CoDetect Financial Fraud Detection with Anomaly Feature Detection

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 effect 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.

**Meanings**

**Money laundering**: illegally obtained money, typically by means of transfers involving foreign banks or legitimate businesses.

**Illegitimately: (illegal )**person making transactions with un authorized (un known user)

**anomaly detection** : identification of unexpected events, observations,

**Synthetic data** : is information that's artificially manufactured rather than generated by real-world events ...

**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, and 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 proceeded 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.

A Support Vector Machine (SVM) performs classification by finding the hyperplane that maximizes the margin between the two classes. The vectors (cases) that define the hyper plane are the support vectors. Extend the above definition for non-linearly separable problems: have a penalty term for misclassifications.

**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.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - Java/J2EE(JSP,Servlet)
* Front End - J2EE
* Back End - MySQL

Modules

-----------

**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 )

CONCLUSION

We propose a new framework, CoDetect, which can performfraud detection on graph-based similarity matrix and featurematrix simultaneously. It introduces a new way to reveal the nature of financial activities from fraud patterns to suspiciousproperty. Furthermore, the framework provides a more interpretableway to identify the fraud on sparse matrix. Experimentalresults on synthetic and real world data sets show thatthe proposed framework (CoDetect) can effectively detect thefraud patterns as well as suspicious features. With this codetection 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 representthese activities into similarity tensor and feature tensor.So we would like to study how to integrate tensor into codetectframework for fraud detection.