Chapter 9
A Review on Applied Data Mining Techniques to Stock Market Prediction

Neslihan Fidan
Istanbul University, Turkey

Beyza Ahlatcioglu Ozkok
Yildiz Technical University, Turkey

ABSTRACT

A portfolio manager considers forecasting the asset prices and measurement of the market risk of an underlying asset. Financial institutions produce datasets to handle their problems by using data mining tools. Recently new technologies have been developed for tracking, collecting, and processing financial data. From a data analysis point of view, this chapter reviews the published articles based upon predictive data mining applications to stock market index. It is observed that hybrid models that combine data mining techniques or integrate an algorithm to a method work efficiently. Finally, the chapter provides likely directions of future researches.

INTRODUCTION

Due to the stock market volatility needs to be captured from hidden information in the database, forecasting stock market movement for intelligent decision making is a quite difficult task. One of the most important problems in stock market is finding efficient mining techniques to have useful information in the stock market movements for investment decisions. A great number of stocks are traded in the stock markets. Therefore the amount of valuable data is generated by the market and this valuable data have some specific dynamic features, for instance expanding in every moment.

Data mining deals with extracting meaningful information from the large amount of data. Investors use historic data of assets, in order to decide how much of their capitals should be al-

located between the assets. They try to find hidden patterns of the historic data, which provide information about future movements of assets. Forecasting financial time series is a very complex and challenging problem. Therefore the forecasting procedure requires specific methods of data mining (Kovalerchuk & Vityaev, 2005).

Data mining involves looking in the data for such factors as (Chung & Gray, 1999):

- **Associations:** Things done together.
- **Sequences:** Events occurring over time.
- **Classifications:** Rules for recognizing patterns.
- **Clusters:** Defining new groups.
- **Forecasting:** Predictions from time series e.g., stock price movements.

In Association, the relationship of a particular item in a data transaction on other items in the same transaction is used to predict patterns. In Classification, the methods are intended for learning different functions, which map each item of the selected data into one of a predefined set of classes. Cluster analysis takes ungrouped data and uses automatic techniques to put this data into groups. Clustering is unsupervised and does not require a learning set. Prediction analysis is related to regression techniques. The key idea of regression analysis is to discover the relationships of an independent variable with the dependent and other independent variables. Sequential Pattern analysis seeks to find similar patterns in data transaction over a business period (Olson & Delen, 2008).

Classification and prediction are two forms of data analysis, which can be used to extract models describing important data classes or to predict future data trends (Han & Kamber, 2006). Practitioners and researches pay attention to analysis and prediction of future values and trends of the financial market. Many studies on stock market prediction using data mining techniques have been performed during the past decade. The stock market data has tremendous noise and complex dimensionality (Kim & Han, 2000). Some data mining applications have been demonstrated on stock markets (Keogh & Kasetty, 2003; Hajizadeh, Ardakani & Shahrabi, 2010; Atsalakis & Valavanis, 2009).

Many researches were presented on financial data mining to solve the problem of knowledge extraction from a database and predict future information. There are comprehensive references including application of data mining techniques on business environments (Kovalerchuk & Vityaev, 2005; Armstrong, 2001; Giudici & Figini, 2009; Bozdogan, 2004). The purpose of all these studies is to review of the data mining techniques, which are applied for stock market prediction in a dynamic and a complex structure. We concentrate on the predictive applications based on stock market index to limit the scope of this paper. The distinctive feature of this review is concentrating on the applications, which includes the prediction of stock market index. Also we try to extract some hybrid model that enhances the data mining techniques by combining them. Accordingly, the sections are organized as follows. The next section summarizes the most used techniques in descriptive and predictive approach for financial data mining. Following section provides the review of the papers with applications in the stock markets. We survey the papers in three subsections under the applications. In the first subsection, we overview the papers which include predictive applications and compare the data mining techniques over the efficiency of predicting stock price movements. The second subsection focuses on the major problem with working on the financial data. Every one of the techniques has its own drawbacks, and the market has tremendous data and complexity correspondingly. The last subsection reviews the papers that include hybrid models. Our survey is finalized with the conclusion by summarizing the plausible future works.

Related Content

Assessing Organizational Readiness for Implementing ERP System Using Fuzzy Expert System Approach

Amir Beirami Hajilari, Mahsa Ghadaksazand Gholamreza Soltani Fasghandis (2017). *International Journal of Enterprise Information Systems (pp. 67-85).*

www.irma-international.org/article/assessing-organizational-readiness-for-implementing-erp-system-using-fuzzy-expert-system-approach/176392

IP Layer Client Puzzles: A Cryptographic Defense against DDoS Attack

Genti Daciand Rezarta Jaupi (2013). *Enterprise Business Modeling, Optimization Techniques, and Flexible Information Systems (pp. 14-25).*

www.irma-international.org/chapter/layer-client-puzzles/77957

Eliciting User Input for Requirements on Personalization: The Case of a Dutch ERP System

Lex van Velsen, Corrie Huijsand Thea van der Geest (2008). *International Journal of Enterprise Information Systems (pp. 34-46).*

www.irma-international.org/article/eliciting-user-input-requirements-personalization/2150

Ubiquitous Integration Architectural Issues

João Henriquesand Paulo Tomé (2015). *Improving Organizational Effectiveness with Enterprise Information Systems (pp. 32-50).*

www.irma-international.org/chapter/ubiquitous-integration-architectural-issues/133085

Approaches for Automating ERP Category Configuration for SMEs

Klaus Wölfel (2017). *Enterprise Information Systems and the Digitalization of Business Functions (pp. 66-98).*

www.irma-international.org/chapter/approaches-for-automating-erp-category-configuration-for-smes/177339