

機械学習を用いた暗号資産の収益率予測モデルの構築

松永 諒 (九州産業大学 理工学部 情報科学科)

Ryo MATSUNAGA, Department of Information Science, Faculty of Science and Engineering, Kyushu Sangyo University

于 海波 (九州産業大学 理工学部 情報科学科)

Haibo YU, Department of Information Science, Faculty of Science and Engineering, Kyushu Sangyo University

1 はじめに

近年、暗号資産市場は急速に拡大し、多くの投資家やトレーダーが参加している。この暗号資産市場は従来の金融市場とは異なる動きを見せ、価格の急激な変動や取引量の増減が顕著である。また、暗号資産取引所は 365 日 24 時間常に動いており、データ取得や取引を行うための公開 API を用意しているところがほとんどである。よって、人間が市場を確認し取引の対応を行うよりもプログラムを用いて自動で取引を行うほうが他の市場参加者に比べて有利になる部分がある。

そこで、本研究では、暗号資産取引所で公開されている過去の暗号資産の取引データや API を利用して、自動で暗号資産の取引を行うプログラムを試作した。このシステムは同研究室の佐々木と 2 人で開発を行った。本人は、暗号資産自動取引システムの中で過去データの収集機能、収益率予測機能の機能設計及び開発を担当した。過去データ収集機能では、暗号資産取引所の公開 API を利用して過去データの収集を行った。また収益率予測機能の設計では、取得したデータの整形、機械学習モデルの構築、学習したモデルの評価を行った。システムの改良を行いながらモデルの評価を行っていき、結果として、ある程度の正確性を持ったモデルの構築ができた。

本論文では、暗号資産や暗号資産取引所について紹介し、収益率予測機能の開発時に使用した関連技術及びシステム設計、システム実装、評価結果などを記す。また、システムを実装する段階で得た知見や、問題点、評価結果から得られた情報を元に、今後の展開やシステムの改善点についての考察などを行う。

2 関連技術紹介

本節では、本研究に関連する技術を紹介する。

2.1 暗号資産及び暗号資産取引所

暗号資産 (Crypto Currency) とは、分散型台帳技術 (ブロックチェーンなど) を基板として作られている電子データでのみやりとりされる通貨である。デジタル通貨 (Digital Currency) と呼ばれることもある。暗号資産の種類には 20,000 種類以上

あり、代表的な物でいうと、BTC (ビットコイン)、ETH (イーサリアム)、XRP (リップル) などがある。現在もその数は増え続けている [1]。

暗号資産取引所とは、日本円や米ドルなどの法定通貨を暗号資産と交換する、または暗号資産同士を交換することのできるサービスである。日本だけではなく様々な国で運営されている。日本国内の代表的な暗号資産取引所には bitFlyer(ビットフライヤー)、Coincheck(コインチェック)[2]、GMO コイン [3] などがある。暗号資産取引所では、公開 API が整備されてあることがほとんどである。このような API を通して世界各国の取引所の相場データがリアルタイムで取得でき、注文を送ることが出来る点が機械学習の題材に適していると言える。暗号資産の取引所は日本国内、海外の両方に大量に存在しているが、取引量、運営元の信頼性、知名度を注目しながら取引所の選定を行った。海外の取引所はデータの取得は許可しているが、日本人の取引を認めていない取引所も存在するため今回は除外した。

今回の研究では Coincheck を取引所として利用した。しかし、今回 Coincheck の API を使用して過去のデータの取得ができなかった。さらにコインチェックでは過去のデータを CSV ファイルにして配布する等も行っていなかったため本研究では過去のデータの取得を GMO コインで行う事とした。

