

# 要旨

## ディープラーニングによる 経済記事テキストデータを用いた株価予測

奥村 順哉

現在，機械学習を用いた株価予測が注目を集めており，数値データだけでなくテキストデータを用いた株価予測も行われている．しかし，テキストデータの処理では，対象となるテキストの特徴の表現が重要である．このような特徴量の設計の問題に対して有効な手法として，多数の階層を有する構造のディープラーニングが提案されており，自然言語処理にも応用されている．本研究では，ディープラーニングの代表的な手法である DBN(Deep Belief Network) を用いて経済記事の単語ベクトルから株価の予測を行い，多くの実問題で用いられている SVM(サポートベクトルマシン) との比較を行い，テキストマイニングにおけるディープラーニングの有効性を検証することが目的である．学習データとして 2014 年のニュース記事 (全 64564 記事) を使用する．各記事に形態素解析を行い，名詞，動詞，形容詞を抽出し，各単語の出現回数を日付毎に求め，単語ベクトルを生成する．教師データとして，日経平均終値の上昇が前日比+0.5%より増加すれば上昇，前日比-0.5%より減少すれば下降，その間を値動きなしとした 3 クラスのデータを用いる．単語ベクトルの各成分と教師データとの相関を求め，相関係数の絶対値の高い順に単語を選出し，SVM (線形およびガウシアン (RBF) カーネル) と DBN について k-分割交差確認法を用いて比較を行った．結果として，DBN が学習時間については SVM より劣るが，識別率については DBN の方が 3 ポイント高くなることを示す．

キーワード 株価予測，Deep Belief Network，テキストマイニング，単語ベクトル

# Abstract

## Prediction of Stock Market Price from Economical News Text Data Using Deep Learning

Junya Okumura

Recently, stock prices prediction using machine learning has been studied. Stock traders use not only numerical data, but also the qualitative data. Qualitative data, however have a problem that it is difficult to extract features. Deep learning has been proposed and it can extract the feature automatically using unsupervised learning as a pretraining. This study propose the prediction method of stock prices using DBN(Deep Belief Network), which is a major deep learning. Word vectors are used as input vector. Word vectors are retrieved from newspaper articles in the Internet. The training data are given as three categories , price up, price down, and same price. Effective words for prediction are selected by the correlation to training data. Proposed method is compared with SVM(Support Vector Machine). As a result, learning of DBN is much slower than that of SVM, and precision of DBN is higher than that of SVM. Therefore DBN is effective for stock prices prediction.

**key words** stock prices prediction , Deep Belief Network , text mining , word vector