Data Driven Fraud Detection

Data driven fraud detection may take many forms. These are just a few examples of how one might go about analyzing data to detect fraud.

*A Comprehensive Survey of Data Mining-based Fraud Detection Research:*

Data mining is about finding insights which are statistically reliable, unknown previously, and actionable from data (Elkan, 2001). This data must be available, relevant, adequate, and clean.

Also, the data mining problem must be well-defined, cannot be solved by query and reporting tools, and guided by a data mining process model (Lavrac *et al*, 2004). This survey paper categorises, compares, and summarises from almost all published technical and review articles in automated fraud detection within the last 10 years. It defines the professional fraudster, formalises the main types and subtypes of known fraud, and presents the nature of data evidence collected within affected industries. Within the business context of mining the data to

achieve higher cost savings, this research presents methods and techniques together with their problems. Compared to all related reviews on fraud detection, this survey covers much more

technical articles and is the only one, to the best of our knowledge, which proposes alternative data and solutions from related domains.

[*http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf*](http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf)

*Fraudulent Network Usage*:

Patent for system and method for detecting fraudulent network usage patterns using real-time network monitoring:

<http://www.google.com/patents?hl=en&lr=&vid=USPAT5627886&id=ws0fAAAAEBAJ&oi=fnd&dq=data+driven+fraud+detection&printsec=abstract#v=onepage&q=data%20driven%20fraud%20detection&f=false>

*Detection of Credit Card Fraud*:

The prevention of credit card fraud is an important application for prediction techniques. One major obstacle for using neural network training techniques is the high necessary diagnostic quality: since only one financial transaction in a thousand is invalid no prediction success less than 99.9% is acceptable. Because of these credit card transaction requirements, completely new concepts had to be developed and tested on real credit card data. This paper shows how advanced data mining techniques and a neural network algorithm can be combined successfully to obtain a high fraud coverage combined with a low false alarm rate.

<http://pdf.aminer.org/000/244/982/credit_card_fraud_detection_with_a_neural_network.pdf>

*Data Analysis*:

The outlier detection problem has important applications in the field of fraud detection, network robustness analysis, and intrusion detection. Most such applications are high dimensional domains in which the data can contain hundreds of dimensions. Many recent algorithms use concepts of proximity in order to find outliers based on their relationship to the rest of the data. However, in high dimensional space, the data is sparse and the notion of proximity fails to retain its meaningfulness. In fact, the sparsity of high dimensional data implies that every point is an almost equally good outlier from the perspective of proximity-based definitions. Consequently, for high dimensional data, the notion of finding meaningful outliers becomes substantially more complex and non-obvious. In this paper, we discuss new techniques for outlier detection which find the outliers by studying the behavior of projections from the data set.

<http://ftp.cse.buffalo.edu/users/azhang/disc/SIGMOD/pdf-files/037/172-outlier.pdf>

*System Fraud Analysis*:

Fraudsters aim to use services without paying or illicitly benefit from the service in other ways, causing service providers financial damage. We report an experiment aimed at generating synthetic test data for fraud detection in an IP based video-on-demand service. The data generation verifies a methodology previously developed by the present authors [E. Lundin et al., (2002)] that ensures that important statistical properties of the authentic data are preserved by using authentic normal data and fraud as a seed for generating synthetic data. This enables us to create realistic behavior profiles for users and attackers. The data is used to train the fraud detection system itself, thus creating the necessary adaptation of the system to a specific environment. Here we aim to verify the usability and applicability of the synthetic data, by using them to train a fraud detection system. The system is then exposed to a set of authentic data to measure parameters such as detection capability and false alarm rate as well as to a corresponding set of synthetic data, and the results are compared.

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.4860&rep=rep1&type=pdf>

*Financial Statement Fraud Analysis:*

Data mining techniques have been used enormously by the researchers’ community in detecting financial statement fraud. Most of the research in this direction has used the numbers (quantitative information) i.e. financial ratios present in the financial statements for detecting fraud. There is very little or no research on the analysis of text such as auditor’s comments or notes present in published reports. In this study we propose a text mining approach for detecting financial statement fraud by analyzing the hidden clues in the qualitative information (text) present in financial statements.

http://thesai.org/Downloads/Volume3No12/Paper\_30-Financial\_Statement\_Fraud\_Detection\_using\_Text\_Mining.pdf