2.2 特徴量

特徴量とは、機械学習やデータ分析において、分析対象のデータの中の予測の手がかりとなる変数の事である。特徴量が必要な理由は、モデルがデータのパターンや関係性を学習し、未知のデータに対して予測を行うためである。今回はその特徴量として RSI[4]、EMA[5]、SMA[6] を選択した。この 3 つは、金融市場の分析や予測において一般的に使われている技術的指標であり、中でも SMA、EMA は計算が単純であるため理解しやすく実装も容易であること、RSI は計算こそ複雑になるが多くのプログラムで利用されている指標のため今回特徴量として選択した。またどの特徴量にもメリット、デメリットが存在するため他の特徴量と組み合わせることも重要である。以下では、本研究で利用した特徴量について紹介する。

2.2.1 RSI (相対力指数)

RSI (Relative Strength Index) [4] とは、金融市場で使用されるテクニカル分析の一種であり、特定の期間内の値動きの強さと早さを示す指標である。RSI は計算式 (1) で求められる。

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

RS (Relative Strength) とは「平均利得 (Average Gain)」を「平均損失 (Average Loss)」で割ったものであり、計算式 (2) で求められる。

$$RS = \frac{AverageGain}{AverageLoss} \quad (2)$$

Average Gain は直近 n 回の上昇幅の合計を n で割った数値であり、Average Loss は直近 n 回の下落幅の合計を n で割った数値である。 n は RSI の計算期間を表しており、通常 14 期間を用いた RSI が使われている。RSI の値は通常、0 から 100 までの範囲で表される。RSI が 70 以上である場合、「過買い (Over bought)」状態を示し、価格が高騰しすぎている事を示す。逆に RSI が 30 以下である場合、「過売り (Over sold)」であることを示す。

RSI のメリットとして、RSI は価格の動きと強さを示すため、価格と RSI の動きが逆行する場合、トレンドの逆転を示す可能性がある。このことを逆強弱指標といい、トレンド転換のシグナルとなる事がある。また、他のテクニカル指標との相性も良いため併用して使用することでより勝率を上げることができる。RSI のデメリットとして、RSI は特に流動性の低い市場や突発的なニュースなどで急激な値動きをした場合、誤ったシグナルを出す場合がある。さらに強いトレンドがずっと続く場合や、相場が一方方向に動く場合に利用することが難しいというデメリットもある。

2.2.2 EMA (指数平滑移動平均)

EMA (Exponential Moving Average) [5] とは、時系列データの平滑化やトレンドの抽出に使用される統計的な指標の 1 つである。EMA は過去のデータに対して指数関数的な重み付けを行い、最新のデータにより大きな重みを与える事でより迅速に変動に対応する特徴がある。EMA は式 (3) で求められる。

$$EMA_t = \alpha X_t + (1 - \alpha)EMA_{t-1} \quad (3)$$

式 (3) 内の X_t は現在の時点 t での観測値で EMA_t は時点 t での EMA の値である。また α は平滑化のための係数であり、通常 0 から 1 の間の値をとる。

EMA のメリットとしては、EMA は最新の価格データにより高い重みを与えるため、滑らかでトレンドをより正確に表示する。変動する市場に適応

しやすいという特性もあり、柔軟に対応することができる。EMA のデメリットとして、EMA は指数的な重みを用いるため、直近の価格データが過去のデータよりも大きな影響を与える。そのため急激な変動が起こった場合、過去のデータが反映されにくくなる。また急激な変動が起こった場合、過敏に反応してしまうため、それがノイズになってしまう場合もある。

2.2.3 SMA (単純移動平均)

SMA (Simple Moving Average) [6] とは、時系列データの平滑化やトレンドの把握に使われる統計的な指標の 1 つである。SMA は単純な平均を求め指標であり、データの一定期間内の平均値を計算する。SMA は式 (4) で求められる。

$$SMA_t = \frac{X_{t-n+1} + X_{t-n+2} + \dots + X_{t-1} + X_t}{n} \quad (4)$$

数式 (4) で SMA_t は時点 t での SMA の値であり、 X_t は時点 t での観測値である。また SMA は単純な移動を計算するため、特定のデータ期間内のポイントの合計を期間の数で割ることで得られる。期間が n である場合、 n 個のデータポイントの合計を計算しそれを n で割り平均を求める。

