Correlation between the Economy News and Stock Market in Turkey

Sadi Evren Seker, Department of Computer Engineering, Istanbul University, Istanbul, Republic of Turkey
Cihan Mert, Department of Electrical Engineering, University of Texas at Dallas, TX, USA
Khaled Al-Naami, Department of Computer Science, University of Texas at Dallas, TX, USA
Nuri Ozalp, Turkish Science Foundation, Istanbul, Republic of Turkey
Ugur Ayan, Turkish Science Foundation, Istanbul, Republic of Turkey

ABSTRACT

Depending on the market strength and structure, it is a known fact that there is a correlation between the stock market values and the content in newspapers. The correlation increases in weak and speculative markets, while they never get reduced to zero in the strongest markets. This research focuses on the correlation between the economic news published in a highly circulating newspaper in Turkey and the stock market closing values in Turkey. In the research several feature extraction methodologies are implemented on both of the data sources, which are the stock market values and economic news. Since the economic news is in natural language format, the text mining technique, term frequency – inverse document frequency is implemented. On the other hand, the time series analysis methods like random walk, Bollinger band, moving average or difference are applied over the stock market values. After the feature extraction step, the classification methods are built on the well-known classifiers support vector machine, k-nearest neighborhood and decision tree. Moreover, an ensemble classifier based on majority voting is implemented on top of these classifiers. The success rates show that the results are satisfactory to claim the methods implemented in this study can be spread to future research with similar data sets from other countries.

Keywords: Classification, k-Nearest Neighbors Algorithm (kNN), Sentimental Analysis, Stock Market Analysis, Support Vector Machines (SVM), Text Mining, Time Series Analysis

1. INTRODUCTION

This study is built on two datasets. The first one is the economic news from one of the high-circulating newspapers in Turkey, which have special pages for economic news. We have collected only economic news, which are separate from other news like sports or local news etc. The second dataset is the stock market closing values and we have applied several time series
analyses over it. The properties of the dataset will be explained in the experiments section. We have processed the news text via the text mining approach called term frequency – inverse document frequency (TF-IDF) which will be explained in the methodology section. On the other hand, we have processed the stock market closing values by using time series analysis, which are random walk (RW), Bollinger band (BB), relative strength index (RSI), moving average convergence / divergence (MACD), momentum, rate of change (ROC), acceleration, difference and their variations. Finally we have investigated the correlation between these two feature vectors by using the support vector machines (SVM), k-nearest neighborhood (KNN) and decision tree-based classification (C4.5). Furthermore an ensemble classification based on majority vote learning (MaVL) is implemented on top of these three classifiers, which is discussed in the section of classification. Also in this paper, we discuss the implementation details and the methodology of evaluation over the classification results, in the evaluation section.

2. PROBLEM STATEMENT

We have two datasets, the economy news and stock market closing values, and we have applied a layered approach in this study, as illustrated by Figure 1. At the bottom level starting from the datasets, we build a feature extraction methodology. For the economic news, which is in natural language, we have applied sentimental analysis via the text mining methodologies. On the other hand, we have applied the random walk method for the feature extraction from the stock market closing values.

The correlation between news and the stock market is one of the indicators of the speculative markets (Nikfarjam, Emadzadeh, & Muthaiyah, 2010).

One of the difficulties in this study is dealing with the natural language data source, which requires a feature extraction. The other difficulty is dealing with a stock market value, which is considered as a signal. The size of data we are dealing with, which can be considered big data, is also problematic. The dataset holds 131,248 distinct words and when the feature vector of Figure 1. Bottom up overview of study
Equilibria for Traffic Network with Capacity Constraints
www.igi-global.com/chapter/equilibria-for-traffic-network-with-capacity-constraints/107286?camid=4v1a