suspicious feature.Financial activities are involving with time. We can represent these activities into similarity tensor and feature tensor.So we would like to study how to integrate tensor into codetect framework for fraud detection

BIBILOGRAHY

[1] C. Sullivan and E. Smith. ``Trade-Based Money Laundering: Risks andRegulatory Responses," Social Sci. Electron. Publishing, 2012, p. 6.

[2] United Press International. (May 2009). *Trade-Based MoneyLaundering Flourishing*.[Online].Available:http://www.upi.co

m/TopNews/2009/05/11/Trade-based-money-laundering-_ourishing/UPI-17331242061466
[3] L. Akoglu, M. McGlohon, and C. Faloutsos, ``OddBall: Spotting anomaliesin weighted graphs," in *Proc. Pacic-Asia Conf. Knowl. Discovery DataMining*, 2010, pp. 410_421.
[4] V. Chandola, A. Banerjee, and V. Kumar, ``Anomaly detection: A survey,"*ACMComput. Surv.*, vol. 41, no. 3, 2009, Art. no. 15.
[5] W. Eberle and L. Holder, ``Mining for structural anomalies in graph-baseddata," in *Proc. DMin*, 2007, pp. 376_389.
[6] C. C. Noble and D. J. Cook, ``Graph-based anomaly detection," in *Proc.9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2003,pp. 631_636.
[7] H. Tong and C.-Y. Lin, ``Non-negative residual matrix factorization withapplication to graph anomaly detection," in *Proc. SIAM Int. Conf. DataMining*, 2011, pp. 1_11.
[8] S.Wang, J. Tang, and H. Liu, ``Embedded unsupervised feature selection,"in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 470_476.
[9] Z. Lin, M. Chen, and Y. Ma. (2010). ``The Augmented lagrange multipliermethod for exact recovery of corrupted low-rank matrices." [Online].Available: https://arxiv.org/abs/1009.5055.
[10] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos, ``Neighborhoodformationand anomaly detection in bipartite graphs," in *Proc. 15th IEEE Int. Conf.Data Mining*, Nov. 2005, p. 8.
[11] A. Patcha and J.-M. Park, ``An overview of anomaly detection techniques:Existing solutions and latest technological trends," *Comput. Netw.*, vol. 51,no. 12, pp. 3448_3470, Aug. 2007.
[12] W. Li, V. Mahadevan, and N. Vasconcelos, ``Anomaly detection andlocalization in crowded scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*,vol. 36, no. 1, p. 18_32, Jan. 2014.
[13] K. Henderson *et al.*, ``It's who you know: Graph mining using recursivestructural features," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 663_671.
[14] F. Keller, E. Müller, and K. Bohm, ``HiCS: High contrast subspaces fordensity-based outlier ranking," in *Proc. ICDE*, Apr. 2012, pp. 1037_1048.