SMA のメリットとして、計算が単純で初心者でも理解しやすい。また最新のデータと過去のデータを平等に扱うため安定している。SMA のデメリットとしては、すべてのデータに同じ重みを与えるため急激な変動があった場合その変動を捉えきれずトレンドの転換を遅れて捉えてしまう事がある。

2.3 機械学習

機械学習 (Machine Learning) [7] とは、コンピュータが膨大なデータをもとにして、複数のルールやパターンを学習し、分類や予測する技術である。機械学習のモデルには様々なものがある。例えば、時系列解析モデル、回帰モデル、深層学習モデル、アンサンブルモデルなどがある。RNN (Recurrent Neural Network) とは、ニューラルネットワークの一種であり、特に時系列データやシーケンスデータの解析に特化したモデルである。LSTM (Long Short-Term Memory) [8] とは RNN の一種であり、過去の情報を長期記憶しておく記憶セルの導入により、RNN が持っていた「長期記憶の消失」というデメリットをある程度改善したものである。GRU (Gated Recurrent Unit) [9] とは LSTM の簡易版である。LSTM ゲート機構の 3 つのうち出力ゲートをなくして 2 ゲート構成にして、計算効率化、軽量化を狙った LSTM の改善版である [10]。本研究ではこの中で、時系列解析モデルに含まれる LSTM モデルと GRU モデルを利用する。ネットにあげられているプログラムの中にも LSTM モデルを用いたものが多く存在しているため、今回は LSTM モデルと、LSTM モデルの計算量を減らしたモデルである GRU モデルを使用することにした。書面の

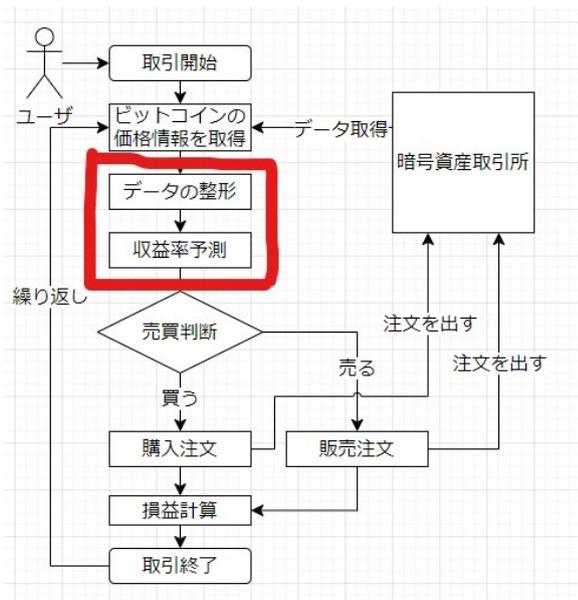


図1 暗号資産自動取引システムの全体像

制限のため、機械学習モデルについての詳細を省略する。

3 暗号資産自動取引システムの設計及び実装

本節では、今回開発する暗号資産自動取引システムについて簡単に紹介して上で、主に本人は担当している暗号資産収益率予測機能の設計及び実装について述べる。

3.1 暗号資産自動取引システムの概要

今回開発した暗号資産自動取引システムの全体像を図1に示す。本システムでは直接取引所からデータを取得し自動で売買を行うシステムである。まず、暗号資産取引所から暗号資産のリアルタイムデータを取得し、データの前処理を行い、今回作成した収益率予測機能を利用して収益率を予測する。その予測結果を基に暗号資産の売買を行い、それを繰り返すというシステムである。

本研究では、主に太い四角の赤線で囲まれている収益率予測部分の機能を実現した。暗号資産収益率予測モデル構築の流れを図2に示す。まず、暗号資産取引所から過去のビットコイン (BTC) の相場データを取得し、選択した機械学習のモデルに合わせたデータの前処理を行う。そのあと前処理したデータをトレーニングデータとテストデータに分割し、学習を行い、予測結果を出す。最後に、予測結果をもとにモデルの評価を行い、より評価の高いモデルを使用する。今回の収益率予測モデル構築にあたり、関連サイト [11], [12] を参考にした。

