

機械学習によるテクニカル分析の影響の調査

Analysis of effect by technical analysis using machine learning

片寄 諒亮^{1*} 吉岡 真治^{1,2}
Katayose Ryosuke¹ Masaharu Yoshioka^{1,2}

¹ 北海道大学大学院情報科学院

¹ Graduate School of Information Science and Tehnology, Hokkaido University

² 北海道大学大学院情報科学研究院

² Faculty of Information Science and Tehnology, Hokkaido University

Abstract: In recent years, stock prices have been predicted in various forms such as technical analysis and fundamental analysis using time series data and financial indicators, and text analysis using news information. In particular, some use text mining to predict whether stock prices will rise or fall from text information, but useful text information does not appear frequently on all stocks. Considering the increase in algorithmic trading using technical analysis, analysis that relies solely on textual information is not appropriate because it does not take into account the impact of such trading. Therefore, in this study, first, stocks that are easily affected by technical analysis were ranked by using machine learning to raise and lower stock prices using indicators that are often used in technical analysis techniques. Moreover, when the analysis did not go well, we analyzed what kind of events occurred and investigated how technical analysis affects the stock price.

1 はじめに

1.1 背景

近年、時系列データや金融指標を用いたテクニカル分析やファンダメンタル分析、さらにニュース情報を用いたテキスト分析など様々な形で株価の予測が行われている。特にテキストマイニングを使ってテキスト情報から株価の騰落を予測するものもあるが、全ての銘柄に対して頻繁に有用なテキスト情報が出るわけではない。また、テクニカル分析などを用いたアルゴリズム取引が増えていることを考慮すると、テキスト情報だけに頼った分析では、このような取引の影響が考慮されないため、適切な分析にならないことが想定される。そこで本研究では、まずテクニカル分析の手法で良く用いる指標を複数用いて、株価の騰落を機械学習によって行うことでテクニカル分析の影響を受けやすい(予想があたりやすい)株式のランク付けや、その分析がうまく行かないときに、どのような事象が発生しているのかを分析することで、テクニカル分析が株価にどのような影響を与えるかについての調査を行なった。

1.2 株価の変動要因

現在株価の変動要因と考えられているものは大きく4つある。

- ファンダメンタルズ (国や企業などの経済活動の状況を表すもの)
- テクニカル (株価の変動パターンのような過去の時系列的動きから作成されるもの)
- センチメント (相場状況や様々な情報に対する投資家心理)
- アノマリー (理論で説明できない規則性)

主にこれらの4つの変動要因により株価は動いていると考えられる。この変動要因を元に株価予測を行なう研究が多く行なわれているが、背景でも述べたように現状ではニュースの分析からの株価予測 [1] や、センチメントに対する分析 [2] など、その1つの要因からの分析によって株価を予測するものが多い。しかし背景でも述べたようにテキストマイニングを元にした予想を例にすると実際にはある銘柄に対して、毎日投資家に影響を及ぼすようなニュースが出ることは少ない。

実際の株価の変動要因は上で挙げたが、参照とするデータが数値データで表せるのかどうかという観点か

*連絡先：北海道大学大学院情報科学院
〒 060-0814 北海道札幌市北区北 14 条西 9 丁目
lk@eis.hokudai.ac.jp

ら変動要因を大きく2つに分けた [3]。数値データ以外のデータというのは全ての銘柄に対して、毎日有用であるようなものではないと考えられることから、数値データを主要因、それ以外を付属要因とした。イメージとしては上の変動要因としたファンダメンタルズが、全体の景気動向を元にした変動と、個別銘柄に関する基本的情報の2種類に分かれているようなものになる。

- 主要因 (数値データを元としたもの)
 1. 企業や国全体の景気動向
 2. テクニカル分析
- 付属要因 (数値データ以外を元としたもの)
 1. 個別銘柄に関する基本的情報 (業績の情報や不祥事などのニュース)
 2. センチメントに関するもの (情報に対しての投資家の反応)
 3. アノマリー (理論で説明できない規則性)

