Data Mining for Fraud Detection

Roberto Marmo
Universita' Pavia, Italy

INTRODUCTION

The term fraud involves one or more people, they deliberately act secretly to take away some one valuable thing for their benefit. The statement fraud is as old as human being, it takes various forms. Because of expansion of modern technology, number and scenario of fraud are increasing dramatically. Reputation blemish and losses caused are primary motivations for technologies and methodologies for fraud detection. The Association of Certified Fraud Examiners (ACFE) estimates that 5% of revenue is lost to fraud and embezzlement every year. Applied to the U.S. GDP, this amounts to $730 billion every year (https://www.acfe.com/press-release.aspx?id=4294973129).

Fraud detection is based on huge logged data sets, it involves: monitoring and tracking the behavior of users, constructing models and profiles of fraudulent behavior.

The problem is very difficult because fraudsters takes many different forms and they are adaptive, in order to avoid security measure.

Data mining refers to extracting knowledge from a mass of data, in order to acquire the rules that provide decision support and determine what action should be taken. Data mining techniques are usually applied to fraud detection, due to their great flexibility, adaptivity and speed.

The aim of this contribution is to describe data mining techniques for fraud detection, in order to achieve higher cost savings and legal evidence. Solutions and recommendations regarding automobile insurance, credit card, e-commerce, telecommunications and others are also discussed.

BACKGROUND

The Association of Certified Fraud Examiners ACFE defines fraud as: "The use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization's resources or assets". Therefore, a fraud is a criminal deception, use of false representations to obtain an unjust advantage, or to injure the rights and interests of another.

The main reason behind the commitment of fraud is to achieve gain on false ground by an illegal means. This has a dramatic impact on the economy, law and the human moral values.

There are different kinds of fraud: internal/occupational or external. Internal frauds happen when an employee commits fraud against his or her organization. External frauds involve a wide range of schemes, including vendors, customers or thefts by other third parties.

Frauds can be classified according to three areas:

- **Motive:** The main reason behind commit fraud;
- **Means:** The nature or form of the fraud involved to achieve the goal;
- **Methods:** The facilities and tools that are used to commit fraud.

DOI: 10.4018/978-1-7998-3473-1.ch079

Diversity of fraud regards organizations, governments, and individuals such as external parties, internal employees, customers, service providers and suppliers.

Each fraud area has its specific characteristics and faces different challenges, therefore it is important to analyze the fraud scenario in order to establish:

- What is a fraudulent behavior and modus operandi over time;
- What is a fraudulent person;
- Degree of available knowledge about known fraud;
- Quantity and type of available data.

Planning audit strategies is a posteriori fraud detection problem with prevention purpose of analyzing historical audit data and constructing models of planning effectively future audits. A case study is presented by Bonchi (1999) which illustrates how techniques based on classification can be used to support the task of planning audit strategies.

Implementing the right technology is the key factor in order to analyze and to prevent fraud. Fraud prevention and detection are the proper protection mechanism against fraud, but fraud prevention alone is not sufficient, fraud detection is required to protect vital services in business systems. Some of the major benefits of investing in fraud detection software include:

- Detecting and preventing loss;
- Tracking of vulnerable points in a system that can leave it open to fraudsters, such as lack of data on behavior of users;
- Efficiently safeguarding organizational, security, operational and financial data;
- Providing a sense of reliability in an organization that has succeeded in avoiding catastrophes in an environment, where fraud attacks can be expected due to risk factors and complexity.

Applying data mining to fraud detection is suitable to search for patterns indicative of fraud, as stated in next paragraphs of this chapter.

A common obstacle for predictive fraud detection models is the rare occurrence of the target event. The proportion of identified fraud cases in the model input data is usually very small, makingit difficult to extract the fraud signal in the data.

It is difficult to provide estimates since some fraud may never be detected, and the operators are reluctant to reveal fraud losses, due to show an appearance of reliability and security in business operations, to avoid reputation blemish.

It is necessary to take into account cost of fraud detection and cost of fraudulent behavior, because stopping a fraud of few dollars can require a very expensive system. This is possible by introducing a decision layer on top level system, in order to decide the action taking into account factors, like the amount of transaction and the risk associated to user doing the transaction.

The development of new detections methods is more difficult, due to the severe limitation on privacy and on exchange of ideas due to confidentiality reasons. Moreover, data sets are not available and results are often not disclosed.