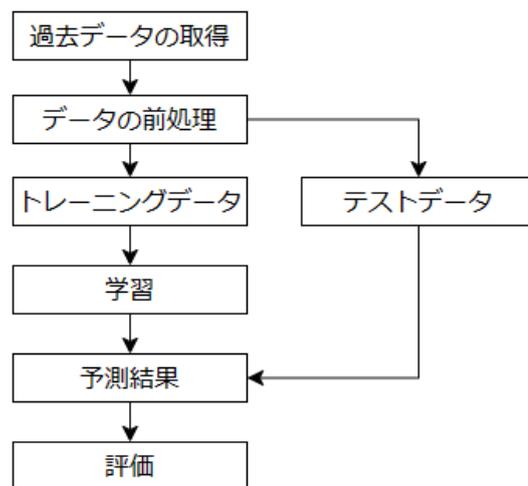


図2 暗号資産収益率予測モデル構築の流れ

3.2 暗号資産収益率予測モデルの構築

本研究では、BTCの過去のデータから次のBTCの収益率を予測するモデルを設計し、実装した。学習モデルとして、LSTMモデル [8] と GRUモデル [9] の2種類のモデルを構築し、実装した結果を比較した。以下では収益率予測モデルの構築の処理の流れについて説明する。

3.2.1 過去データの取得

今回はGMOコインのAPIドキュメントを参考にして、BTCの始値 (Open)、高値 (High)、低値 (Low)、終値 (Close)、出来高 (Volume) の5つのOHLCVの過去のデータを取得した。今回は2023年10月2日までの100日分の1分足のデータを取得した。取得したデータはCSVファイルとして保存した。

3.2.2 データの前処理

データの前処理は、以下の順番で行った。

- データの読み込み: 取得した過去データをCSVファイルとして保存しているため、それを読み込み、始値、高値、低値、終値、出来高の100日分の1分足のデータを取得する。
- 目的変数の生成: 10分後の収益率を目的変数とする。
- 特徴量の生成: BTCの始値、高値、低値、終値、出来高、RSI、SMA、EMAを特徴量として用いる。またどれぐらい前までの情報を含むかが重要となるが、今回は過去10分間のデータを用いて10分後の収益率を予測する。
- 欠損値の穴埋め: 今回RSI、SMA、EMAを計算する過程で欠損値が出現する。そのため欠損値を埋める必要がある。欠損値の穴埋めには様々なやり方があるが、今回は前の時間の値を利用して欠損値の穴埋めを行う。

- データの正規化: 機械学習を行う際にはデータの正規化を行う必要がある。データの正規化にも様々なやり方があるが、今回は最小-最大スケールリング (Min-Max Scaling) というデータを 0 から 1 の範囲にスケールリングする方法を用いてデータの正規化を行う。
- データの分割: 今回は全データをトレーニングデータとテストデータに分割する必要がある。一般的に 7 : 3 の割合で分割することが多いため、今回はそれにならない、全体の 70 % をトレーニングデータ、30 % をテストデータとする。

3.2.3 学習モデルの構築

以下では学習モデルの構築の流れについて説明する。まず、使用する機械学習のモデルを選択する。今回は LSTM[8] と GRU[9] という 2 つの機械学習モデルを利用する。今回 LSTM と GRU の 2 つを選択して理由として、ネットにあげられているプログラムの中にも LSTM モデルを用いたものが多く存在しているためである。また、ファムホアントウイリンらの研究 [13] では、LSTM モデルの性能が良いとされていたため、LSTM モデルと LSTM の計算量を改善したモデルの GRU モデルを選択した。

次に、モデルの作成、各パラメータの設定、学習を行う。今回は LSTM モデルと GRU モデルの 2 つのモデルを作成し予測を行ったが、LSTM モデルを例にして、以下でモデルの作成、学習等について説明する。

