

2024 年度 卒業研究論文

テクニカル指標を用いた株価予測精度の検証

指導教員

五島洋行 教授

法政大学

経営システム工学科

21X4012 大山明莉

学科名	経営システム工 学科	学籍番号	21X4012
申請者名		大山明莉	
指導教員		五島洋行	

論文要旨

論文題目	テクニカル指標を用いた株価 予測精度の検証
------	--------------------------

本研究では、機械学習を活用し、複数のテクニカル指標を組み合わせた株価予測手法の有効性を検証する。テクニカル指標は、個人、機関投資家、トレーダー、ブローカー、ディーラー、その他市場参加者の大多数にとって、投資行動から最大のリターンを求め、取引から最大の利益を得る機会を上げるべく株価予測に広く用いられているが、複数の指標を効果的に組み合わせる手法に関する体系的な研究は少ない。そこで本研究は、特にテクニカル指標の組み合わせ選択に着目し、ランダムフォレストと遺伝的アルゴリズム (GA) を用いて最適な特徴量の選択を試みる。

実験の結果、単一のテクニカル指標を用いた場合に比べ、指標の成り立ちや原理を基に適切に組み合わせた指標群が株価予測の精度を大幅に向上させることが示された。特に、単純移動平均 (SMA) と MACD の組み合わせは、他の組み合わせに比べて最良の精度を示した。一方で、遺伝的アルゴリズムによる特徴選択では、相互に高い相関を持つ指標や不要な指標が選ばれることがあり、予測精度の向上に寄与しない結果であった。これにより、指標間の整合性や情報の冗長性を考慮した特徴選択の重要性が明らかになった。

目次

第1章 はじめに

1.1 研究背景と目的

1.2 論文構成

第2章 先行研究

2.1 テクニカル分析のパフォーマンス考察

2.2 テクニカル分析を用いた株価予測の研究

2.3 深層学習と株価予測

第3章 関連知識

3.1 機械学習

3.2 ランダムフォレスト

3.3 テクニカル分析

3.4 貪欲法(グリーティー法)

3.5 遺伝的アルゴリズム(GA)

第4章 実験概要

4.1 使用データ

4.2 評価関数

4.3 貪欲法による最適化

4.4 GA による最適化

4.5 実験環境

第5章 単一指標を用いた株価予測

5.1 予測結果

第6章 複数のテクニカル指標の組み合わせ最適化

6.1 貪欲法による予測結果

6.2 GA による予測結果

第7章 考察

7.1 結果における考察

7.2 今後の課題

第8章 おわりに

参考文献

第1章 はじめに

1.1 背景と目的

現在の日本経済は少子高齢化問題や低金利環境などの様々な不安要因や、政府が「NISA（少額投資非金融制度）」や「iDeCo（個人型確定拠出年金）」などの制度を導入したことにより、貯蓄から投資へという資産の流れが到来してきている。株式市場において、投資家が効果的な意思決定を行うためには迅速かつ正確な将来の株価動向予測が最も重要である。しかし、株価は経済状況、国内外の政治情勢、国際的な経済動向など、多くの複雑な社会的要因によって影響を受けるため、正確な予測は困難であるとされている[1]。その中で、古くから多くの投資家は株価の動向を予測するために、時系列データや金融指標を用いたテクニカル分析や企業の決算書の数値やニュース情報を活用するファンダメンタル分析といった手法を活用している。特にテクニカル分析は、経済的な専門知識を必要とせず、視覚的な観点から分析を行うため初心者でも始めやすい手法とされており、実務的にも学術的にも有効であるとされている。一方で、多くの投資家は「テクニカル分析＝複数のテクニカル指標を組み合わせて行うもの」ということを当たり前のように解釈し、指標の原理原則を十分に理解せず、書籍やインターネット、情報商材などで得た知識、ノウハウに基づいて有名なテクニカル指標を組み合わせて利用して組み合わせているのが現状である[2]。そこで本研究では実際によく用いられているテクニカル指標を組み合わせることでその整合性や合理性を調査することを目的とする。

時を同じくして、近年ではコンピューター技術の発展に伴い大容量で高速なデータ処理が可能になったことから、市場のデータをそれらの技術を利用しデータ分析に役立つアルゴリズムの開発などがされている。中でも金融分野における機械学習の活用は国内問わず広く研究されている。また、各金融機関にとっても機械学習を金融商品価格や貸倒リスクなどの予測精度向上、レポート作成や不正検知による業務効率化、AIによる顧客サポートなど幅広い分野にて活用し、競争力を確保することが重要なテーマとなっている[3]。

本研究では、ソフトバンク（株）の株価データを対象に、「現在の価格変動パターンが過去のある時点と似ていれば、そのパターンに基づいて将来の株価予測ができる」という仮説を基本とし、幾何学的な価格変動パターンを機械的に抽出し株価予測へ活用することでテクニカル指標単体、または組み合わせが予測精度に与える影響を明らかにすることを目的とした将来の株価予測を行う。さらに、複数のテクニカル指標を組み合わせることで、株価変化率の予測精度向上を目指し、その有効性を体系的に検証する。

1.2 本論文の構成

本論文は、全 8 章で構成している。

第 2 章では先行研究を述べる。

第 3 章では本研究で用いる関連研究について述べる。

第 4 章では本研究の実験の概要及び使用するデータと最適化問題としての定式化、実行環境について述べる。

第 5 章では単一指標を用いた予測結果を述べる。

第 6 章では複数指標や特徴量の組み合わせによる予測結果を述べる。

第 7 章では第 5 章と第 6 章の結果を用いた考察を述べる。

第 8 章では本研究のまとめを述べる。

第2章 先行研究

2.1 テクニカル分析のパフォーマンス考察

文献[1]では、テクニカル分析が投資パフォーマンスの向上にどの程度役立つのかを目的として、ストキャスティクス、RSI、MACD、75日移動平均、ゴールデン・デットクロス、一目均衡表の6つの指標を用いて分析が行われた。具体的には、過去2年間の市場データをもとに買いと空売り（信用取引）による運用成果が120%以上となる銘柄群着目し、単一指標ごとのパフォーマンス性を評価している。研究の流れは、まずテクニカル指標を用い、過去の市場データをもとにシミュレーションを実施し、その結果をもとに実際に市場の状況でこれらの指標が利益を生むか統計的に分析する。また、個人投資家へのアンケートや市場データを利用した実証分析も行われた。

