

要 旨

テキストマイニングを用いた株価予測

仲矢 浩二

株価予測の際に使用される手法として、ニューラルネットワーク、遺伝的アルゴリズム、遺伝的プログラミング、Q-Learning、フラクタル次元、Support-Vector-Machine などの研究が行われている。それらの研究は株価や為替などの定量的な数値データのみの場合が多かった。しかし、実際のファンドマネージャは数値データのみではなく、むしろ積極的に定性的なニュース記事の情報も用いて株式の売買を行っている。そこで本研究では定量的データと定性的データを用いた株価予測を行うことを目的として、テキストマイニングを用いた。定性的データとして、NIKKEI NET、Infoseek と MSN 産経ニュースの 10 月 1 日から 12 月 22 日までの経済に関連する記事を 170 記事入手し、株価に影響を与える可能性のある単語を抽出し 121 単語の単語ベクトルを作成する。定量的データとして記事発表から 10 日分前の 9 月 15 日から 12 月 22 日までの日経平均終値の時系列データを使用する。株価予測にはニューラルネットワークによるテキストマイニングを使用する。ニューラルネットワークの学習手法にはバックプロパゲーションを使用する。実験用データとして 85 セットの単語ベクトルと株価の時系列データを使用する。教師データを翌日の株価が上昇か下降の 2 値とする。検証用データとして残りの 85 セットの単語ベクトルと株価の時系列データを使用する。単語ベクトルと株価の併用、単語ベクトルのみ、株価の時系列データ、認識率の比較を行った。その結果、単語ベクトルと株価の併用の認識率が 10 ポイント程度高い結果となる事を示す。

キーワード ニューラルネットワーク、株価予測、単語ベクトル

Abstract

Stock price forecast using text mining

Koji Nakaya

The stock prices forecast was done by the numeric character data. . However, Fund manager is having dealings over the stocks by using information in not only the numeric character data but also a qualitative news article. Then, text mining was used aiming to forecast stock price that used the quantitative data and qualitative data in this research. Qualitative data extracts the word with the possibility of influencing stock prices by using the article that relates to the economy of NIKKEI NET, Infoseek, and MSN Sankei news from October 1st to December 22nd by 170 articles and makes word vector of 121 words. The quantitative data uses the longitudinal data of Nikkei average closing share price from September 15th to December 22nd the tenth ago from the article announcement. Text mining by the Neural Network is used for the stock prices forecast. The learning technique of Neural Network uses Back Propagation. Verification data uses the longitudinal data of word vector and stock prices of 85 sets. Whether the teacher data is a rise of next day's stock prices is assumed to be binary of the descent. Data for the verification uses the longitudinal data of word vector and stock prices of the remainder of 85 sets. Only using together and word vector of word vector and stock prices compared the recognition rates of the longitudinal data of stock prices. The results show that using of word vector and stock prices together became the highest result.

key words Neural Network, stock prices forecast, word vector