- LSTM 層: 3 つの LSTM 層を構築した。各 LSTM 層は 32 のユニットを持つ。
- ドロップアウト層: 各 LSTM 層の後にドロップアウト層が 2 層ある。ドロップアウトは過学習を抑制するための手段で、ランダムなユニットを無効にする。ドロップアウト率を 0.2 に設定している。この値は状況に応じて変更できる。
- 全結合層: 最後に全結合層がある。この層のユニット数は 1 で、出力層となる。
- モデルのコンパイル: 最適化アルゴリズムを Adam に設定し、学習率を 0.0005 に設定している。Adam は一般的に使われている最適化アルゴリズムで学習率の調整を自動で行う。損失関数を MSE (平均二乗誤差) に設定している。
- モデルの学習: 今回はエポック数を 150 に設定し、150 回学習させた。ミニバッチのサイズを 64 に設定し、学習は全データを一度に使わず、小さなバッチに分割して学習させ、メモリの使用効率を向上させる。検証用のデータを指定し、モデルがトレーニングデータに過剰に適合していないかの確認を行う。

4 暗号資産収益予測モデルの評価

本研究では構築した暗号資産収益予測モデルを実装し、評価を行った。実装にあたって、エディタに Visual Studio Code を使用し、開発言語として Python を使用した。また、Python の数値計算を支援するライブラリである NumPy、データ操作と解析を行うためのライブラリである Pandas、機械学習のためのライブラリである Sklearn と TensorFlow 等を利用した。

4.1 評価の概要

評価では、機械学習のモデルがどれだけ良くしているかを測定するプロセスである。モデルの評価には、訓練データ (モデルの学習に使用されたデータ) とテストデータ (未知のデータに対する性能評価に使用されるデータ) の 2 つに分割される。モデルが訓練データに過度に適合する「過学習」を避けるために、訓練データとテストデータは別々に保持する。

今回は評価指標として R2[7]、MAE[14]、RMSE[15] の 3 つを使用する。この 3 つを選択した理由として、それぞれ異なる観点からモデルの評価を行うため、総合的なパフォーマンス評価ができるからである。また、参考しているファムホアントウイリンらの研究 [13] でも使用されている評価指標だったため、今回はそれにならない、この 3 つの評価指標を使用する。

4.2 R2 による予測機能の評価

まず、R2 (R-squared) [7] による予測モデルの機能の評価を行った。以下では、評価方法と評価結果について述べる。

4.2.1 評価方法

R2 (R-squared) とは、統計学や機械学習などで使用される評価指標の 1 つである。R2 は、モデルが目的変数の変動をどれだけ説明出来るかを示す指標である。R2 は式 (5) で求められる。

$$R2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n ((y_i - \bar{y})^2)} \quad (5)$$

数式 (5) において、各 y_i は実際の目的変数の値、 \hat{y}_i はモデルによる予測値、 \bar{y} は目的変数の平均値、 n はデータポイントの数を表す。一般的に、 $0 \leq R2 \leq 1$ の範囲で評価され、高いほどモデルの性能が良いことを示す。ただし R2 は外れ値の影響を受けやすいため、他の評価指標と併用する事が推奨される。

4.2.2 評価結果

R2 (R-squared) による予測モデルの機能の評価結果は表 1 と表 2 で示す。表 1 は GRU モデルの R2 の評価結果であり、表 2 は LSTM モデルの R2 の評価結果である。

表 1 GRU モデルの R2

特徴量	訓練データ	テストデータ
なし	0.6749322925444785	0.6651717009689513
RSI	0.738622460018331	0.729981510177558
SMA	0.6955070654469848	0.6818013951883636
EMA	0.6804925596024708	0.6706496945695288
RSI SMA	0.782430750876208	0.776792664065447
RSI EMA	0.781083560370537	0.770992709180339
SMA EMA	0.683036049700632	0.6697838171951207
RSI SMA EMA	0.7508283973584714	0.7408636292074918

