Literature Survey: -

UPI Fraud Detection Overview: -

UPI fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on data mining and neural networks, have been suggested. Ghosh and Reilly have proposed UPI fraud detection with a neural network. They have built a detection system, which is trained on a large sample labelled UPI account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and no received issue (NRI) fraud. Recently, Syed et al. have used parallel granular neural networks (PGNNs) for improving the speed of data mining and knowledge discovery process in UPI fraud detection. A complete system has been implemented for this purpose. Stolfo et al. Suggest a UPI fraud detection system (FDS) using Metal earning techniques to learn models of fraudulent UPI transactions. Metal earning is a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A met classifier is thus trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection. They use Python agents for Metal learning (JAM), which is a distributed data mining system for UPI fraud detection. A number of important performance metrics like True Positive—False Positive (TP-FP) spread and accuracy have been defined by them. Askerov et al. present CARDWATCH, a database mining system used for UPI fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Kim and Kim have identified skewed distribution of data and mix of legitimate and fraudulent transactions as the two main reasons for the complexity of UPI fraud detection. Based on this observation, they use fraud density of real transaction data as a confidence value and generate the weighted fraud score to reduce the number of misdetections. Fan et al. suggest the application of distributed data mining in UPI fraud detection. Brause et al. have developed an approach that involves advanced data mining techniques and neural network algorithms to obtain high fraud coverage. Chiu and Tsai have proposed Web services and data mining techniques to establish a collaborative scheme for fraud detection in the banking industry. With this scheme, participating banks share knowledge about the fraud patterns in a heterogeneous and distributed environment. To establish a smooth channel of data exchange, Web services techniques such as XML, SOAP, and WSDL are used. Phua et al. have done an extensive survey of existing data-mining-based FDSs and published a comprehensive report. Prodromitids and Stolfo use an agent-based approach with distributed learning for detect-Ing frauds in UPI transactions. It is based on artificial intelligence and combines inductive learning algorithms and Metal earning methods for achieving higher accuracy. Phua et al. suggest the use of met classifier similar to in fraud detection problems. They consider naive Bayesian C4.5, and Back Propagation neural networks as the base classifiers. A met classifier is used to determine which classifier should be considered based on skewness of data. Although they do not directly use UPI fraud detection as the target application, their approach is quite generic. Vatsa et al. have recently proposed a game-theoretic approach to UPI fraud detection. They model the interaction between an attacker and an FDS as a multistage game between two players, each trying to maximize his payoff. The problem with most of the abovementioned approaches is that they require labelled data for both genuine, as well as fraudulent transactions, to train the classifiers. Getting real-world fraud data is one of the biggest problems associated with UPI fraud detection. Also, these approaches cannot detect new kinds of frauds for which labelled data is not available. In contrast, we present a Hidden Markov Model (AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING)-based UPI FDS, which does not require fraud signatures and yet is able to detect frauds by considering a cardholder's spending habit. We model a UPI transaction processing sequence by the stochastic process of an AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING. The details of items purchased in individual transactions are usually not known to an FDS running at the bank that issues UPI to the cardholders. This can be represented as the underlying finite Markov chain, which is not observable. The transactions can only be observed through the other stochastic process that produces the sequence of the amount of money spent on each transaction. Hence, we feel that AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING, is an ideal choice for addressing this problem. Another important advantage of the AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING-based approach is a drastic reduction in the number of False Positives (FPs)transactions identified as malicious by an FDS although they are actually genuine. Since the number of genuine transactions is a few orders of magnitude higher than the number of malicious transactions, an FDS should be designed in such a way that the number of FPs is as low as possible. Otherwise, due to the "base rate fallacy" effect, bank administrators may tend to ignore the alarms. To the best of our knowledge, there is no other published literature on the application of AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING for UPI fraud detection.

Title: - AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING Background

An AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING is a double embedded stochastic process with two hierarchy levels. It can be used to model much more complicated stochastic processes as compared to a traditional Markov model. An AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING has a finite set of states governed by a set of transition probabilities. In a particular state, an outcome or observation can be generated according to an associated probability distribution. It is only the outcome and not the state that is visible to an external observer. AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING-based applications are common in various areas such as speech recognition, bioinformatics, and genomics. In recent years, Joshi and Phoba have investigated the capabilities of AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING in anomaly detection.

Title: BLAST-SSAHA Hybridization for UPI Fraud Detection Author: Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar

Description:

A phenomenal growth in the number of UPI transactions, especially for online purchases, has recently led to a substantial rise in fraudulent activities. Implementation of efficient fraud detection systems has thus become imperative for all UPI issuing banks to minimize their losses. In real life, fraudulent transactions are interspersed with genuine transactions and simple pattern matching is not often sufficient to detect them accurately. Thus, there is a need for combining both anomaly detection as well as misuse detection techniques. In this paper, we propose to use two-stage sequence alignment in which a profile analyzer (PA) first determines the similarity of an incoming sequence of transactions on a given UPI with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyzer are next passed on to a deviation analyzer (DA) for possible alignment with past fraudulent behavior. The final decision about the nature of a transaction is taken on the basis of the observations by these two analyzers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA

Title: - Fast algorithms for mining association rules in large databases

Author: R. Agrawal and R. Srikant.