実際にこのように数値データとそれ以外に分けた研究 [4] は過去にされており、この研究では主に数値データの方の要因に加えてニュースで報道される社会的イベントを考慮した株価予測を行なっている。

本研究では、この過去の研究の内容から株価予測について、次のような新たな仮説を考えた。

仮説 株価変動の主要因としたものの影響を調べることで、今回付属要因として位置付けたものの要因が実際の株価変動要因になっている部分を調べることができるのではないかと考えた。

詳しく説明すると、企業や国全体の景気動向、そしてテクニカル分析から株価を予想した際にある程度の予想ができるモデルができたとする、それが適用できない範囲、すなわち予測が外れる時期であったり、条件などを洗い出すことができれば、データ量が少ない業績に関するニュースも、研究が多く行われているテキストマイニングによる株価予測もこれに適用しやすくなるのではないかと考えた。

その上で、以下のような検証を行うことにより、本仮説の有用性が検討できると考えた。

1. 企業や国全体の景気動向は個別銘柄に対して影響を与えているのかを、日経平均との相関性を調べることにより検証する。
2. テクニカル分析においても有名な手法をモデル化することによって、それがうまく当てはまるのかを検証する。

本論文では、1と2についての検証方法とそれを行なった結果と考察を述べる。

2 テクニカル分析モデル

2.1 作成したモデルについて

今回提案する手法はシステムトレードにおいてよく使われているテクニカル分析の特徴量を用いて株価の予測を行うモデルを作成した。今回使用したデータは2018年時点で日経225に入っている225銘柄と日経225の株価と連動して動く日経225連動型上場投資信託の合わせて226銘柄についての2009年から2018年の10年間の株価データである。2009年以降に上場した企業についてはこれよりもデータが少ないこともある。これらの数値データをもとに様々なテクニカル指標を算出した。また作成したモデルは226銘柄すべての2009年以降のデータから2017年までのデータをトレーニングデータとして使用して作成し、これを2018年のデータをテストデータとして株価の予測を行なった。今回、用いたテクニカル分析の特徴量として、変動率に加えて、システムトレードにおいてよく使われているテクニカル指標の中から、移動平均線乖離率、RSI、MACD、ATR、モメンタム、パラボリック、DMI、ストキャスティクスを使用した。またこれらの指標から15個の特徴量を作成し、翌日の変動率が上がるか下がるかを予測した。この特徴量には次に説明するテクニカル指標によって判断したシグナルの有無を入れている。また、Pythonの機械学習ライブラリであるscikit-learnを使用し、機械学習の一種であるrandomforestを使用してモデルの作成を行い予測を行なっている。

2.2 テクニカル指標

今回有名なテクニカル分析をモデル化するためにあたって、機械学習に説明変数として用いるテクニカル分析の手法を説明する。移動平均線乖離率、RSI、MACD、パラボリック、DMI、ストキャスティクスについてはマネックス証券のテクニカル指標一覧 [5] にもあるがいくつかを説明する。また [5] では紹介されていないATR、モメンタムについての説明も行う。

ATR(Average True Range) ATRとはTRの移動平均をとったもので、TRは当日の高値から当日の安値を引いたもの、当日高値から当日安値を引いたもの、当日高値から前日終値を引いたものそれぞれ以下の式で表すことができる。

$$ATR = Avg(TR)$$

$$TR = Max \begin{pmatrix} T_{high} - T_{low} \\ T_{high} - Y_{close} \\ Y_{close} - T_{low} \end{pmatrix}$$

ただし、 T_{high} : 当日高値、 T_{low} : 当日安値、 Y_{close} : 前日終値これによって株価の動く大きさを、すなわちボラティリティが高いか低いかを表すことができる。つまり ATR は株価の上昇、下降それぞれの局面における勢いの強さを判断することができる指標であり、ATR が上昇している時には株価のトレンドが持続する可能性が高く、ATR が上昇から下降に転ずる時には株価のトレンドが弱まり、反転する可能性があることを示している。