表 2 LSTM モデルの R2

特徴量	訓練データ	テストデータ
なし	0.7266783812254729	0.7172616257296095
RSI	0.770878586338471	0.7650348516411895
SMA	0.7239908181506745	0.7135706859935746
EMA	0.73885384742825	0.7291627314738351
RSI SMA	0.7959045696849647	0.782478303889052
RSI EMA	0.7906108779525182	0.7802905680805978
SMA EMA	0.7424865128989024	0.7313102216635474
RSI SMA EMA	0.7837917390137639	0.7702482020379458

4.3 MAE による予測機能の評価

次に、MAE (Mean Absolute Error) [14] による予測モデルの機能の評価を行った。以下では、評価方法と評価結果について述べる。

4.3.1 評価方法

MAE (Mean Absolute Error) とは、回帰モデルの性能を評価するための指標の 1 つである。MAE は予測値と実際の値との絶対値の平均を計算し、モデルの予測が実際のデータとどれだけ離れているのかを示す。MAE は式 (6) で表される。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

数式 (6) において、 n はデータポイントの数を表し、 y_i は実際の目的変数の値、 \hat{y}_i はモデルによる予測値を表す。MAE は絶対値を使用しているため、予測誤差の単位に敏感でないという特徴がある。例えば、予測が通常よりも 10000 円高い場合と 100000 円高い場合でも同じくらいの影響を持つ。また MAE は外れ値がある場合でも、その影響を比較的受けにくいという特徴がある。

4.3.2 評価結果

MAE (Mean Absolute Error) による予測モデルの機能の評価結果は表 3 と表 4 で示す。表 3 は

GRU モデルの MAE の評価結果であり、表 4 は LSTM モデルの MAE の評価結果である。

4.4 RMSE による予測機能の評価

最後に、RMSE (Root Mean Squared Error) [15] による予測モデルの機能の評価を行った。以下では、評価方法と評価結果について述べる。

4.4.1 評価方法

RMSE (Root Mean Squared Error) とは、回帰モデルの性能を評価する指標の 1 つである。RMSE は予測値と実際の値との差の平方の平均を計算し、その後平方根をとることで、誤差の絶対値の平均を示す。RMSE は式 (7) で表される。

$$RSME = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

式 (7) において、 n はデータポイントの数を表し、 y_i は実際の目的変数の値、 \hat{y}_i はモデルによる予測値を表す。RMSE は誤差を平方するため、元のデータと同じ単位を持つ。平方を用いているため、大きな誤差が生じた場合に影響が大きくなるため、外れ値の影響を考慮することが出来る。

表 3 GRU モデルの MAE

特徴量	訓練データ	テストデータ
なし	0.050227986160070355	0.05121246803902003
RSI	0.044045224578800564	0.04442748482031604
SMA	0.049416153475860355	0.04979394110662729
EMA	0.052575282920683045	0.053627303170900836
RSI SMA	0.04101977038441864	0.041180107072966876
RSI EMA	0.04149273756397863	0.04188118148380276
SMA EMA	0.05248781691391784	0.053348590394212456
RSI SMA EMA	0.04504523692404792	0.046043624752074395

表 4 LSTM モデルの MAE

特徴量	訓練データ	テストデータ
なし	0.04581846200215947	0.046739969770889955
RSI	0.0448039020541127	0.0455486349658181
SMA	0.044352569446400605	0.04524741272751368
EMA	0.04477239856000979	0.04562683819960324
RSI SMA	0.04063198222407326	0.04046958420563298
RSI EMA	0.04236332916196131	0.04322542356244045
SMA EMA	0.044597866617397654	0.04553965578779903
RSI SMA EMA	0.044539640297429602	0.04483013899748297

4.4.2 評価結果

RMSE (Root Mean Squared Error) による予測モデルの機能の評価結果は表 5 と表 6 で示す。表 5 は GRU モデルの RSME の評価結果であり、表 6 は LSTM モデルの RSME の評価結果である。

