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 difcult 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 reeffect 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.

**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.

**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