



Information Technology and Quantitative Management (ITQM 2016)

Application of outlier mining in insider identification based on Boxplot method

Aihua Li*, Mengyan Feng, Yanryu Li, Zhidong Liu

School of Management science and Engineering, Central University of Finance and Economics, Beijing, 100081, China

Abstract

With the rapid development of Chinese capital market, securities and futures trading activities have become increasingly active. Insider trading, disclosure of insider information and other illegal and criminal activities are prone to happening. This paper explores the application of Boxplot in outlier mining. Moreover, we put forward the improved scheme based on the Boxplot itself by using R software to add density curve and scatter for the purpose of more intuitive. And the result shows the feasibility of the Boxplot method for outlier

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of ITQM 2016

Keywords: Outlier; Data mining; Boxplot; Insider Trade; Stocks

1. Introduction

With the rapid development of Chinese capital market, securities and futures trading activities have become increasingly active. Insider trading, disclosure of insider information and other illegal and criminal activities are prone to happening. Peng has a more comprehensive definition of insider trading. He thinks insider trading refers to an activity that people who get sensitive information of securities, which is confidential and has a significant impact on the prices of securities, through improper or intimate access, use this information to purchase a large number of the securities and spread this information to recommend others to conduct securities trading in the sensitive period [1]. Insider trading disrupts security market rules, which has great harm. What's more, the serious one will break the whole balance of the capital market.

Chinese securities market is relatively young, and market discipline mechanism has not been fully formed, which lead to insider trading sometimes. In July 2008, GF Securities chairman Dong was involved in insider trading case, becoming the first case of insider trading in China. In May 2010, Huang was involved in the

* Corresponding author. Tel.: +86-10-62288622; fax: +86-10-62288622.
E-mail address: aihuali@cufe.edu.cn.

Zhongguancun Stock (000931) insider trading case, which is the highest amount of illegal income and fines of insider trading in China. In April 2011, Zhongshan Public Utilities case became the first insider trading case involved with officials. In April 2012, Shenzhen Topway Video Communication insider trading case became the case which has the largest number of involved people in China, up to 15.

A lot of researchers are looking for ways to identify insider trading behavior. Nowadays, data mining has been widely used in various industries, including the securities industry. Because data mining has been the forefront in the financial field, detecting outliers in the stock market is still in the exploratory stage. The data of insider trading does not belong to the scope of normal transaction data, so the detection of insider trading belongs to the outlier mining category, which is the focus of this paper.

2. Research Status of Outlier Mining

Barnett has made a definition of anomaly points: an anomaly point is the point that the data concentration is obviously different from others. The different measurement methods lead to the different results of the abnormal point [2]. Sometimes people are interested in outliers, because it contains some important information. Simply culling will make us lose some significant information.

Outlier mining has already been applied in the stock market. The data is basically time series, so there are many scholars focusing on outlier mining algorithm from the perspective of time series. WTMM method proposed by Struzik and Siebe is a wavelet method. It pointed out that the method was feasible in detecting the financial market of outliers in time series, but it did not mention whether it could be used in the high-frequency data stream [3]. Ane and Rangau proposed an AR-GARCH model to detect abnormal returns of Asia-Pacific stock markets [4]. Zhao proposed a method to detect the outlier of stock time series based on density and proved to be feasible and effective on SSE Composite Index [5]. Cao et.al demonstrated applicability the VSO outlier system for supporting detection of outliers in big data stream, and the system performed well when it helped analysts to identify and respond in the high volume streams of stocks [6].

Chinese scholars have also been studying the anomaly detection of the stock market. Sun et.al proposed an outlier pattern mining considered as an optimization segmentation problem by using fractal theory, from the view point of outlier affecting orderliness of data set of time series [7]. Du and Zhang presented a detection algorithm for outlier subsequence based on Local Linear Mapping (LLM) to improve the effectiveness of outlier subsequence detection algorithm for time series [8]. Du et.al proposed an algorithm for the prediction of stock data to find the abnormal data, with the introduction of Gaussian process machine learning method. And they used colony algorithm to solve the adaptive mechanism for the parameters of Gaussian process [9].

And there are also some applications of boxplot in anomaly detection. Sim *et al.* introduced outlier labeling with boxplot procedures, improving the feasibility of the method [10]. Dovoedo and Chakraborti studied boxplot-based outlier detection for the location-scale family and do a simulation study on the performance of the outlier detection rules [11]. However there are nearly no studies about the application of boxplot on insider trading detection.

3. Outlier Mining Based on Boxplot

3.1. The Introduction of Boxplot

Boxplots are among the most widely used exploratory data analysis (EDA) tools in statistical practice. It's a visualization method of the degree of dispersion of the data, which is invented by the American mathematician John W. Turkey in 1969. Boxplot can intuitively reflect outliers. Boxplot outlier judgment scope is outside of $Q3+1.5IQR$ and $Q1-1.5IQR$. This kind of determination method seems to be random, but in fact it is a method derived from a great deal of experience, proved to be effective and valid. Usually, the empirical method of the

Download English Version:

<https://daneshyari.com/en/article/488330>

Download Persian Version:

<https://daneshyari.com/article/488330>

[Daneshyari.com](https://daneshyari.com)