4.5 考察

今回は LSTM モデルの RSI、SMA の特徴量を追加したモデルが最も評価が高かった。しかし、どのモデルも R2 を見てみると 0.6 以上を示しており、一定以上の正確性をもっていると言える。また 3 つの評価指標を確認すると学習データとテストデータが大幅に乖離していないため過学習をしていないことが確認できる。そのため今回のモデルの構築は成功したと考えられる。

今回使用した特徴量では RSI を追加すると目に見えて評価が向上したが、SMA、EMA を追加してもあまり評価が変わらなかった。RSI は過去の一定期間の価格変動を元に算出されるため、価格の特徴を抽出するための役に立ったと考えられる。これによって価格の変動パターンや、サイクルを学習しやすくなり、適切なタイミングで予測が可能になったため、モデルの精度が向上したと考えられる。しかし、どのモデルでも価格が大きく上昇や下降する場合にその変化を捉える事が出来ていなかった。そのためその変動を捉えることの出来る特徴量を追加

できるとさらなる評価の向上につながると考えられる。

3 つの特徴量を組み合わせたものよりも RSI、SMA の 2 つを組み合わせたモデルの方が評価が高かった。EMA は最新のデータに重みを大きく置くため、急激な変動に対して敏感だが、SMA はすべての評価に対して均等に重みを置く。そのため今回のデータで市場の変動が急激だった箇所が少なかったため EMA がノイズとなってしまいモデルのパフォーマンス向上に寄与しなかったと考えられる。

評価指標として R2、RMSE、MAE の 3 つを使用した。今回はどの評価指標を見ても LSTM モデルの RSI、SMA の特徴量を追加したモデルが一番良い評価だったが、モデルを改善していく上で一番評価の高いモデルがばらけてしまうことも考えられる。そこでどの評価指標を重要視するかは目的によって変更していく必要がある。一般的には外れ値の影響が大きい場合には RMSE、MAE が適しており、予測誤差の大きい場合も RMSE、MAE が適しており、変動理由の説明が重要な場合には R2 が最も適している。今回の収益率予測モデルでは目的変数の変動を正確に捉えられていることが最も重要だと考えられるため 3 つの評価指標の一番高いモデルがばらけた場合には R2 を重要視するのが良いと考える。

表 5 GRU モデルの RMSE

特徴量	訓練データ	テストデータ
なし	0.08675475276067778	0.08102379390046484
RSI	0.07122767571808837	0.06787396519727097
SMA	0.08396435808456196	0.07798685678958711
EMA	0.08995691951096763	0.083938365058498
RSI SMA	0.06824213284287953	0.06730549572658496
RSI EMA	0.06878485107076622	0.06863413099436132
SMA EMA	0.08566655404182509	0.08046382487380561
RSI SMA EMA	0.07338204757382938	0.07147107946836288

表 6 LSTM モデルの RMSE

特徴量	訓練データ	テストデータ
なし	0.07955058072044241	0.0744550011908708
RSI	0.07779299560097867	0.07276093121531924
SMA	0.07698996693190652	0.07227573218308717
EMA	0.07775855441434507	0.07287116441979366
RSI SMA	0.0684914703772196	0.06615394207853025
RSI EMA	0.06854928509979566	0.06700792216197486
SMA EMA	0.07721583143996898	0.07258168883703013
RSI SMA EMA	0.07531037964103964	0.07173913604204792

5 おわりに

本研究の目的は暗号資産の自動取引システムを作成し、市場参加者が常に監視し市場に張り付く必要なく、暗号資産を売買出来るようにする事である。本研究ではそのシステムの中で、暗号資産の収益率予測機能を構築し、3つの特徴量を組み合わせてモデルの評価やモデルの見直しを行った。

今回作成したモデルを今後さらに改善する事が重要だと考えられる。例えばモデルのハイパーパラメータ（学習率、正則化項の重みなど）を調整したりモデルの複雑性を調整したりなど様々な方法でモデルの改善を行える。また、どのモデルが良いのか適宜評価を行いながら改善していく必要がある。また、モデルについても、今回は LSTM モデルと GRU モデルを利用したが、これらのモデル以外のモデルの使用も検討する必要がある。