検証の結果、運用パフォーマンスは企業によって異なり、市場がトレンドが明確な局面やボラティリティが高い相場など特定の条件下では短期的な利益が見込めるが、長期的な効果は限定的であることが示された。また、主観的判断が多くなるため、一貫した成果を上げるのは難しい点や過剰な分析によるだましの可能性も指摘された。これを踏まえ、伊藤博志氏は多くの人々にとってこうしたテクニカル分析を軸とした株式市場の客観的な考察がもっと盛んにおこなわれるべきであり、自らの投資手法をもった賢明な個人投資家層の拡大が株価市場の発展となると主張した。

そこで本研究では、主観的判断を機械学習によって排除し、複数のテクニカル指標を組み合わせることでテクニカル分析の有効性を検証する。

2.2 テクニカル分析を用いた株価予測の研究

類似の研究として、文献[4]ではテクニカル分析の手法でよ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株式のランク付けや、その分析がうまくい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のかについて調査を行っている。代表的なテクニカル指標とは移動平均乖離率、RSI、MACD、パラボリック、DMI、ストキャスティクス、ATR、モメンタムを使用し、それらの指標から計算される15個の特徴量を作成し、翌日の株価変動率の騰落を2値分類問題として予測する実験を行っている。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ランダムフォレストを使用してモデルの作成を行い、予測を行なっている。この際、テクニカル指標の組み合わせは考慮されておらず、テクニカル指標から算出した15個の特徴量すべてを用いた際の予測精度は51.750%と

なり株価の騰落をうまく表すことが出来る有用なモデルはできなかつたとされている。

そこで本研究では、テクニカル指標との相性がいい個別銘柄を分析したうえで複数のテクニカル指標や特徴量の組み合わせを行い、さらなる予測精度の向上を試みる。

2.3 深層学習と株価予測

株価予測において機械学習技術の活用が広く行われている。特に、高い特徴発見能力を持つ Convolutional Neural Network(CNN)が株価予測に適していると考えられている。文献 [5]では、株価データのみを用いて 9 種類のモデルを作成し、日経平均株価の騰落を 2 値分類問題として予測し、それぞれのモデルの予測精度を比較する実験を行っている。比較するモデルの種類は以下の 9 種類である。

- Rand-1: 学習データ 1 年間における日経平均の騰落率でテストデータの騰落を予測
- Rand-2: 1/2 の確率で騰落を予測
- LogisticRegression: ロジスティック回帰
- RandomForest: ランダムフォレスト
- MLP: 多層パーセプトロン (Dense4 層)
- LSTM: LSTM2 層+Dense2 層
- CNN: (Conv+Maxpooling)2 層+Dense2 層
- PCA-CNN: 前処理 PCA+ (Conv+Maxpooling)2 層+Dense2 層
- CNN-LSTM: (Conv1D+Maxpooling1D)1 層+LSTM2 層+Dense1 層

ニューラルネットワークは 4 層に統一し、それぞれ活性化関数は ReLU を使っており、0.2 の割合で各層で Dropout を組み込み、学習回数は 100 epoch としている。結果としては、LSTM が最も予測結果が高く、次にランダムフォレストが高かった。一定以上の Accuracy スコアがあることから、長期的なトレンドの予測を行うことを可能にした。

この研究ではファンダメンタルズやテクニカル指標、センチメント指標を含めずに純粋な株価データのみを使用し、モデルの性能を比較検証している。そこで本研究では、この研究で予測精度の高かったランダムフォレストモデルを使用しテクニカル指標を考慮した株価予測を行う。

第3章 関連知識

3.1 機械学習

機械学習は、過去のデータを使ってモデルをトレーニングし、未来の予測を行うことができる。株価データのような経済指標、企業業績、投資家心理など多数の要因に影響される複雑で非線形的なデータでも従来のモデルでは非線形性や複雑な相関関係を解析し、予測ができるため本研究では株価の予測に機械学習を活用する。本節では、機械学習の概要について説明する。

3.1.1 機械学習とは

機械学習 (Machine Learning) とは、コンピューターに膨大なデータを読み込ませ、データ内に潜むパターンを反復的に学習させることでデータの中に潜む特徴や規則性を見つけ、未知 (将来) のデータを判断するためのルールを獲得することを可能にするデータ解析技術である。機械学習による「学習」とは、データから意味のある規則や分類のための基準を取り出し、データが生成されるメカニズムやデータが示している個体の特徴をアルゴリズムとしてとらえることを意味する。機械学習の活用は、これまで人々の勘や経験に基づいて判断されてきた事象などを定量的に判断することを可能とする[8]。

3.1.2 機械学習の方法

機械学習の手法は「教師あり学習」と「教師なし学習」、「強化学習」の3つに分けることができる。

① 教師あり学習 (Supervised Learning)

教師あり学習とは、訓練データによる「学習」とテストデータによる「予測」の2段階に分けて機械学習を行う手法である。あらかじめ正解ラベルが付与されている訓練データから、データの特徴やパターン、ルールを学習させ、モデルとして出力する。次に、正解がわからないテストデータと訓練データで得た予測モデルを適応させ予測を行う。コンピューターが自動で作成し、識別力を向上させていくことで入力されたデータに対して正解となるデータの出力をすることが出来る。代表的な活用事例として、「回帰」や「分類」が挙げられる。

② 教師なし学習 (Unsupervised Learning)

教師なし学習とは、正解となるデータは存在せず、入力されたデータを利用して基本構造や分布をモデル化する手法である。大量のデータを読み込むことにとってコンピューター自身がデータの持つ特徴を導き出し、結果を出力する。代表的な活用事例として、「クラスタリング」

や「異常検知」が挙げられる。

③ 強化学習 (Reinforcement Learning)

強化学習とは、最初からデータがあるわけではなく、コンピューター自身が試行錯誤しながら、精度を上げていくための学習方法であり、正解を与える代わりに将来の価値を最大化することを学習する。

本研究では①教師あり学習を用いて機械学習を行う。

3.1.3 予測モデル

機械学習の代表的なモデルは以下の図1に表される。

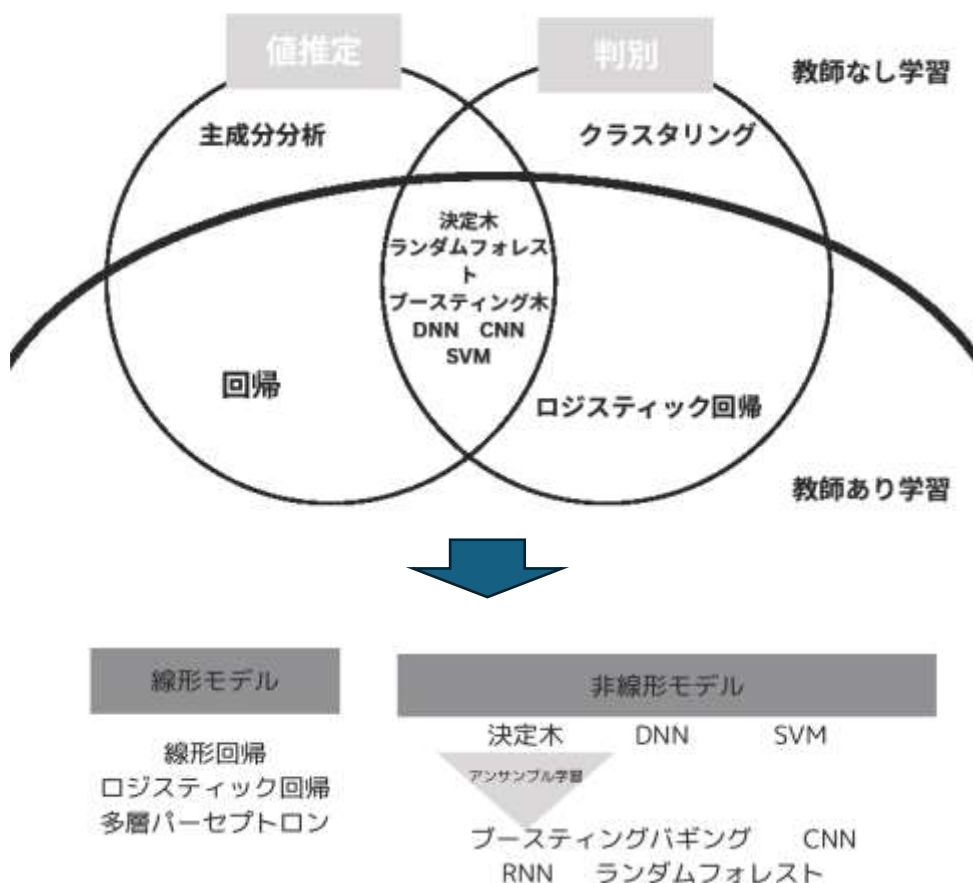


図1 機械学習のモデルの分類

3.2 ランダムフォレスト

ランダムフォレスト (Random Forest) は、機械学習におけるアンサンブル学習手法の 1 つで、複数の決定木 (Decision Tree) を組み合わせて予測を行うモデルである。機械学習における分類問題と回帰問題の両方に使用でき、高い精度と汎化性能を持っている。また、ランダムフォレストはモデルが学習において各特徴量をどの程度利用したかを計測する仕組みが組み込まれている。特に MDI (決定木の分割基準として使用された際の不純度の減少量をもとに重要度を算出) や MDA (特徴量をシャッフルして精度の低下量を評価) といった指標を活用することで、特徴量の重要度や寄与度を簡単に可視化・解釈することが可能である [9]。そのため、本研究ではランダムフォレストモデルを用いて株価予測を行い、テクニカル指標の影響調査を行う。

3.2.1 決定木

決定木とは、「予測」、「判別」、「分類」といった目的で用いられる手法である。Yes/No など二者択一の質問を階層構造につなげて 1 つずつ答えていくことでいくつかのグループに分けていき、最終的に目的となる答えにたどり着けるといった仕組みを持つ。この階層構造が枝の木のように見えることから「決定木」という名前が付いたとされている。ランダムフォレストでは多数の決定木を集めて、組み合わせることで精度を高めている。

3.2.2 アンサンブル学習

アンサンブル学習は、より良い予測を得るために複数の異なる学習アルゴリズムを組み合わせる技術である。「分類」の場合は各木が出力するクラスの多数決で決まり、「回帰」では各木の出力値の平均を取るのが一般的である。アンサンブル学習においてよく用いられる手法には「バギング」と「ブースティング」の 2 種類がある。

① バギング

バギングとは、データの分散を減らしてモデルの過学習を防ぐことを目的とした手法である。この方法では、トレーニングデータセットの中からランダムにサンプリングして、複数のデータセットを作りそれぞれ別々にモデルを訓練させ、最後に予測結果を統合させる方法である。バギングでは、複数のモデルが独立して動作するため、並列計算ができ計算効率が良いという利点がある。バギングの代表例として、先述したランダムフォレストが挙げられる。

② ブースティング

ブースティングとは、モデルのバイアスを減らして予測精度を高めることを手法とした方法である。ブースティングは、最初のモデルを訓練させうまく予測が出来なかったデータに重みをつけ、重みをつけたデータを使って 2 つ目のモデルを訓練させる。このプロセスを繰り返

し、誤差を徐々に修正するようにモデルを強化していく方法である。弱学習木を繰り返し改善することで非常に精度の高い結果が得られる可能性があることが利点である。ブースティングの代表例として、AdaBoost, Gradient Boosting, XGBoost, LightGBMなどが挙げられる。

3.3 テクニカル分析

3.2.1 テクニカル分析とファンダメンタル分析

相場の分析には主に「テクニカル分析」と「ファンダメンタル分析」という2つの分析手法がある。

① テクニカル分析

「テクニカル分析」とは、投資商品の価格、出来高、健玉の水準といった数値を積み上げ、過去の市場の動きを示したチャートを分析することによって、価格の値動きや将来の市場の方向性を予測する手法である。テクニカル分析におけるメリットとデメリットは表1にまとめる。

表1 テクニカル分析のメリットとデメリット

メリット	デメリット
テクニカル分析は、ほとんどすべての市場で広範な投資商品を追跡するのに利用できる。	主観的要素を含むため、経験豊富なアナリストであってもデータに対する見解が分かれることがある。
チャートは時間単位から100年単位まで幅広い期間のデータ分析に使用することが出来る。	テクニカル分析はパターンや規則性に基づいているが単純に将来が過去から推測できることには限界がある。テクニカル分析はある事象の必然性ではなく事象が起り得る確率に重きを置く。
リアルタイムで、あるいは必要に応じてわずかな時間差に伴った正確なデータに基づいている。	
異なるニーズを満たす多くのテクニカル分析手法と技術がある。	現在のテクニカル分析手法のいくつかは非常に複雑な数学や統計学の考え方に基づいている。
テクニカル分析は市場の動きに合わせて開発されてきたため基本的原理が理解しやすい。	利用される情報が迅速かつ正確であることがテクニカル指標の成否を決める。

② ファンダメンタル分析

ファンダメンタル分析は、投資商品の基礎となる需給関係に影響を与える経済・政治・環境あるいはそれらに関連する様々な要因や統計に基づき、投資商品の将来の価格変動を予測する手法である。例えば、アナリストが企業を投資対象として取り上げる場合、その企業の製品や

サービスにおける需要関係，経営の質や方針，業績，将来の展望や計画などに対して調査をする。

テクニカル分析とファンダメンタル分析の違いを以下の表 2 にまとめる。

表 2 テクニカル分析とファンダメンタル分析の違い

テクニカル分析	ファンダメンタル分析
<ul style="list-style-type: none"> ・市場で実際に起きていることに焦点を置く ・市場動向に基づくチャートは以下のもので構成されている <ul style="list-style-type: none"> ・ 価格 ・ 出来高 ・ 建玉 	<ul style="list-style-type: none"> 市場で何か起こるであろうことに焦点を置く ・ 価格分析において考慮される要因は以下のものである <ul style="list-style-type: none"> ・ 需要と供給 ・ 季節循環 ・ 天候 ・ 政府の政策

3.2.2 テクニカル指標

本研究では，SBI 証券とマネックス証券のテクニカル指標の関するサイト [6] から，特にメジャーな 6 つのテクニカル指標を用い，それらについて説明する。

テクニカル分析に使用する指標をテクニカル指標と呼び，大きく分けて市場の方向性を示すトレンド系指標と価格の過熱感や反転ポイントを示すオシレーター系指標の二つがある。

・トレンド系指標

1. 移動平均線

移動平均線とはある一定期間の終値を合計してその期間数で割った平均値を結んだ線のことである。

- ① 単純移動平均線 (SMA) 一株価の過去一定期間 (通常 5 日, 25 日, 50 日) の価格の終値の平均値をグラフにした線.
- ② 加重移動平均 (WMA) 一直近の価格にデータにより大きな重みをかけて計算する
- ③ 指数平滑移動平均 (EMA) 一直近の価格にデータにより大きな重みをかけて計算するが古い価格は除外せず期間中の価格をすべて平均値の計算に使用する.

2. 一目均衡表

一目均衡表は相場のトレンドや反転のタイミングを視覚的に判断するために用いられる。主に基準線，転換線，先行スパン 2 本，遅行線の五本線で構成されている。

- ① 基準線 = (当日を含めた過去 26 日間の最高値 + 最安値) ÷ 2
- ② 転換線 = (当日を含めた過去の 9 日間の最高値 + 最安値) ÷ 2
- ③ 先行スパン 1 = {(転換地 + 基準値) ÷ 2} を 26 日先行させて表示

- ④ 先行スパン 2 = {当日を含めた過去 52 日間の最高値 + 最安値} ÷ 2}
- ⑤ 遅行線 = 当日の終値を 26 日遅行させて表示

・オシレーター系指標

3. MACD

移動平均線を組み合わせて相場のトレンドや売買のタイミングを見極める指標である。

MACD ライン：短期の移動平均線（通常 12 日）と長期の移動平均線（通常 26 日）の差 MACD

シグナル：MACD ラインの指数平滑移動平均（EMA）をとったもの

MACD ヒストグラム：MACD ラインと MACD シグナルの差

4. RSI

ある銘柄の現在価格を同じ銘柄の過去データと比較する。

$$RSI = 100 - \left(\frac{100}{1 + RS} \right)$$

$$RS = \frac{N \text{ 日間で株価が上昇した日の上昇額の平均}}{N \text{ 日間で株価が下落した日の下落額の平均}} \quad (N = 14)$$

