A survey of anomaly detection techniques in financial domain

Mohiuddin Ahmed\textsuperscript{a,*}, Abdun Naser Mahmood\textsuperscript{a}, Md. Rafiqul Islam\textsuperscript{b}

\textsuperscript{a} School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2600, Australia
\textsuperscript{b} School of Computing and Mathematics, Charles Sturt University, Albury, NSW 2640, Australia

HIGHLIGHTS

- Recently, in the financial sector, there has been renewed interest in research on detection of fraudulent activities.
- This paper presents an in-depth survey of various clustering based anomaly detection techniques and compares them from different perspectives.
- In addition, we discuss the lack of real world data and how synthetic data has been used to validate current detection techniques.

ARTICLE INFO

Article history:
Received 30 November 2013
Received in revised form 28 August 2014
Accepted 6 January 2015
Available online 14 January 2015

Keywords:
Clustering
Fraud detection
Anomaly detection

ABSTRACT

Anomaly detection is an important data analysis task. It is used to identify interesting and emerging patterns, trends and anomalies from data. Anomaly detection is an important tool to detect abnormalities in many different domains including financial fraud detection, computer network intrusion, human behavioural analysis, gene expression analysis and many more. Recently, in the financial sector, there has been renewed interest in research on detection of fraudulent activities. There has been a lot of work in the area of clustering based unsupervised anomaly detection in the financial domain. This paper presents an in-depth survey of various clustering based anomaly detection technique and compares them from different perspectives. In addition, we discuss the lack of real world data and how synthetic data has been used to validate current detection techniques.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Anomaly detection is an important aspect of data mining, where the main objective is to identify anomalous or unusual data from a given dataset. Anomaly detection is interesting because it involves automatically discovering interesting and rare patterns from datasets [1]. Anomaly detection has been widely studied in statistics and machine learning [2], where it is also known as outlier detection, deviation detection, novelty detection, and exception mining. Anomalies are important because they indicate significant but rare events, and they can prompt critical actions to be taken in a wide range of application domains. For example, an anomaly in an MRI image may indicate the presence of a malignant tumour [3]. Similarly, abnormal behaviour in a credit card transaction could indicate fraudulent activities [4], an unusual traffic pattern in a network could mean that a computer is hacked or under attack, e.g., using worms and Denial of Service (DoS) attacks [5]. In this paper, we focus on anomaly detection research in the financial domain. The key contribution of this paper is it provides a structured and broad overview of extensive research on anomaly based fraud detection using clustering techniques, while providing insights into the effectiveness of these techniques in detecting anomalies. The issue of dearth of financial data is also discussed, along with a discussion on how to generate synthetic data for this domain. Although, there are a few survey papers on anomaly detection (see Table 1), however, there is a lack of focus on unsupervised techniques to detect fraud, which is important since new fraudulent attacks are being invented every day, and existing signature based techniques [6,7] are unable to detect them.

Table 1 shows the set of methods, application domain and data covered by this survey and related surveys. Chandola et al. [2] provided an extensive survey encompassing various techniques and application domains but did not focus on clustering based fraud detection, instead they discussed signature based fraud detection techniques. Andrei et al. [8] also reviewed the fraud detection techniques using clustering but with limited details of the individual approaches. The survey also proposed a technique which was claimed to be clustering-based, while also using supervised learning methods, such as, Support Vector Machine [9]. Phua
2. Preliminaries

Section 7 concludes the paper. Data scarcity issue and how to address it using synthetic datasets. Key features of the discussed techniques. Section 6 describes the clustering based fraud detection. In each section, we summarize the detection techniques. Section 5 includes miscellaneous types of clustering based fraud detection. In each section, we summarize the key features of the discussed techniques. Section 6 describes the data scarcity issue and how to address it using synthetic datasets. Section 7 concludes the paper.

2.1. Areas of fraud

Fraud occurs when users misuse the resources provided by organization in an illegal way. However, organizations are interested in prompt action against fraudulent activities. Some of the most common applications of fraud detection are discussed below.

- **Credit card**: Everyday business, online shopping, electronic banking are largely dependent on credit cards. Meanwhile, the misuse of credit or debit card is on the rise and a source of common fraud. The advancement of technology made it easier to produce counterfeit cards and conduct fraudulent transactions. There are a number of ways to fraud credit cards. Capturing card details using hidden device in an ATM (Automated Teller Machine) or shoulder surfing are examples of how credit cards can be fraudulently used. Fig. 1 shows a forged card reading device installed over the original one. The device also has a pinhole camera which can record the video of a person entering PIN (Personal Identification Number) on the keypad. At a later time, the device is used to retrieve card details for illegal use. Credit card frauds anomalies can be identified using transactional records and generally correspond to any of a number of indicators including unusually high payments, purchase of extremely unusual items, and high frequency of purchase. Customer profiling as well as detecting anomalies in consumer's transactional records are possible ways to detect fraud. The challenge associated with detecting fraudulent transactions are detecting them in an online and unsupervised manner, so that the new types of fraud are detected as soon as they happen.

- **Mobile phone**: Mobile phone fraud includes a variety of scams that persuade consumers to buy various products. For example, offer of free ring tones or chance to win fantastic prizes by replying to messages or calls. After responding to this mobile call and signing up to this service, one could be left facing a huge phone bill. To prevent the misuse of mobile phone account, it is necessary to detect any unusual usage pattern. The basic technique is to monitor the usage pattern and create customer profile of each of the account. Consequently, any deviation from the common usage pattern will issue an alarm to warn the customer to avoid probable scams. The challenge here is to create a representation of a normal user profile to identify unusual behaviours.

- **Insurance claim**: Insurance fraud is simply another form of theft which involves trick or deception. A recent study estimated that insurance fraud costs more $2 billion annually in Australia [22]. Automobile related fraud is a common type of insurance fraud. Criminal rings of illegal claimants manipulate the claim processing system by submitting forged documents.
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات