

株価時系列画像を用いた株価予測の有効性検証

遠藤博人^{†1} 穴田一^{†1}

概要：近年、人工知能技術の発展が目覚ましく、様々な分野に応用されてきている。金融分野においても人工知能が応用されており、株価時系列データやニューステキストを用いた株価予測に関する研究が行われている。株価予測の研究では、株価をそのまま数値データを用いて学習を行う方法や株価時系列画像に変換してから学習を行う方法があるが、どちらのアプローチが有効なのかは検証されていない。そこで本研究では、株価時系列の画像データと数値データを用いて比較実験を行い、どちらが有効か検証した。その結果、株価を数値データのまま入力データとして学習するよりも時系列画像に変換して学習を行う方が高い精度で予測できることを確認した。

キーワード：株価時系列画像、株価予測、畳み込みニューラルネットワーク

Effectiveness Validation of Stock Price Prediction Using Stock Price Time Series Image

HIROTO ENDO^{†1} HAJIME ANADA^{†1}

1. はじめに

近年、人工知能技術の発展が目覚ましく、様々な分野に応用されてきている。金融分野においても人工知能が応用されており、株価時系列データ[1][2]やニューステキスト[3]を用いた株価予測に関する研究などが盛んに行われている。

宮崎らは、東証株価指数(TOPIX)と TOPIX 主要構成銘柄である Topix Core 30 の株価を入力とした畳み込みニューラルネットワーク(Convolutional Neural Network : CNN)を用いて株価の上昇下落を予測し、CNN による予測可能性を示した[1]。株価をそのまま数値データとして学習を行う利点としては、画像データよりも少ない計算量で済む点や余計なノイズが入らない点などが挙げられる。また池田は、日経平均株価のデータをローソク足チャート画像に変換し、それを CNN の入力として株価の上昇下落を予測し、約 60% の正解率を出している[2]。株価を画像データに変換してから学習を行う利点としては、実際に人間が行っているようなチャートの形状を目で見てトレンドを把握することで、今後の動向を予測するといった行動を学習できる。しかし、数値データでも同様の学習が行われているのではないかと考えられ、どちらのアプローチが有効か検証されていない。

本研究では、株価時系列の数値データと画像データを用いて比較実験を行い、どちらのアプローチが有効かを検証した。その結果、株価を数値データとして扱うよりも時系列画像に変換してから扱う方が高い精度で予測できることを確認した。

2. 比較実験

2.1 使用するデータ

本研究では、株価時系列の数値データと画像データを入力とした株価予測の比較実験を行い、どちらのアプローチが有効かを検証する。使用するデータは、日経平均株価の 2004 年 1 月 1 日~2017 年 12 月 31 日を学習期間、2018 年 1 月 1 日~2021 年 12 月 31 日をテスト期間として用いた。

数値データに関しては、終値を使用したデータとローソク足(始値、終値、高値、安値)を使用したデータの 2 種類を用意した(図 1)。

終値	26574	26427	25748	26213	27453
終値							
始値	26753	25918	25954	26045	27197
高値	26836	26479	26028	26290	27453
安値	26438	25904	25688	26003	27023
終値	26574	26427	25748	26213	27453

ローソク足

図 1 使用する数値データ

画像データに関しては、終値を表示した時系列画像とローソク足を表示した時系列画像の 2 種類を用意した(図 2)。

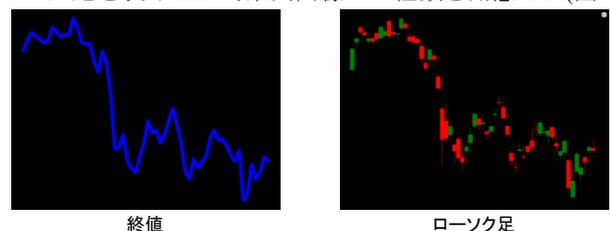


図 2 使用する画像データ

^{†1} 東京都市大学大学院 総合理工学研究科
Graduate School of Integrative Science and Engineering, Tokyo City University

予備実験の結果が良かったため、数値データの入力日数は 200 日、画像データの表示日数は 60 日、画像の入力サイズは 128 ピクセル×128 ピクセルとした。

