Information Technology and Quantitative Management (ITQM 2017)

Market Manipulation Detection Based on Classification Methods

Aihua Li^a,*, Jiede Wu^a, Zhidong Liu^a

^aSchool of Management Science and Engineering, Central University of Finance and Economics, Beijing, 100081, China

Abstract

In this paper, we use supervised machine learning methods to detect the market manipulation in China based on the information released by China Securities Regulation Commission (CSRC) and data in the security market. Among the supervised machine learning, we mainly use classification methods to detect the anomaly from the daily and tick trading data of manipulated stocks. As a result, we find that the supervised machine learning methods are good at detecting market manipulation from daily trading data and have poor performance on tick data, based the measure method of accuracy, sensitivity, specificity and area under the curve (AUC). The best used supervised machine learning models are K-Nearest Neighbor (KNN) and Decision Tree (DT) which have over 99% of them.

© 2017 The Authors. Published by Elsevier B.V. Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

Keywords: Market manipulation, Classification methods, Anomaly detection;

1. Introduction

Generally, market manipulation was defined that a deliberate attempt to intervene in the free and fair operation of the market which can create artificial, false or misleading appearances with respect to price of security [1]. Market manipulation, broadly defined, has existed since the infancy of financial markets [2], which has become an important issue of emerging and developed market. Some researchers regarded that market manipulated could enhance the mobility. More focused on the disadvantage aspect such as that manipulated security market not only distorted the prices and transactions in the security market, but also undermined the function of security market. What’s more, many investors would lose significantly because of most manipulators’ illegal profit making. Thus, it is meaningful to detect the action before or after it happened.

There has been increasing research on data mining techniques in detecting market manipulation. Allen and Gale [3] classified three kinds of types, which are manipulation as action-based manipulation, information based manipulation, and trade-based manipulation. Here, we studied how to detect the trade-based manipulation that a trader attempts to manipulate a stock price simply by buying and then selling (or vice versa), without releasing any false information or taking any other publicly observable actions designed to alter the value of security, by using supervised and unsupervised machine learning methods based on information about the manipulated stocks from China Securities Regulation Commission (CSRC). To obtain this goal, daily stock data and tick stock data of 64 manipulated stocks published in CSRC website were selected in this paper. After data cleaning and transformation,
supervised machine learning models and rules were built, such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Tree (DT), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (LR), Artificial Neural Networks (ANN) and so on. What’s more, K-folds cross validation methods were used here to test the robustness of these supervised methods.

The main contributions of our work are as follows. First, currently, most scholars’ study of detecting market manipulation are theoretical and pattern description which is still hard to accurately and fast detect market manipulation. Here, in order to increase the possibility and efficiency of detecting market manipulation, we apply different supervised machine learning models to detect market manipulation in real time stocks data and find that the most effective supervised machine learning methods are K-Nearest Neighbor and Decision Tree. Second, we use daily and tick real time trading stock data to evaluate those supervised machine. Third, it would be benefit to the investors and regulators on stock market because of the high performance of market manipulation detection which is over 99% accuracy, sensitivity and specificity for some supervised machine learning models.

The remainder of the article is organized as follows. In section 2, we summarized the main result of prior study on market manipulation detection. In section 3, we introduced 7 kinds of supervised machine learning models that was used on detecting the market manipulation. Section 4 is the main section, in which we used these supervised machine learning models to detect market manipulation and evaluate the result of these models. Section 5 gave the conclusion and some future works.

2. Literature Review

Allen and Gale [3] and Jarrow [4] were the first researchers to study manipulation. After studying the history of stock-price manipulation, Allen and Gale classified the manipulation as action-based manipulation, information-based manipulation and trade-based manipulation[3]. They defined trade-based manipulation as manipulating stocks through actual trading or trading orders by distorting stock market prices, rather than changing company values or issuing false information. In a dynamic model of asset markets, by investigating large traders’ manipulation of trading strategies in the securities market, Jarrow found that large investors had a greater impact on stock prices [4]. Carhart and Reed [5] found that a large trader can manipulated the market or lure the market to “manipulate” their own trades. And he developed a theory for option pricing in an economy with a larger trader showing that the standard option pricing models holds, but with a random volatility price process. Hanson and Oprea [6] developed an experimental model to study whether the manipulators could distort the prices in a prediction market.

Some researchers have attempted to detect manipulation in different methods. Aggarwal and Wu [7] studied the US stock market from 1990 to 2001 to detect market manipulation and found that the manipulated stock had abnormal stock prices, liquidity, volatility and return. H.Öğüt and Aktaş [8] compared the performance of Artificial Neural Networks (ANN) and Support Vector Machine (SVM) with discriminant analysis and logistics regression on detecting the market manipulation and found that data mining techniques were better than multivariate techniques. According to Mongkolnavin and Tirapat [9], association rules were applied to detect mark-the-close in intraday trades from the Thai Bond Market Association. Price variation in the market and behavior of investors were integrated to analyze warning signals in real time. And the method showed that a list of investors were in the market, who perhaps are manipulators. Fallh and Kordlouie [10] used logit model, artificial neural network, and multiple discriminant analysis to create stock price manipulation models in Tehran stock exchange, and the performances of three aforesaid models were effective. The selected data were thoroughly analyzed by runs test, skewness test, and duration correlative test and the selected samples were divided into two sets: manipulated and non-manipulated companies. The factors that were related to stock price manipulation were defined such as: size of company, P/E ratio, liquidity of stock, status of information clarity, and structure of shareholders. In Yangs paper [11], logistic regression model was chosen to detect stock price manipulation activities in Shanghai and Shenzhen market. They analyzed independent variables based on primary component analysis, which increased performance for forecasting the model. And the model was better than the linear regression model. Cao and McGinnity [12] proposed the Adaptive Hidden Markov Model with Anomaly States (AHMMAS) to detect intraday stock price manipulation activities. The stock tick data were level 2 data from NASDAQ and London stock exchange and the model was tested with simulated data and real market data. The performance evaluation of AHMMAS outperforms other benchmark algorithms such as Gaussian Mixture Models (GMM), K-Nearest
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