

## ローソク足のパターン分析を考慮した LSTM による経済指標予測 Economic Indicators Forecasting using LSTM considering Candlestick Pattern Analysis

仲岡 拓哉<sup>†</sup>  
Takuya Nakaoka

浦野 昌一<sup>†</sup>  
Shoichi Urano

### 1. はじめに

近年、国内外において経済情勢が目まぐるしく変化している中で、企業の今後の戦略立案の参考、個人の投資活動におけるリスク回避や投資対象選定への活用、更には国家の政策決定の参考に役立てていくためにも景気動向を掴む事は必要となってくると考えられる。景気動向を測るための指標としては、日経平均株価・GDP(国内総生産)・景気動向指数などが挙げられる。

これまで株価予測に機械学習の手法を提案している先行研究として、CNN-Autoencoder を用いたローソク足チャートからの価格変動予測<sup>[1]</sup>、ローソク足を複数組み合わせたローソク足チャートを使用した株価予測<sup>[2]</sup>などが報告されている。しかし、株価の上がり下がりやを考慮しつつ、回帰的に経済指標を予測する研究はあまりされていない。

これまで筆者等は、景気動向を測るためのテクニカル分析の一つとして経済指標を高精度に予測し、景気動向の特徴を捉えるための LSTM(Long Short Term Memory)を用いた経済指標の予測手法<sup>[3]</sup>や、エンコーダ・デコーダモデルを用いた短期の予測及び中長期の予測<sup>[4]</sup>を検討してきた。しかし、これまでは回帰的な予測のみで、株価の騰落についてなどの分類予測については考慮してこなかった。そこで、本稿では経営方針の参考と投資活動の参考に用いられる景気動向を測るためのテクニカル分析の一つとして、更なる予測精度の向上のために、これまでの経済指標予測に加えて上り下がりやのトレンドを加味する事を検討する。経済指標の上り下がりやのローソク足のパターン分析により捉えられると考えられる。これまでの数値予測に加え、株価のパターン分析も考慮した予測を行うことによって、トレンドの推移を踏まえた高精度な予測が可能となると期待される。

### 2. 経済指標予測

経済指標とは、各国の政府や中央銀行、関係省庁などが金利や景気など、その国の経済状況を反映する事柄を数値化した指標であり、現在の経済の状況や過去からの変化を把握するためには必要であると考えられている。将来の景気動向を分析する方法は大きく分類すると、景気動向や金融政策などの変化が市場に及ぼす影響を分析する「ファンダメンタル分析」と株価や通貨などの値動きの推移をグラフ化したチャートを読み、分析する「テクニカル分析」が挙げられ、近年では機械学習などの先端技術を用いた様々な研究も活発に行われている。本稿では、テクニカル分析に役立てるための経済指標予測の手法を検討する。

#### 2.1 日経平均株価

日経平均株価とは、日本の株式市場の中でも代表的な経済指標の一つであり、東京証券取引所一部に上場する約 2000 銘柄の株式の内、日本経済新聞社が 225 銘柄を選出して算出したものである。日経平均株価の動きを見ることに

<sup>†</sup> 明治大学 Meiji University

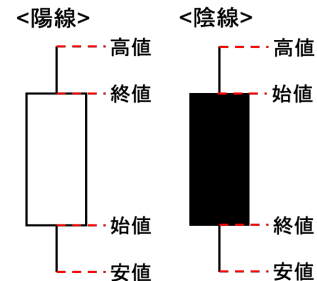


図 1 ローソク足の構造

よって、株式市場全体の大まかな値動きを把握できると言われている。また、日経平均株価には「終値」「始値」「高値」「安値」の 4 つの価格があり、四本値と総称される(以下、四本値と記載する)。本稿では、日経平均株価の翌日の上り下がりやの分類と終値を予測する。

### 3. ローソク足

ローソク足とは、日本で使われている代表的なチャートの内の一つであり、四本値を使用して図 1 のようにローソクの形に表したものである。ローソク足のパターンには何種類もあり、大きく分けると始値に対して終値が高い場合の「陽線」と始値に対して終値が低い場合の「陰線」に分けられる。

### 4. 予測手法

#### 4.1 LSTM

LSTM とは、再起型ニューラルネットワーク(RNN:Recurrent Neural Network)の一種であり、再帰的に時系列データを扱うことが出来る。LSTM は、短期依存性の学習が可能なおよび、図 2 に示すように LSTM 層の中のループ構造により過去の情報を保存することで、長期依存性の学習ができないという従来の RNN の問題点を改善した手法でもある。経済指標予測に対して LSTM を用いる事で、日経平均株価の時系列の特徴を捉え、より精度の高い予測を行える事が期待される。

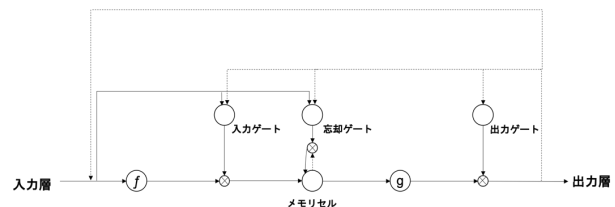


図 2 LSTM の構造例

#### 4.2 提案法

検討している提案法について説明する。以下が提案法の手順となる。

- (i). 複数のモデル化期間で行った分類予測において最も精度が高かった期間での回帰予測結果を選択する
- (ii). 分類予測を考慮した回帰予測として 1 月の予測結果、2 月の予測結果・・・12 月の予測結果といったように、それぞれの予測結果を組み合わせる新たな 1 年分の予測結果を出力する

