

深層学習を用いた新聞記事分析による市場動向予測

松本 一樹*

* 中部大学 工学部 情報工学科

松井 藤五郎†

† 中部大学 生命健康科学部 臨床工学科

1 はじめに

金融市場の動向を予測する研究は古くから行われている。これまで金融市場の動きを予測する研究として、過去の取引実績の時系列パターンから株価を予測するテクニカル分析や、財務諸表などから株価を予測するファンダメンタル分析が主流であった。これらの手法はいずれも定量的データを用いた分析である。しかし、定量的データを用いた分析だけでは予測できない市場の動きがある。例えば、ある企業の不正会計が発覚したとき、その企業の株価は急落する。このような株価の動きは定量的データの分析だけで予測することはできない。そこで、近年テキスト分析を用いて金融市場の動きを予測する研究が盛んに行われている。テキスト分析による金融市場の動向予測において、株価や金利といった市場データである目的変数が時系列を成すように、説明に用いられるテキストも時系列をなしている。そのため、新聞などの定期的に発行されるテキスト・データは、一種の時系列データであると考えられる [1]。

また、近年パターン認識の分野において、近年深層学習が注目を集めている。深層学習は、画像認識や音声認識において高い認識精度を発揮しており、テキスト分析においても従来手法と比べ高い性能であることが確認されている [2]。

そこで、本研究では、新聞記事を時系列データと捉え、テキストの差分に着目した分析を行い、深層学習を用いて株価が上昇したか下落したかを予測する。本論文では、従来用いられていたサポートベクターマシンによる結果と比較することで、深層学習を用いた新聞記事分析による市場動向予測の有効性を確認する。

2 深層学習を用いた時系列テキスト分析による市場動向予測

本論文では、新聞記事を対象とし、深層学習を用いた時系列テキスト分析によって市場動向を予測する手

Market trend prediction by newspaper article analysis using deep learning

*Kazuki Matsumoto †Tohgoroh Matsui

*Department of Computer Science, College of Engineering, Chubu University

†Department of Clinical Engineering, College of Life and Health Sciences, Chubu University

法を提案する。

対象とする新聞記事は、日本経済新聞の見出しとリードとし、市場動向に影響を与えるテキストであると考えられる予測対象日の前営業日の夕刊から当日の朝刊を結合したテキストを用いる。

テキスト分析に用いる深層学習は、ニューラルネットワークを多層化したものであり、画像認識や音声認識で高い認識精度を発揮している。テキスト分析においても従来手法と比べて高い性能であることが確認されている。深層学習への入力、先行研究 [3] で作成された特徴量ベクトルのデータを CSV 形式に変換したものをを用いる。深層学習の出力は、予測対象日の日中の利益率が正または 0 のとき (0,1)、そうでないときは (1,0) とする。新たに予測を行うときは、テキストを特徴量ベクトルに変換してニューラルネットワークへ入力し、その出力 (o_1, o_2) を求め、 $o_1 < o_2$ なら利益率は正になると予測し、そうでないなら負になると予測する。

3 実験結果と考察

日本経済新聞を対象として、予測対象日の前営業日の夕刊から予測対象日の営業日の朝刊までを 1 つのテキストとし、その見出しとリードを用いた。予測対象は 2014 年から 2015 年までの東証株価指数 (TOPIX) 連動型上場投資信託 (ETF) とし、予測対象日の寄りから引けにかけて TOPIX ETF の取引価格が上昇するか下落するかを予測した。訓練データの期間は予測対象日の直近過去 5 年間とした。

特徴語はテキストの差分に基づいて新出、続出、消滅の 3 つの出現パターンに分類する。分類した特徴語を出現パターンごとに出現回数をカウントし、50 回以上出現したパターンのみを対象とし、パターンが出現した日に上昇した割合が 0.55 以上または 0.45 以下のパターンを特徴として用いた。深層学習のツールは TensorFlow を使用した。深層学習のユニット数は、入力層を特徴語数分のユニット、出力層を 2 ユニット、隠れ層は 2 層とし、1 層のユニット数を 64 ユニットとした。また、活性化関数には ReLU を用いた。