RSI が 70%以上：買われすぎの状態と判断

RSI が 30%以下：売られすぎの状態と判断

5. ストキャスティクス

ストキャスティクスは RSI と同様で「買われすぎ」や「売られすぎ」を表す指標で %K 線と %D 線という 2 本のラインがある。

$$\%K = \frac{(\text{直近の終値} - \text{過去 } n \text{ 日間の最安値})}{(\text{過去 } n \text{ 日間の最高値} - \text{過去 } n \text{ 日間の最安値})} \times 100 \quad (n = 14)$$

$$\%D = m \text{ 日の \%K の単純移動平均}$$

%K, %D が 20%以下：売られすぎの状態と判断

%K, %D が 80%以上：売られすぎの状態と判断

6. ボリンジャーバンド

株価が通常どの範囲で動いているかや、異常な動きが起きた時のサインを表す。

中央線（移動平均線）と上限線（中央線 + 標準偏差 × 2）、下限線（中央線 - 標準偏差 × 2）の 3 本線で構成される。

3.4 貪欲法（グリーディー法）

貪欲法は、近似アルゴリズムの基本的設計技法の一つである。このアルゴリズムは問題を複数の部分問題に分割し、それぞれ独立に評価を行い、その時点で最も評価値の高い選択肢を解に取り込んでいくことで最適解を得るという方法である。使用例として、最短経路問題、スケジューリング問題、ダイクストラ法、ナップサック問題などがある。

3.5 遺伝的アルゴリズム（GA）

進化生物学の概念を応用して、複雑な選択肢の中から最適な解を探索するのに有効なアルゴリズムである。このアルゴリズムでは、最初にランダムに生成した複数の組み合わせ（個体）を進化させながら、最適解を探索する。具体的には、まず個体の「適応度」（予測精度）を評価し、適応度の高い個体を次世代に引き継ぐ。次に異なる個体を組み合わせて新しい候補を生成する「交叉」と、個体の一部をランダムに変更する「突然変異」といった操作を繰り返し行う。このような操作を繰り返し行うことで、適応度が最も高い組み合わせを見つけ最適解を導くことが可能となる。

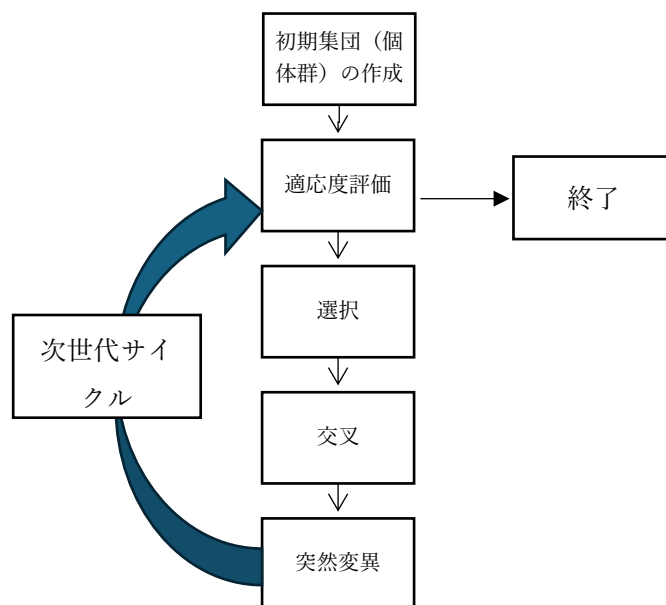


図2 GAの仕組み構造

第4章 実験概要

本章では、本研究の実験の概要及び、使用するデータ、評価関数、実行環境について述べる。

予備実験として、第5章では単一のテクニカル指標から算出される特徴量を組み込んだデータを用いてランダムフォレストモデルで株価変化率の予測を行い、特徴量重要度を測定し、ベースラインとしての個々の指標の有効性を理解する。6章では、複数のテクニカル指標の最適な組み合わせを考え予測精度の向上を目指す。具体的な最適な組み合わせの探索方法は、問題を解く際にその場で最適である解を繰り返して最終的に全体の最適解に到達させる貪欲法と次世代に優れた解を引き継いでいくことで、最適解を導く遺伝的アルゴリズムを活用する。

4.1 使用するデータ

Yahoo!ファイナンス[7]に掲載されている日経225の掲載銘柄の一つであるT9434ソフトバンク(株)の2018年12月19日から2024年11月1日までの約6年間の株価データを用いる。株価データの列は「始値」「終値」「高値」「終値」を使用し、

目的関数としては「翌日の株価変動率： $\frac{\text{翌日の終値}-\text{当日の終値}}{\text{当日の終値}} \times 100(\%)$ 」を組み入れる。

また、talibというライブラリをインポートすることでそれぞれのテクニカル指標の特徴量の計算を行う。本研究で用いる株価の値動きと特徴量をまとめたものについて下記の図3と表3に示す。機械学習における訓練データとテストデータの分割は8:2とする。



図3 ソフトバンク(株)の終値の過去6年間の値動き

表3 本研究で用いる特徴量

テクニカル指標	特徴量	説明
単純移動平均線	sma01	短期（5日）移動平均線
	sma02	中期（25日）移動平均線
	sma03	長期（50日）移動平均線
MACD	macd	MACD
	macdsignal	MACD シグナル
	macdhist	MACD ヒストグラム
RSI	rsi	RSI
ボリンジャーバンド	upper	上限線
	middle	中央線
	lower	下限線
一目均衡表	basic_line	基準線
	turn_line	転換線
	slow_line	遅行線
	span1	先行スパン1
	span2	先行スパン2
ストキャスティクス	%K	%K ライン
	%D	%D ライン

4.2 評価関数

本研究では、予測精度を比較する際は以下の3つの評価関数を用いる。数値が小さいほど誤差が小さく、精度が良いと判断される。 r を実測値、 \hat{r} を予測値、 N をデータ数とする。

① 平均絶対誤差（MAE：Mean Absolute Error）

MAEとは、予測値と実測値の誤差の絶対値を平均した値の和であり、外れ値の影響を受けにくい評価関数である。MAEは以下の式で求められる。

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{r} - r|$$

② 平均二乗誤差（MSE：Mean Squared Error）

MSEとは、予測値と実測値の誤差を二乗して平均をとった値の和であり、大きな誤差を重視するため外れ値の影響を受けやすい評価関数である。MSEは以下の式で求められる。

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{r} - r)^2$$

③ 二乗平均平方根誤差（RMSE：Root Mean Squared Error）

RMSEとは、MSEの平方根を取った値であり、MSEのスケールを元データの単位に戻した評価

関数である。RMSE は以下の式で求められる。

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{r} - r)^2}$$

4.3 貪欲法による最適化

4.3.1 定数・変数の定義

定数の定義は以下のとおりである。

F : 使用するテクニカル指標の集合

w_1, w_2, w_3 : 評価指標の重み

k : 選択可能な指標数の上限

$x_{\text{train}}, y_{\text{train}}$: 訓練データ

$x_{\text{val}}, y_{\text{val}}$: 検証データ

また、設定した定数の値は以下の通りになる。

$F = \{\text{単純移動平均, MACD, RSI, ボリンジャーバンド, ストキャスティクス, 一目均衡表}\}$

$|F|=6$

$k=|F|$

$$w_1 = w_2 = w_3 = \frac{1}{3}$$

変数の定義は以下のとおりである。

x_i : 指標 F_i が選択されているかを表すバイナリ変数

$x_i \in \{0,1\}, \forall_i \in \{1,2,3, \dots, |F|\}$

4.3.2 定式化

Minimize

$$J(S) = w_1 \cdot \text{MAE}(S) + w_2 \cdot \text{MSE}(S) + w_3 \cdot \text{RMSE}(S)$$

subject to

$$1 < \sum_{i=1}^{|F|} x_i \leq k$$

$$x_i \in \{0,1\}, \forall_i \in \{1,2,3, \dots, |F|\}$$

4.4 GA による最適化

4.4.1 定数・変数の定義

定数の定義は以下のとおりである.

F : 使用するテクニカル指標の集合

w_1, w_2, w_3 : 評価指標の重み

k : 選択可能な指標数の上限

$x_{\text{train}}, y_{\text{train}}$: 訓練データ

$x_{\text{val}}, y_{\text{val}}$: 検証データ

N : 個体数

ngen : 世代数

cxpb : 交叉確率

mutpb : 突然変異確率

また, 設定した定数の値は以下の通りになる.

$F = \{\text{sma01}, \text{sma02}, \text{sma03}, \text{macd}, \text{macdsignal}, \text{macdhist}, \text{RSI}, \text{upper}, \text{middle}, \text{lower}, \text{basic_line}, \text{turn_line}, \text{span1}, \text{span2}, \text{slow_line}, \%K, \%D\}$

$|F| = 17$

$k = |F|$

$w_1 = w_2 = w_3 = \frac{1}{3}$

N=20

ngen=10

cxpb=0.7

mutpb=0.2

変数の定義は以下のとおりである.

x_i : 指標 F_i が選択されているかを表すバイナリ変数

$x_i \in \{0,1\}, \forall_i \in \{1,2,3, \dots, |F|\}$

X : 個体 (特徴量選択の集合)

$X = [x_1, x_2, \dots, x_{|F|}]$

4.4.2 定式化

Minimize

$$J(X) = w_1 \cdot \text{MAE}(X) + w_2 \cdot \text{MSE}(X) + w_3 \cdot \text{RMSE}(X)$$

subject to

$$1 < \sum_{i=1}^{|\mathcal{F}|} x_i \leq k$$
$$x_i \in \{0,1\}, \forall i \in \{1,2,3,\dots,|\mathcal{F}|\}$$

4.4 実験環境

本研究における実験環境は以下のとおりである。

表 4 実験環境

CPU	11th Gen Intel® Core(TM) i3-1115G4
OS	Windows 10 Education
Memory (CPU)	8GB
GPU	Intel(R) UHD Graphics
Memory (GPU)	0.6GB
使用言語	Python3.8

第5章 単一指標を用いた株価予測

本章では、ソフトバンク(株)の株価データを用いて単一のテクニカル指標（単純移動平均、MACD、RSI、ボリンジャーバンド、ストキャスティクス、一目均衡表）に関する特徴量をそれぞれ含め、ランダムフォレストモデルを使用して株価の変化率を予測し、各テクニカル指標が株価予測精度にどのように寄与するかを評価する。

5.1 予測結果

予測値に基づいた評価関数の値は表5に記す。

表5 テクニカル指標別予測値の評価関数

テクニカル指標	MAE	MSE	RMSE
ストキャスティクス	0.3706	0.261	0.5109
単純移動平均	0.3822	0.2875	0.5362
一目均衡表	0.4022	0.3139	0.5602
なし	0.4301	0.3821	0.6181
MACD	0.4321	0.3778	0.6147
ボリンジャーバンド	0.4352	0.3851	0.3851
RSI	0.44	0.391	0.6253

表1より、ストキャスティクスがすべての指標で他のモデルを上回る性能を発揮した。次に、単純移動平均と一目均衡表も比較的良好な結果を示した。これらの指標は指標を使用しない場合や他の指標と比べて誤差が小さく、予測精度の向上に一定の効果を発揮していることがわかる。一方で、MACD、RSI、ボリンジャーバンドは、テクニカル指標を使用しない場合と大差がない結果となった。

これらの結果から、テクニカル指標によって予測精度の向上に寄与するものとししないものがあることがわかる。特に、トレンド系指標は名前の通り相場のトレンドに乗った指標であるため株価の値動きに反応しやすいことから予測精度が高くなる傾向がある。一方、オシレーター系指標は相場の過熱感や買われすぎ・売られすぎを表すものであり、株価のトレンドを直接示すわけではないため予測精度が低くなることわかる。

文献[10]より、オシレーターの中でも、MACDはレンジ相場に、ボリンジャーバンドやRSIは変動率の高い相場に弱く、それに対してストキャスティクスは短期的な変動を敏感に捉えることがわかった。そのため、オシレーター指標の中でもストキャスティクスは高い予測結果になったと考える。