本稿では、提案法に使われる分類予測と回帰予測をそれぞれ行ったので、その結果について示す。

## 5. シミュレーション

本稿では、日経平均株価の四本値に着目し、日経平均株価の四本値を用いて株価の上がり下りの分類及び日経平均株価の予測を行う。

### 5.1 シミュレーション条件

本稿では、公開されている日経平均株価の四本値を入力変数とする。今回は、1 日を午前 9:00 から午後 17:00 と設定する。始値は午前 9:00 の価格、終値は午後 17:00 の価格、安値はその日の中で最も安い価格、高値はその日の中で最も高い価格である。シミュレーションには、2010 年 1 月 4 日から 2019 年 12 月 30 日までの日経平均株価のデータを用いた。また、日経平均株価は平日のデータのみでもあることから 1 ヶ月を 20 日、1 年を 240 日と設定した。

#### 5.1.1 データセット

分類予測における教師信号は one-hot エンコーディングとした。one-hot エンコーディングとは、ベクトル成分のどれか一つだけが 1 でそれ以外を 0 と表現するものである。下記に one-hot エンコーディングの例を記載する。

「1」→(1, 0, 0) 「2」→(0, 1, 0) 「3」→(0, 0, 1)

また、本稿では分類予測において、始値より終値が上回った際には(1, 0)、始値より終値が下回った際には(0, 1)として、騰落の 2 値を予測する分類問題とした。

#### 5.1.2 予測における各パラメータ・予測期間について

入力変数は、日経平均株価の四本値の 4 つの日次データとする。以下に中間層のユニット数・学習回数・予測期間・活性化関数について示す。

LSTM 中間層：3 学習回数：1000 予測期間：60, 120 日

分類予測における活性化関数：sigmoid

回帰予測における活性化関数：linear

また、モデル化期間は以下の 5 パターンに設定した。

- (i). 2010 年 1 月 4 日から 2019 年 6 月 30 日
- (ii). 2015 年 1 月 4 日から 2019 年 6 月 30 日
- (iii). 2016 年 1 月 4 日から 2019 年 6 月 30 日
- (iv). 2017 年 1 月 4 日から 2019 年 6 月 30 日
- (v). 2018 年 1 月 4 日から 2019 年 6 月 30 日

### 5.2 シミュレーション結果

表 1 に予測期間を 60 日と 120 日とした際のそれぞれのモデル期間で分類予測した結果を示す。60 日予測では、モデル化期間を 2016 年 1 月 4 日からにした時、120 日予測では、

表 1 分類予測のシミュレーション結果

	2010~	2015~	2016~	2017~	2018~
60 日予測[%]	50.8	56.7	59.2	55.8	44.2
120 日予測[%]	53.3	58.3	60.0	63.3	56.6

表 2 終値の 60 日予測のシミュレーション結果

	2010~	2015~	2016~	2017~	2018~
平均誤差率[%]	0.6208	0.7909	0.9597	1.1334	1.9252
最大誤差率[%]	1.6079	3.0998	4.3160	3.6745	5.3704
最小誤差率[%]	0.0331	0.0082	0.0217	0.0103	0.0832

表 3 終値の 120 日予測のシミュレーション結果

	2010~	2015~	2016~	2017~	2018~
平均誤差率[%]	0.5473	0.7061	0.6533	1.0752	2.5668
最大誤差率[%]	1.9489	3.7468	2.4901	4.2207	16.378
最小誤差率[%]	0.0149	0.0026	0.0233	0.0243	0.0770

モデル化期間を 2017 年 1 月 4 日からにした時が最良の結果となった。モデル化期間が短い時、またはモデル化期間が長い時に予測精度が悪くなりやすい傾向にあることが分かる。これは、モデル化期間が短い時はデータ数が少なく日経平均株価の騰落の時系列の特徴を捉えられていないと考えられる。更に、モデル化期間が長い時はデータ数が多いことによって、始値と終値がほぼ同じ値である時のデータも多くなり、正しく分類できていない事が考えられる。

表 2 に予測期間を 60 日、表 3 に予測期間を 120 日とした際のそれぞれのモデル期間で日経平均株価の終値を予測した際のシミュレーション結果を示す。予測期間に関わらず、平均誤差率及び最大誤差率は 2010 年からモデル期間とした場合の精度が良い結果となった。回帰予測においては、より多くのデータを扱うことにより精度が良くなる傾向にあると考えられる。

## 6. まとめ

本稿では、日経平均株価の予測に対して、分類予測と回帰予測を行った結果、それぞれの予測においてモデル期間によって予測精度の違いがある事が確認できた。

今後は、提案法に示した分類の予測結果を踏まえた回帰予測を行なっていきたい。また、今回は最も簡素な 2 値分類を行ったが、今後は始値と終値の差が微小である場合についても考慮した上で経済指標予測を行っていきたく考えている。

### 参考文献

- [1] Siou Jihh Guo, Chih-Chieh Hung, Fu-Chun Hsu, "Deep Candlestick Predictor: A Framework Toward Forecasting the Price Movement from Camdlestick Charts", International Symposium on Parallel Architectures, Algorithms and Programming(2018).
- [2] Yoshihisa Udagawa, "Predicting Stock Price Trend Using Candlestick Chart Blending Technique", IEEE International Conference on Big Data(2018).
- [3] 仲岡 拓哉, 浦野 昌一, "LSTM を用いた経済指標予測", 電気学会 C 部門大会, G55-4(2019)
- [4] 仲岡 拓哉, 浦野 昌一, "中長期的な経済指標予測のための LSTM によるエンコーダ・デコーダモデルの検討", 人工知能学会全国大会, 113-GS-2-05(2020)