

CLASSIFICATION WITH DEGREE OF IMPORTANCE OF ATTRIBUTES FOR STOCK MARKET DATA MINING

¹Rashid Hafeez Khokhar, ²Mohd Noor Md Sap
Faculty of Computer Science and Information System
University Technology of Malaysia, K.B. 791
81310 Skudai, Johor, Malaysia
Tel: (607)-5532419, Fax: (607) 5565044
¹rashid@siswa.utm.my, ²mohdnoor@fksm.utm.my

Abstract

With the increase of economic globalization and evolution of information technology, financial time series data are being generated and accumulated at an unprecedented pace. As a result, there has been a critical need for automated approaches to effective and efficient utilization of massive amount of financial data to support companies and individuals in strategic planning and investment for decision-making. Many statistical and data mining techniques have been used to predict time series stock market. However, most statistical and data mining methods suffer from serious drawback due to requiring long training times, results are often hard to understand, and producing inaccurate predictions. We present another modification of fuzzy decision tree (FDT) classification techniques that aims to combine symbolic decision trees in data classification with approximate reasoning offered by fuzzy representation. The intent is to exploit complementary advantages of both: ability to learn from examples, high knowledge comprehensibility of decision trees, and the ability to deal with uncertain information of fuzzy representation. In particular, the proposed predictive fuzzy decision tree is based on the concept of degree of importance of attribute contributing to the classification. We extend this idea with the expressive power of fuzzy reasoning method. After constructing predictive FDT, weighted fuzzy production rules (WFPRs) can be extracted from predictive FDT. The predictive FDT has been tested using three data sets including KLSE, NYSE and LSE. The experimental results show that predictive FDT algorithm can generate a relatively optimal tree without much computation effort (comprehensibility), and WFPRs have a better predictive accuracy of stock market time series data.

Keywords: Data Mining, Classification, Time Series, Fuzzy Logics, Decision Tree

1. Introduction

Many attempts have been made for meaningful prediction from real time stock market data by using data mining and statistical techniques such as Support Vector Machine [1, 2], and Linear and Non-Linear Statistical Models [3, 4], Neural Networks [5, 6]. Alan Fan et al., [2] use Support Vector Machine (SVM) to stock market prediction. The SVM is a training algorithm for learning classification and regression rules from data [7]. However the predictive accuracy of SVM achieved by [2] in stock market is relatively lower than other classification applications [8, 9]. Also the existing relationship between the future stock returns and its accounting information, one would expect it to be a weak relationship. Support Vector Regression (SVR) is the extended form of SVM that can be applied in financial time series prediction [8, 9]. In financial data, due to the embedded noise, one must set a suitable margin in order to obtain a good prediction [9]. Haiqin et al., [9] has extended the standard