第6章 複数のテクニカル指標の組み合わせ最適化

本章では、複数のテクニカル指標を組み合わせることで株価予測の精度を向上させる方法を探る。第5章で扱ったテクニカル指標の特徴量を基に、ランダムフォレストモデルを用いて株価の変化率を予測し、予測精度を評価する。最適化のアプローチとして、貪欲法と遺伝的アルゴリズム (GA) を活用し、株価予測精度を最大化することを目指す。

6.1 貪欲法による予測結果

予測値に基づいた評価関数の値は表6、7に記す。

表6 予測精度の高い組み合わせ指標

組み合わせ指標	MAE	MSE	RMSE
単純移動平均と MACD	0.3136	0.1911	0.4302
単純移動平均と MACD とストキャスティクス	0.3287	0.1941	0.4406
全指標	0.3271	0.1949	0.4415
単純移動平均と MACD と RSI とストキャスティクス	0.3329	0.1954	0.4420
単純移動平均と MACD と RSI とボリンジャーバンドとストキャスティクス	0.3374	0.2005	0.4478

表7 予測精度の低い組み合わせ指標

組み合わせ指標	MAE	MSE	RMSE
RSI とボリンジャーバンド	0.4343	0.3865	0.6217
MACD と RSI	0.4534	0.3657	0.6047
MACD と一目均衡表	0.4262	0.3529	0.5941
MACD と RSI と一目均衡表	0.4376	0.3444	0.5869
RSI と一目均衡表	0.4032	0.3130	0.5595

以上の結果から、以下のことが明らかになった。

「単純移動平均と MACD」の組み合わせが最も高い精度を示した。次に僅差で「単純移動平均

と MACD にストキャスティクス」を加えた指標が高い予測精度を記録した。したがって、単に数多くのテクニカル指標を組み合わせれば精度が上がるわけではなく、予測精度を上げることのできる相性の良い組み合わせがあることがわかった。また、単一指標の予測精度が低かった「RSI とボリンジャーバンド」の組み合わせが最も予測精度が低く、次に僅差で「MACD と RSI」の組み合わせが低い予測精度を示した、これらの結果は表 5 よりテクニカル指標を特徴量として含まないものよりも予測精度が低くなった。これにより、オシレーター指標同士の組み合わせは予測精度の向上に寄与しないことがわかった。

6.2 遺伝的アルゴリズムによる予測結果

GA による特徴量カラムの最適な選択結果は、

```
['sma01', 'sma02', 'macdsignal', 'macdhist', 'turn_line', 'span2', 'slow_line', '%K', '%D']
```

これらの特徴量を用いた評価指標の値を以下の表 8 に表す。

表 8 GA により選択された特徴量を用いた予測精度結果

MAE	0.4229
MSE	0.3598
RMSE	0.5998

上記の結果から、貪欲法による最適解と比較して、GA を用いた場合の予測精度は大幅に低い結果となった。また、単純移動平均やストキャスティクス、一目均衡表を単体で用いた際の予測精度と比較しても、GA による予測精度のほうが劣る結果となった。この結果は、本研究での GA の構築および適用において、パラメータ設定や適応度関数の設計などいくつかの課題を解決していく必要があることが示された。

第7章 考察

本研究では、複数のテクニカル指標を用いた株価予測において、貪欲法と遺伝的アルゴリズムを活用して特徴選択を行い、最も予測精度が高くなる組み合わせを検討した。本章では、これらの結果を基にしたまとめと考察を述べる。表9は表5,6,7,8をまとめたものである。

表9 各予測における評価関数のまとめ

カテゴリ	テクニカル指標	MAE	MSE	RMSE
指標なし	なし	0.4301	0.3821	0.6181
最高精度の単体指標	ストキャスティクス	0.3706	0.261	0.5109
最低精度の単体指標	RSI	0.44	0.391	0.6253
最良精度の指標の組み合わせ	単純移動平均+MACD	0.3136	0.1911	0.4302
最悪精度の組み合わせ	RSI+ボリンジャーバンド	0.4343	0.3865	0.6217
全指標の組み合わせ	全指標	0.3271	0.1949	0.4415
遺伝的アルゴリズムによる組み合わせ	sma01, sma02, macdsignal, macdhist, turn_line, span2, slow_line, %K, %D	0.4429	0.3598	0.5998

7.1 結果における考察

実験結果から、単体指標だけでは限界がある一方で、指標を適切に組み合わせることで予測精度の向上を図れることがわかった。ただし、多すぎる指標の組み合わせや相性の悪い指標同士の組み合わせは、逆に精度が低下してしまうケースも見受けられた。そのため「有効とされているテクニカル指標を単純に併用すれば、予測精度が改善する」というわけではなく、指標同士の整合性や補完関係を考慮し、株価データの特徴に適した組み合わせを選定することでよい投資成果が得られる可能性が高まる。

単純移動平均は長期的なトレンドをとらえるのに有効な指標であり、MACDは短期EMA（12期間）と長期EMA（26期間）をベースに算出されるためこの指標はオシレーターとしてだけでなく、相場のトレンドの把握にも活用される指標である。しかし第5章の結果からもわかるようにMACD単体では、トレンド相場では機能しやすいがレンジ相場には適応しにくいことが確認された。一方、ストキャスティクスは価格の過熱感を示すオシレーター系指標であり、トレンド系指標とオシレーター系指標を組み合わせることで、相場の「方向性」と「過熱感」の

両方をとらえることができる。これにより、トレンド相場とレンジ相場の両方の特徴を示すソフトバンクの株価の動きに対して多面的なアプローチが可能となり、予測精度が向上したと考える。この結果は文献[10]でも示されており、トレンド系+オシレーター系の組み合わせは有効であるといえる。