DMI(Directional Movement Index) この指標の特徴は相場の強弱を可視化し、価格の変動幅(ボラティリティ)からトレンド分析を行える部分にあります。DMIは通常、3つの指標(DI^+ , DI^- , ADX)から総合的に売買の判断をします。

$$DM^+ = T_{high} - Y_{high}, \quad DM^- = Y_{low} - T_{low}$$

ただしそれぞれ負の値になる場合や DM^+ であれば DM^- を下回った場合、 DM^+ は0となる。逆も同様である。この値により、3つの指標が以下のように導かれる。また ADX が上昇している際に DM^+ が DM^- を上に突き抜けた場合に買いシグナル、逆の場合に売りシグナルとなる。

$$DI^+ = \frac{\sum^n DM^+}{\sum^n TR} \times 100, \quad DI^- = \frac{\sum^n DM^-}{\sum^n TR} \times 100$$

$$ADX = \frac{1}{n} \sum^n \frac{|DI^+ - DI^-|}{DI^+ + DI^-}$$

モメンタム (Momentum) 10日間のモメンタムは、当日の終値から10日前の終値を引くだけで求めることができ、これによって相場の短期的な動きを判断する指標である。すなわち、相場が上昇している時の勢いが弱くなってきているのか、また相場が下降している時の勢いが強くなってきているのかを捉えることができる。

移動平均線 (Moving Average) 移動平均線とは一定期間の平均終値を計算し、それぞれの結果を線をつないだものである。例えば、5日移動平均値は5日間の平均価格であり日々の移動平均値を線をつないだものが移動平均線となる。

これを用いて、テクニカル分析でよく行なわれているのは、ゴールデンクロスとデッドクロスを利用したものである。ゴールデンクロスとは短期の移動平均線が長期の移動平均線を下から上へ追い越した場面のことを言う。デッドクロスはその逆である。一般的にはゴールデンクロスが起こった際には今後の株価は上がっていくことが予想され

る。デッドクロスはその逆で起これば、それ以降の株価は下がっていくことが予想される。

RSI(Relative Strength Index) 相場において買われすぎか売られすぎかを値上がり率の比率から求める有名な指標である。一般的には30%を下回ると売られ過ぎていると言うことで株価は好転することが多く、逆に70%を越えると買われすぎだと判断する。

$$RSI = \frac{Avg_{sum}}{Avg_{gain} + Avg_{loss}} \times 100$$

ただし、

Avg_{sum} : 一定期間の値上がり幅の平均、

Avg_{gain} : 一定期間の値上がり率の平均、

Avg_{loss} : 一定期間の値下がり率の平均

MACD(Moving Average Convergence Divergence) 短期の移動平均線と長期の移動平均線の差をMACDと言う。このMACDの移動平均線をとったものがシグナルと呼ばれ、MACDがシグナルを上突き抜けたら「買い」と判断する。すなわち今後の株価は上がっていくと予想できる。一方、MACDがシグナルを下に突き抜けたら「売り」と判断し、今後の株価は下がっていくと予想する。移動平均線のゴールデンクロスやデッドクロスによるシグナルに先行する形で、MACDの売買シグナルが出現する特徴がある。

特にここで短期の移動平均線と長期の移動平均線の差を乖離といい、その割合をとったものを乖離率という。

3 調査結果と考察

3.1 主要指標との連動性

日経225の変動率それぞれの個別銘柄の変動率データを使用し、相関係数を計算した。ここでの変動率は前日の終値と当日の終値の差の割合を表している。実際には株式の分割や併合を行なった銘柄については相関係数が極端に低くなってしまったため、それに該当する51銘柄を除いた174銘柄についての検証を行なった結果を表1,2に示す。

相関関係の高い銘柄に関してはあまりニュースにもならず手堅い企業のイメージがある企業が名を連ねた。一方で相関関係が低い銘柄に関してはデータ改ざんにより株価を大きく下げた時期があった神戸製鉄所や比較的新しい事業のニュースが多く出るIT企業が入った。