比較として、朝刊のみを用いたパターン、見出しのみを用いたパターン、テキストの差分を使わず語が出

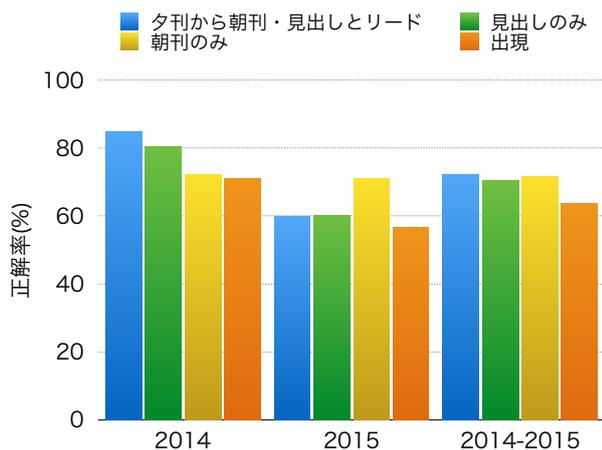


図1: 予測正解率

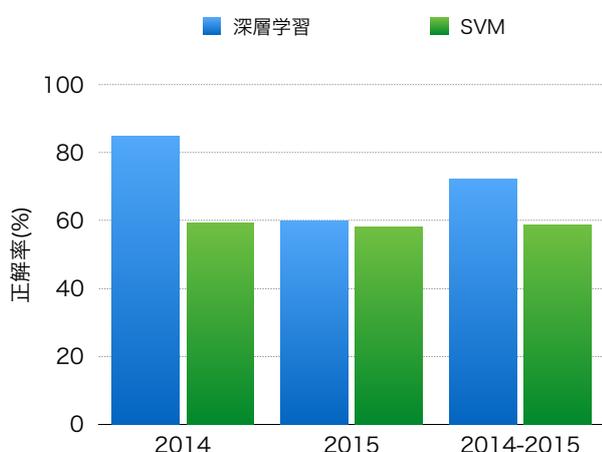


図2: 深層学習とSVMの比較

現したかどうかを特徴として用いたパターンで予測を行った。

実験結果を図1に示す。2014年から2015年を通した全体の予測正解率は、夕刊から朝刊の見出しとリードを用いたパターンが最も高く72.3%であった。テキストの差分を使わず語が出現したかどうかを特徴として用いたパターンでは、予測正解率が63.7%となり最も低くなった。見出しのみを用いたパターンでは予測正解率が70.3%となり、朝刊のみを用いたパターンでは71.5%となった。提案手法である夕刊から朝刊の見出しとリードを用いたパターンにおいて、予測正解率の低かった2015年においても予測正解率は59.8%であり、予測正解率が55%以上あれば有用であると考えられることから、提案手法は有用であるといえる。

また、テキストの差分を用いなかったパターンでは、予測正解率が2014年では70.9%、2015年では56.6%となり、両方の年で他のパターンに比べて低い結果となった。このことから、テキストの差分を用いることにより、より高い精度で予測できることが確認できた。

次に、深層学習を用いた予測で最も予測正解率の高かった、夕刊から朝刊の見出しとリードを用いたパターンの予測正解率と、従来用いられていたSVMでの予測正解率を比較したグラフを図2に示す。2014年から2015年を通した全体の予測正解率が、SVMを用いた予測では58.8%であったのに対し、深層学習を用いた予測の予測正解率は72.3%となった。また、2014年、2015年ともに深層学習の予測正解率がSVMの予測正解率を上回った。これらのことから、深層学習を用いることによりSVMよりも高い精度で予測できることが確認できた。

4 まとめ

本論文では、深層学習を用いた新聞記事分析による市場動向予測について述べた。日本経済新聞を対象として、夕刊から朝刊までを1つのテキストとし、その見出しとリードに対して時系列テキスト分析を行い、TOPIX ETFの取引価格が上昇するか下落するかを予測した。実験の結果、全体の予測正解率は72.3%となり、予測正解率が低い2015年においても予測正解率は59.8%であった。金融の分野では正解率が55%以上であれば有用であると考えられるため、提案手法は有用であるといえる。

参考文献

- [1] 松井藤五郎, 石田智也, 中嶋啓治, 和泉潔, 吉田稔, 中川祐志. 新聞記事を対象とした時系列テキスト分析による市場予測. 第7回人工知能学会ファイナンスにおける人工知能応用研究会 (SIG-FIN), pp. 15-19, 2011.
- [2] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. *EMNLP 2013*, pp. 1631-1642, 2013.
- [3] 松井藤五郎, 和泉潔. 新聞記事の時系列テキスト分析による市場動向予測. 第30回人工知能学会全国大会, 3L3-OS-16a-6, 2016.