Future Generation Computer Systems 55 (2016) 278–288

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A survey of anomaly detection techniques in financial domain

Mohiuddin Ahmed [a,*], Abdun Naser Mahmood [a], Md. Rafiqul Islam [b]

[a] School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2600, Australia

[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

* Corresponding author.
E-mail addresses: Mohiuddin.Ahmed@student.adfa.edu.au (M. Ahmed), a.mahmood@adfa.edu.au (A.N. Mahmood), mislam@csu.edu.au (Md. Rafiqul Islam).

http://dx.doi.org/10.1016/j.future.2015.01.001
0167-739X/© 2015 Elsevier B.V. All rights reserved.

Table 1
Comparison of this survey to other related surveys.

Survey	Supervised	Unsupervised	Clustering	Fraud detection	Data issue
Ahmed[a]		√	√	√	√
Andrei [8]		√	√	√	
Chandola [2]	√	√		√	
Patcha [11]	√	√	√		
Bakar [15]	√	√			
Phua [10]	√	√	√	√	
Hodge [12]	√	√	√		
Markou [13]	√				
Beckman [14]	√				

[a] This survey.

Table 2
The summary of fraud profiles on different sectors [20].

Sector	Level of financial loss	Investment on anti-fraud
Natural resource	Medium	High
Medical, Healthcare	High	High
Retail, Wholesale	Medium	Low
Financial services	High	Medium
Manufacturing	High	Low
Information technology	Medium	Low

et al. [10] also categorized, compared, and summarized a good number of published technical and review articles in automated fraud detection, however, newer attacks have emerged in the last 10 years since the report was published. Patcha et al. [11] and Hodge et al. [12] also presented various anomaly detection techniques based on supervised, unsupervised and clustering methods but not for fraud detection. Markou et al. [13] and Beckman et al. [14] also provided surveys on anomaly detection but only on supervised methods. Interestingly, none of these surveys discussed the issues of training and test data for anomaly detection which we have addressed in this survey.

Section 2 provides a background on fraud detection and financial losses occurred due to fraud and, also discusses the fundamentals of clustering based anomaly detection. Sections 3 and 4 include discussion on partition and hierarchical clustering based fraud 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. Preliminaries

According to The American Heritage Dictionary [16], '*Fraud is defined as a deception deliberately practiced in order to secure unfair or unlawful gain*'. Fraud and financial crimes are a category of larceny that takes place when a person or entity embezzles money or property, or uses them in an illicit manner, with the intent to gain benefit from it. Davia et al. [17] stated, '*Fraud always involves one or more persons who, with intent, act secretly to deprive another of something of value, for their own enrichment*'. Wells et al. [18] stressed '*deception*' as the linchpin to fraud. Occupational fraud and abuse may be defined as: '*The use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organizations resources or assets*'. This definition encompasses a wide variety of conducts by executives, employees, managers and principals of organizations. FBI has been treating such crimes as white collar crimes [19] and a number of other organizations are also dedicated to diminish such crimes.

According to the 2013/2014 global fraud report from Kroll [20], the level of fraud increased by every measure in the past 12 months. Overall, 70% of companies reported suffering from at least one type of fraud in the past year, up from 61% in the previous poll. Individual businesses also faced a more diverse range of threats on average, compared with the threats in 2012. Finally, the economic cost of these crimes mounted, increasing from an average of 0.9% of revenue to 1.4%, with one in ten businesses reporting a cost of more than 4% of revenue. Table 2 shows the summary of fraud profiles from different sectors.

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

ISIArticles
مرجع مقالات تخصصی ایران

✓ 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