この2つの対象の銘柄の中でも最も日経225との変動率に差があったのは1333(マルハニチロ)の2月6日

表 1: 日経 225 との相関関係が高い 5 銘柄

銘柄名	相関係数
8015(豊田通商)	0.75100
6367(ダイキン工業)	0.74736
7267(ホンダ)	0.71834
5333(日本ガイシ)	0.71530
6503(三菱電機)	0.71146

表 2: 日経 225 との相関関係が低い 5 銘柄

銘柄名	相関係数
1333(マルハニチロ)	0.30097
9503(関西電力)	0.34878
5406(神戸製鋼所)	0.37308
4689(Z ホールディングス ¹)	0.37921
2432(ディーエヌエー)	0.38702

であった。図 1 についてはもっとも変動率の差が大きかった 2 月 6 日を含んだ 2 月における日経 225、豊田通商、マルハニチロの 3 銘柄の株価変動率を表したものである。この日付について調査したところ、前日の 2 月 5 日に決算予想の発表があり、予想業績を上方修正したことにより個別銘柄の株価は上昇したことが考えられる。一方で、日経平均は同じく前日にアメリカの株価が過去最大の下げ幅を記録したことから、その影響を受け、大幅に値を下げたことが考えられる。

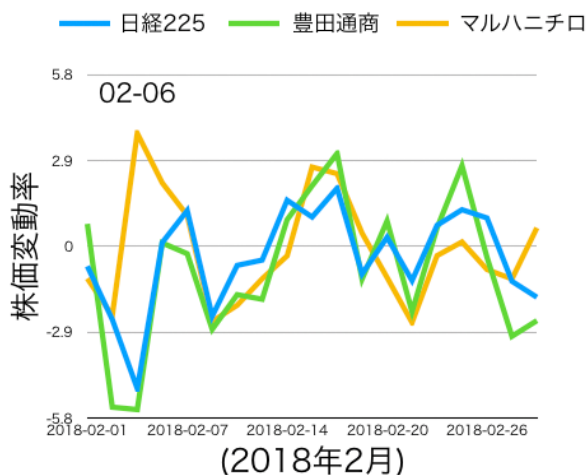


図 1: 相関係数を元にした 2018 年の株価変動率の推移 (2 月)

この結果からもわかるように独自の情報、すなわち決算であったり不祥事のニュースがある場合には連動性が下がる可能性があることが確認できた。このように独自情報がある日付については、センチメント解析

や個別銘柄に関するニュースについてのテキストマイニングなど、他の情報から予測を行う必要がある。

しかし、このような独自の情報は決して多くはないため、これらによる変動が起きた時期を見つけることは、個別要因の影響を分析する上で重要な事例となる。

3.2 テクニカル分析

今回の調査では日経 225 において使われている銘柄についてすべての銘柄について、2 日前と前日の離散化した株価変動（上昇、下降）と 2.2 節で述べた指標に基づく売り買いのシグナルが発生したかどうか（買いシグナル、シグナルなし、売りシグナル）をフィーチャーとして、当日の株価の変動（上昇、下降）を当てるとする二値分類の問題としてモデル化し、機械学習による予測を行なった。具体的には、scikit-learn[6] のランダムフォレストを用い、訓練用のデータとしては、2009 年から 2017 年の株価に基づいたデータを利用し、2018 年の株価の変動を当てるとするタスクを実施した。ただし、日経 225 の銘柄にも変動があるため、トレーニングデータについては、日経 225 に登録された期間のものに限定することとした。この結果、精度の良かった 5 銘柄と悪かった 5 銘柄を表 3,4 に示す。た

表 3: 機械学習に基づく予測上位 5 銘柄

銘柄名	予測精度 (%)
6773(パイオニア)	62.832
6752(パナソニック)	61.504
7004(日立造船)	61.062
4502(武田薬品工業)	59.735
4689(Z ホールディングス)	59.292