2.2 使用する手法

比較実験では、人工知能技術を支えている深層学習の中でも画像データと時系列データの両方に強みを持つ CNN を用いた。実際に日足のデータを用いて投資判断を行う際、多くの場合は数日から数週間の短いスパンで取引を行うため、予測対象日は 5 営業日後と設定した。本実験では、図 1 及び図 2 のデータを CNN の入力として学習し、株価の上昇下落を予測した。

3. 実験結果

評価指標には、以下の表 1 のような混同行列から算出できる正解率(Accuracy)、再現率(Recall)、適合率(Precision)、F 値(F-measure)を用いた。

表 1 混同行列

	実際に上昇	実際に下落
上昇と予測	True Positive(TP)	False Positive(FP)
下落と予測	False Negative(FN)	True Negative(TN)

正解率は、全データに対してどれだけ正解できたかの割合を表す。再現率は、実際に株価が上昇したうち、どの程度予測できたかの割合を表す。適合率は、株価が上昇すると予測したうち実際に上昇した割合を表す。再現率と適合率はトレードオフな関係であるため、両者をバランス良く評価できる F 値も用いる。図 1 及び図 2 を CNN の入力とし、5 営業日後の株価の上昇下落を行った実験結果を以下の表 2,3 に示す。

表 2 終値を用いた際の実験結果

データ	正解率	再現率	適合率	F値
数値	49.09%	49.77%	54.87%	52.16%
画像	55.16%	61.61%	59.51%	60.54%

表 3 ローソク足を用いた際の実験結果

データ	正解率	再現率	適合率	F値
数値	50.65%	54.19%	56.01%	55.08%
画像	54.29%	59.62%	56.10%	57.81%

終値を用いた際の実験結果は、数値データの正解率が 49.09%で、画像データの正解率が 55.16%となり、画像データの方が高い結果を出すことが確認できた。ローソク足を用いた際の実験結果も、数値データの正解率が 50.65%で、画像データの正解率が 54.29%となり、同様の結果が確認できた。再現率や適合率、F 値といった指標についても、画像データの方が良い結果となった。また、終値を用いた画像データの方がローソク足を用いた画像データよりも僅か

に高い精度で予測できることが確認できた。

4. 考察

比較実験の結果から、数値データよりも画像データの方が高い精度で予測できることが確認できた。これは、株価時系列画像に表示されているチャートから上昇下落の特徴を形状で学習できたからだと考えられる。実際に人間が投資判断を行う際、自分の目でチャートに表示されている形状からトレンドを把握し、今後の動向を予測している。CNN は、人間の視覚野をモデルにした手法であるため、人間が見るデータ(時系列画像)を入力とした方が特徴を発見しやすくなり、高い精度で予測ができたと考えられる。

また、ローソク足用いた画像よりも終値を用いた画像の方が高い精度で予測できた。これは、終値を用いた画像の方が比較的シンプルな形状になっているため、CNN が特徴を発見しやすかったのではないかと考えられる。

5. おわりに

本研究では、CNN への入力データとして、株価時系列の数値データと画像データのどちらを用いることが株価予測に有効かを検証するため、比較実験を行った。その結果、画像データを用いる方が高い精度で予測できることを確認した。

今後の課題としては、予測精度の向上を図る必要がある。画像データの方が株価予測に有効であることは確認できたが、決して精度の高い予測ができたというわけではない。人間がチャートを分析する際は、移動平均線や出来高などの他の指標も参考にしているため、今後は株価時系列画像に表示する情報量を増やすなどを行い、予測精度の向上を目指していきたい。

さらに、CNN の構造解析を行う必要がある。画像データの方が高い精度で予測できる理由として、株価時系列画像に表示されているチャートから上昇下落の特徴を形状で学習できたからだと考察したが、数値データでも同じような学習をしている可能性がある。各データを入力とした CNN が何を学習したかを解析することで、画像データの有効性の源泉を確かめていきたい。

参考文献

- [1] 宮崎邦洋, 松尾豊. 深層学習を用いた株価予測. 第 31 回人工知能学会全国大会. 2017, 2D3-OS-19a-3.
- [2] 池田欽一. 株価ローソク足チャート画像を用いた畳み込みニューラルネットワークによる株価変動予測. 北九州市立大学『商経論集』. 2019, Vol. 54, No. 1,2,3,4. p. 1-18.
- [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株式市場の動向予測. 第 30 回人工知能学会全国大会. 2016, 3L3-OS-16a-6.