Description:

The major consequences are loss of billions of dollars each year, investor confidence or corporate reputation. A study area called Financial Fraud Detection (FFD) is obligatory, in order to prevent the destructive results caused by financial fraud. In this study, we propose a new method based on Grammar-based Genetic Programming (GBGP), multi-objectives optimization and ensemble learning for solving FFD problems. We comprehensively compare the proposed method with Logistic Regression (LR), Neural Networks (NNs), Support Vector Machine (SVM), Bayesian Networks (BNs), Decision Trees (DTs), AdaBoost, Bagging and LogitBoost on four FFD datasets. The experimental results showed the effectiveness of the new approach in the given FFD problems including two real-life problems. The major implications and significances of the study can concretely generalize for two points. First, it evaluates a number of data mining techniques by the given real-life classification problems. Second, it suggests a new method based on GBGP, NSGA-II and ensemble learning.

Title: Why we tag: Motivations for annotation in mobile and online media

Author: M. AMES AND M. NAAMAN.

Description:

Financial fraud is a criminal act, which violates the law, rules or policy to gain unauthorized financial benefit. The major consequences are loss of billions of dollars each year, investor confidence or corporate reputation. A study area called Financial Fraud Detection (FFD) is obligatory, in order to prevent the destructive results caused by financial fraud. In this study, we propose a new method based on Grammar-based Genetic Programming (GBGP), multi-objectives optimization and ensemble learning for solving FFD problems. We comprehensively compare the proposed method with Logistic Regression (LR), Neural Networks (NNs), Support Vector Machine (SVM), Bayesian Networks (BNs), Decision Trees (DTs), AdaBoost, Bagging and LogitBoost on four FFD datasets. The experimental results showed the effectiveness of the new approach in the given FFD problems including two real-life problems. The major implications and significances of the study can concretely generalize for two points. First, it evaluates a number of data mining techniques by the given real-life classification problems. Second, it suggests a new method based on GBGP, NSGA-II and ensemble learning.

Title: Fuzzy Darwinian detection of UPI fraud

Author: PETER J. BENTLEY, JUNGWON KIM, GIL-HO JUNG AND JONG-UK CHOI

DESCRIPTION:

By the exponential growth of UPI user fraudulent transactions also have increased dramatically. The genuine transaction and fraudulent transactions are almost similar, so it is very hard to discover a fraudulent transaction form the genuine one. In this paper we have proposed fraud detection algorithm based on Fuzzy-ID3. Intermediate nodes we split using attribute having highest information gain. The leaf nodes classifies the transactions as fraud, doubtful or normal. Experimental result exhibits that the technique is efficient one in detecting frauds.

Title: UPI fraud detection using Big data

Author: Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, Bernard Manderick,

Description:

Big data is the frontier of a firm's ability to store, process and access all the data and it needs to operate effectively for decision making, reduce risks and serve customers. The 3 main characteristics of big data volume (data quantity), velocity (data speed), variety (data type), Big data can handle more than 1 million customer transactions per hour. Hadoop is an Apache top level project, open-source implementation of frameworks for reliable, scalable, distributed computing and data storage. It is a flexible and highly-available architecture for large scale computation and data processing on a network of commodity hardware. Big data helps financial institutions to approach fraud in different ways and possibly get different results. For the UPI fraud detection we need bank, transaction and customer data.

SURVEY TABLE

Ref.No	Purpose	Dataset	Method	Advantage	Disadvantage	Future Scope	Tools	Evaluation
[1]	Detect UPI fraud using CNN	UPI transactions	Convolutional Neural Network (CNN)	High accuracy in fraud detection	Requires high computational power	Can be improved with real-time processing	Python, SQL Server	Accuracy close to 80%
[2]	Fraud detection using Dempster-Shafer Theory & Bayesian Learning	UPI fraud cases	Bayesian learning & fusion approach	Enhances detection by combining multiple evidences	Requires extensive labeled data	Can integrate deep learning for better insights	Python	Improved fraud detection rates
[3]	UPI Fraud Detection using Hidden Markov Model	UPI spending patterns	Hidden Markov Model (HMM)	Detects anomalies based on spending habits	Cannot detect new fraud types	Can be extended with deep learning models	Python, SQL Server	Reduces false positives
[4]	Fuzzy Darwinian Detection of UPI Fraud	UPI transaction logs	Fuzzy-ID3 Algorithm	Handles complex fraud scenarios effectively	High processing time	Optimization using AI techniques	Python	Efficient fraud classification
[5]	Fraud detection using Bayesian & Neural Networks	Financial fraud databases	Bayesian Networks, Neural Networks	High predictive accuracy	Needs large labeled datasets	Incorporate reinforcement learning for adaptive detection	Python, SQL Server	Effective for fraud pattern recognition
[6]	Two-Stage UPI Fraud Detection using Sequence Alignment	UPI transaction sequences	BLAST-SSAHA Hybridization	Enhances anomaly & misuse detection	Requires complex implementation	Extendable to multi-modal transaction analysis	Python	Faster and accurate fraud detection
[7]	Big Data-based UPI fraud detection	Large-scale UPI transactions	Hadoop & distributed computing	Processes high-volume transactions efficiently	High infrastructure cost	Future enhancements using AI & ML	Hadoop, Python	Real-time fraud monitoring