表 4: 機械学習に基づく予測下位 5 銘柄

銘柄名	予測精度 (%)
8028(ファミリーマート)	37.168
9735(セコム)	42.035
4506(大日本住友製薬)	43.363
4902(コニカミノルタ)	43.805
9613(NTT データ)	44.248

だし、全体の結果としては 51.750% となり、使用したテクニカル指標の良し悪しや予想する時期を長期でみるのか短期でみるのかなどについて、再検討が必要であると考えている。また予測精度が低かった、8028(ファミリーマート)と 9735(セコム)、4506(大日本住友製薬)

¹旧ヤフー

については2018年1年間での株価の上昇率が高く、急激な上昇もいくつかみられる。このような銘柄は、業績その他の材料が多い銘柄だと考えられる。これに対し、7004(日立造船)、6752(パナソニック)、4502(武田薬品工業)は、全体的に株価の下落率が高い銘柄になっている。

4 まとめと今後の課題

実験を行う前の予想としてはシステムトレードを行う人が少なからずいることからテクニカル分析によるシグナルが出た後には、ある程度の割合でシグナルなどに沿って株価が変動するものだと考えていた。現段階ではそれをうまく示すことができる有用なモデルはできなかったものの、各銘柄においてそれぞれの特徴があることや一概にテクニカルな分析が全てにおいて有用ではないということが議論できた。

今回のモデルの予測があまり良くなかった要因として考えられるものは以下に示されるように2018年以外の株価変動と2018年の株価の変動率が違うということが考えられる。以下の表5は今回用いた全銘柄の変動回数を表している。あるいは急激な値下がりについてはリーマンショックなどでデータがそれなりに多く存在するのに対し、急激な値上がりというのは、あまりデータがないことも一因だと考えられる。

表 5: 全 226 銘柄の株価変動回数

	上昇	下落
2018年以外の株価変動	253681回 (52.03%)	233932回 (47.97%)
2018年の株価変動	24356回 (47.69%)	26720回 (52.31%)

今後の課題としては予測に使う手法や特徴量の入れ方を見直すことや使うテクニカル指標の種類を変えることで、今回株価変動の主要因として考えてきた数値データによる予測の精度を上げることが一番に考えられる。特に近年LSTM[7]やCNN[8]を予測に用いる手法が成果を出していることから今回作成したテクニカル指標のデータを用いることで予測精度の改善に繋げていきたい。これができる上で、これらの予測をさらに補正していけるように、今回は付随要因として考えた、数多くは出ない企業の決算情報であったり、ニュースのテキスト記事などを加えて行くことにつなげていくことが目標となる。

5 参考文献

- [1] 敷地琢也, 和泉潔. 複数のテキスト情報を用いた株式市場動向の分析. 人工知能学会研究会資料 (第13回金融情報学研究会).SIG-FIN-013-09. (2014).
- [2] 河合継, 新田翔, 大川堯郁, 西山昇. 感情によるマルチモーダルAIを利用したIPO株価推定. 人工知能学会研究会資料 (第20回金融情報学研究会).SIG-FIN-020-03. (2018).
- [3] 三菱UFJトラスト投資工学研究所 [編]. 実践金融データサイエンス. 日本経済新聞出版社. (2018).
- [4] 中菅章浩, 関和広, 上原邦昭. ニュース記事で報道される社会的イベントを考慮した株価動向予測の補正. 人工知能学会研究会資料 (第10回金融情報学研究会).SIG-FIN-010-03. (2013).
- [5] マネックス証券 テクニカル指標一覧
<https://info.monex.co.jp/technical-analysis/indicators/>
- [6]scikit-learn
<https://scikit-learn.org/>
- [7]David M.Q. Nelson,Adriano C.M.Pereira,Renato A. de Oliveira.Stock market's price movement prediction with LSTM neural networks.International Joint Conference on Neural Networks (IJCNN). (2017)
- [8] 宮崎邦洋, 松尾豊. 深層学習を用いた株価予測の分析. 人工知能学会全国大会論文集 (第31回人工知能学会全国大会).2D3-OS-19a-3. (2017)