今回使用した特徴量は SMA、EMA、RSI の 3つである。この 3つ以外にもビットコインのヒストリカルデータから計算できる特徴量もあり、そのほかにも様々な特徴量がある。それらの特徴量を組み合わせることより評価の良いモデルの作成ができると考えられる。また、ヒストリカルデータから計算できる特徴量の他にもニュースやソーシャルメディアからの情報や、外部要因なども特徴量として扱う

ことができる。適宜評価を行いながら、どの特徴量が重要なかを確認し、特徴量の選定を行っていく必要がある。また、特徴量もただ追加すれば良い訳ではなく、今回のように組み合わせによっては追加することで逆にノイズとなってしまう場合もあるため、特徴量の選定は注意を払いながら行う必要がある。

大きく分けて、上の 2つを改善していく必要があるが、この 2つの他にも損失関数の変更や、評価指標の見直し、使用する過去データの量など改善できる部分があるため、これらの課題に取り組みながらより信頼性の高い収益率予測モデルを構築していきたい。

参考文献

- [1] CoinMarketCap, 暗号資産時価総額上位 100, <https://coinmarketcap.com/ja/>.
- [2] Coincheck, <https://coincheck.com/ja/>.
- [3] GMO コイン, <https://coin.z.com/jp/>.
- [4] au カブコム証券, テクニカル分析 abc, “第 8 回 RSI (Relative Strength Index)”, <https://kabu.com/investment/guide/technical/08.html>.
- [5] マネーパートナーズ, “EMA (Exponential Moving Average) 指数平滑移動平均線”, <https://www.moneypartners.co.jp/support/tech/ema.html>.

- [6] ThinkMarkets, “単純移動平均線 (SMA) とは?”, <https://www.thinkmarkets.com/jp/learn-to-trade/indicators-and-patterns/indicators/simple-moving-average-sma-indicator/>.
- [7] Sebastian Raschka, Vahid Mirjalili, 福島真太郎 (監修), 株式会社クイープ (翻訳), “[第 3 版]Python 機械学習プログラミング達人データサイエンティストによる理論と実践”, 株式会社インプレス, 2020 年 10 月 22 日.
- [8] CVML エキスパートガイド, “LSTM (Long Short-Term Memory)”, <https://cvml-expertguide.net/terms/dl/rnn/lstm/>.
- [9] CVML エキスパートガイド, “GRU (Gated Recurrent Unit, ゲート付き再帰ユニット)”, <https://cvml-expertguide.net/terms/dl/rnn/gru/>.
- [10] Sakib Ashraf Zargar, “Introduction to Sequence Learning Models: RNN,LSTM,GRU”, April 2021, https://www.researchgate.net/publication/350950396_Introduction_to_Sequence_Learning_Models_RNN_LSTM_GRU.
- [11] Qiita, “WaveNet を用いた仮想通貨自動売買 (前編-学習-)”, <https://qiita.com/SY122095/items/4cc734debbd4f78a4dcf>.
- [12] Zenn, “【PyTorch+LSTM】仮想通貨の予測”, <https://zenn.dev/dumbled0re/articles/744c783a8b3992>.
- [13] ファムホアントウイリン, 荒平高章, ‘長短期記憶ネットワークを用いたビットコインの価格予測’, 情報処理学会九州支部, 研究報告書, 2022 年. <https://www.ipsj-kyushu.jp/page/ronbun/hinokuni/1011/Papers/A12-1.pdf>
- [14] Sebastian Raschka, Vahid Mirjalili, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance”, *Inter Research CR* vol.30, pp79–82, 2005.
- [15] T. Chai, R.R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature,” *Geosci. Model Dev.*, vol.7, issue 3, pp.1247–1250, 2014. <https://gmd.copernicus.org/articles/7/1247/2014/>