また、遺伝的アルゴリズムで最適な特徴量として選ばれたのは `sma01`, `sma02`, `macdsignal`, `macdsignal`, `turn_line`, `span2`, `slow_line`, `%K`, `%D` の 9 つであった。これは、最適な組み合わせの結果にはならず改善が見られた。この結果から、遺伝的アルゴリズムは探索空間が広く複雑な問題に対しては有効である一方、一度に多くの特徴量を扱うため不要な特徴量や似た特徴量が選ばれることがありモデルの精度低下に寄与する可能性があることがわかった。今回のデータセットでは、株価予測に使用されるテクニカル指標は過去の価格データを基に計算される特徴量であるため相互に強い相関を持つものが含まれており、これが情報の偏りを生じさせ、予測精度の低下につながった可能性があると考えられる。

7.2 今後の課題

1. 特徴選択手法の改善

本研究で用いた遺伝的アルゴリズムは、探索能力が高いものの不要な特徴量や高相関の特徴量が選ばれる問題が見られた。今後、遺伝的アルゴリズムによって組み合わせ最適化を行う際は LASSO 回帰のような重要度選択や過学習防止効果をもつ回帰分析手法や PCA などの次元削減手法を活用し精度の高いモデル構築を目指す。

2. 異なる市場条件での検証

本研究は特定の市場データをもとに行われたため、結果の一般化可能性が限定的である。今後、異なる市場状況や銘柄での検証を行い、予測モデルの適応可能性を確認することが必要である。

3. モデル構造の改善

本研究では特徴量の選択手法の比較に重点を置いたが、予測モデルそのものの構造改善や比較も必要である。具体的にはブースティング系アルゴリズム (XGBoost, LightGBM) やニューラルネットワークなどを用いて比較し、最適なモデル選択を行う。

第8章 おわりに

本研究では、機械学習手法を用いて複数のテクニカル指標の組み合わせから株価予測精度を検証しテクニカル指標の組み合わせの重要性について考察した。

テクニカル分析は実務的にも学術的にも認められつつある一方で、人間の主観性や恣意性に依存しやすいものであるとの指摘を多く受けている。投資家やトレーダーにとって株価予測は正しい売買判断を行うために必要不可欠であり、多くの投資家やトレーダーは複数のテクニカル指標を組み合わせで行うものを当たり前のように解釈し、テクニカル指標の原理原則を理解せず組み合わせで利用しているのが現実である。そこで本研究は、「現在の価格変動パターンが過去のある時点と似ていれば、そのパターンに基づいて将来の株価予測ができる」という仮説を基本とし、幾何学的な価格変動パターンを機械的に抽出し予測へ活用することでテクニカル指標単体、または組み合わせによる有効性を検証した。これまでの先行研究にもテクニカル指標を用いた株価予測を行う研究はあるものの、テクニカル指標の組み合わせ選択を重点に置いた研究は少ない。また、予測手法としても各説明変数の目的変数に対する重要度・寄与度が評価できるランダムフォレストを使用した。

結果として、テクニカル指標を用いない場合や単体の指標を用いた場合に比べ、複数の組み合わせ指標を組み合わせたほうが予測精度が向上することが示された。一方で、単純に数多くの指標の組み合わせや、相性の悪い指標同士の組み合わせは冗長性や非補完性が生じ、かえって精度が低下する場合もあった。したがって、各銘柄のチャートの特徴や各テクニカル指標の成り立ちや理論を考慮した組み合わせ選択をすることが、さらに精度の高い予測を実現するために重要であることが示された。

また、本研究では遺伝的アルゴリズム (GA) を用いて、最適なテクニカル指標の組み合わせを探索する試みも行った。GA は広範な探索空間に対して効果的であり、理論的には適切な特徴選択が期待される。しかし、今回の結果では、選ばれた特徴量の一部に不要なものや相互に高い相関を持つものが含まれたことから、GA による特徴選択が予測精度の向上に必ずしも寄与しない場合があることが分かった。これは、GA が特徴間の相関や情報の冗長性を十分に考慮できないためであり、特に株価予測のような強い相関を持つデータでは課題となった。今後は、GA に加えて次元削減手法を併用することでこの課題を克服し、より効果的な特徴選択を行う必要がある。

本研究で提案した手法は、株価予測だけでなく、他の時系列データ分析や異なる分野においても応用可能であると考えられる。今回の研究を活かし、異なる市場条件や銘柄データでの検証を進める予測精度向上を図るとともに、他分野でも機械学習を応用させさらなる理解と発展に努めたいと考える。

参考文献

- [1]伊藤博士：テクニカル分析のパフォーマンス考察ーその有効性と問題点の視座からー，大阪論集，Vol.60，No.5，pp.79—91，2010
- [2]ロイター・リミテッド，（訳）小島秀雄，小川真路：ロイター・フィナンシャル・トレーニングシリーズ日本版，テクニカル分析，経済法令研究会，pp.3—38,138,169—204，2001
- [3]徳山相賢，神田裕樹：金融市場における機械学習の活用について，三菱UFJ信託銀行，資産運用情報，pp.10—13，January，2021
- [4]片寄諒亮，吉岡真治；機械学習によるテクニカル分析の影響調査，人工知能学会資料，SIG-FIN-024，pp.144—148，2020
- [5]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，2D3-0S-19a-3，pp.1—3，2017
- [6]マネックス証券，はじめてのテクニカル分析
<https://info.monex.co.jp/technical-analysis/>（参照 2024-11-22）
- [7]Yahoo!ファイナンス，Yahoo!ファイナンス-株価・最新ニュース <https://finance.yahoo.co.jp/>
（参照 2024-11-22）
- [8]横内大介，青木義充：「イメージでつかむ機械学習入門～豊富なグラフ，シンプルな数学，Rで理解する」，技術評論社，pp.8—10，2017
- [9]津田博史，島田康史，鶴田大，藤原暢，河合竜也：「ディープラーニング入門ーPythonではじめる金融データ解析ー」，朝倉書店，pp.5—50，2018
- [10]福永博之，：「テクニカル分析 最強の組み合わせ術」，日本経済新聞出版社，pp.136—169，2024