

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

ファムホアントウイリン¹, 荒平高章¹

概要: 暗号通貨には、ビットコイン、リップル、イーサリアム、イーサリアムクラシック、ライトコインなどの種類がある。本研究では、特に人気のある暗号通貨、すなわちビットコインに焦点を当てた。暗号通貨の中でもビットコインは、投資家、研究者、トレーダーなど、さまざまな団体に受け入れられている。本研究では、ビットコインの価格変動に対応し、高い精度を得るために、LSTM (Long short-term Memory) を用いた深層学習ベースの予測モデルを実装することを目的とする。平均絶対誤差 (MAE)、R 二乗スコア (R²)、二乗平均平方根誤差 (RMSE) に応じて、LSTM モデルは ARIMA モデルや Facebook Prophets モデルよりも優れた予測値を示すことがわかった。したがって、LSTM モデルはビットコインの価格予測に対して最適なアルゴリズムであると考えられる。

キーワード: Long Short-Term Memory, LSTM, Price Prediction, Bitcoin, Cryptocurrency, ARIMA, FB Prophet

1. はじめに

暗号通貨は、デジタルマーケティングにおける印象的な技術的成果であるデジタル通貨または仮想通貨の形態である。暗号通貨は、世界中の商取引で最も好まれ、利用されている[1,2]。2020年4月現在、市場には5,000以上の暗号通貨が存在している。その中でも、最も有名で優れたものはビットコインである[3]。あるウェブサイトによると (<https://coinmarketcap.com/>)、仮想通貨の市場価値はおよそ4,213億円であるが、その時々で変動していると報告されている。ビットコインはすべての取引は第三者によって規制・管理されていない。また、顧客間の第三者による介入は不可能である。ビットコインの時価総額は時々刻々と増加し、現在ではおよそ6,700億円以上が公開されている。さらにビットコインはオープンソースのため、明確性・透明性・単純性だけでなく、それらゆえの時間の節約効果もあり、世界のすべての暗号通貨をリードしている。

機械学習 (ML) とは、未来を予測する人工知能の一種である。過去のデータに基づいて ML ベースのモデルは、他の予測モデルと比べて様々な利点がある。先行研究により、正確な結果を得ることができるばかりでなく、より良い結果を得ることができる。実際の結果と同じになるばかりか、その精度を高めることができる。文献[3, 4]では、LSTM モデルを使用して、予測後の数日間に 12,600 USD 以上の結果を予測する可能性のある Yahoo Finance を通じた株式市場での Bitcoin の予測と方法を発見している。近年は、暗号通貨の価格を予測するための信頼性の高い方法の開発が重要であるため、研究者はより革新的なモデルに焦点を当てている。Saxena の研究[10]は LSTM と ARIMA モデルを用いて、ビットコイン価格の最小精度を検討した。Ben Letham, Sean J. Talor の研究[6]は Facebook のオープンソースに基づいて、FB Prophet モデルを用いて、ビットコイン価格を予測している。しかし、これらの予測モデル間において複数

での予測精度について検証した研究はほとんどない。

そこで、本研究では、ヤフーフাইナンスのビットコイン価格データを用いて、30 タイムステップ、合計 6 年以上の情報を取り、次のタイムステップの終値を予測した。データはタイムスタンプによってトレーニングセットとテストセットに分け、モデルフィッティングの前に分析が行われる。予測手法として、LSTM モデル、ARIMA モデル、FB Prophet モデルを用いた。これらのモデル間の予測結果について考察を行ったので報告する。

2. 原理と方法

2.1 予測方法

Bitcoin (BTC) の過去のデータから暗号通貨の価格を予測した。LSTM (Long Short-Term Memory), ARIMA (Auto Regressive Integrated Moving Average), Facebook Prophet の 3 種類のアプローチを提示し比較した。Fig 1 はデータセットの処理方法である。データの収集から始まり、データの可視化プロセスにより、暗号通貨のデータの挙動を説明し、探索する。Fig 1 に基づき、データの収集からプロセスを開始した。データセットは、2014年9月17日から2021年7月31日まで7年間で、米ドル為替レートに基づいてヤフーフাইナンス株式市場から収集され、本研究では、この時系列データを使用して、データセットとして 2510 個のデータを CSV 形式で出力した。Fig 2 にデータセットのサンプルデータを示す。Fig 3 は、収集したデータセットを用いた BTC の終値の推移である。終値は 2020 年末まで徐々に増加し、その後急激に増加し、以前のデータのピークを上回ったことがわかる。

¹ 九州情報大学
Kyushu Institute of Information Sciences

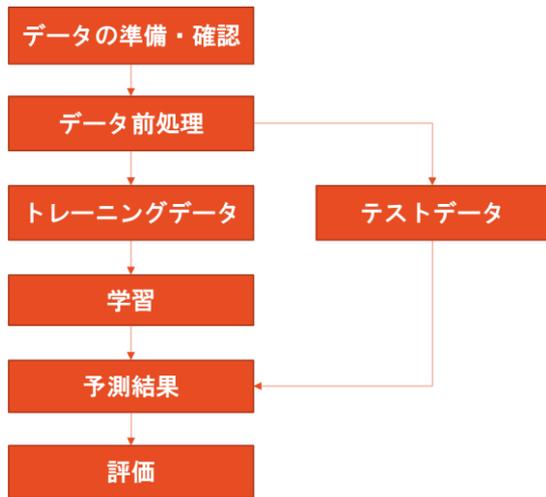


Fig 1 Prediction Processing

	Date	High	Low	Open	Close	Volume	Adj Close
0	2014-09-17	468.174011	452.421997	465.864014	457.334015	21056800.0	457.334015
1	2014-09-18	456.859985	413.104004	456.859985	424.440002	34483200.0	424.440002
2	2014-09-19	427.834991	384.532013	424.102997	394.795990	37919700.0	394.795990
3	2014-09-20	423.295990	389.882996	394.673004	408.903992	36863600.0	408.903992
4	2014-09-21	412.425995	393.181000	408.084991	398.821014	26580100.0	398.821014

Fig 2. Sample data

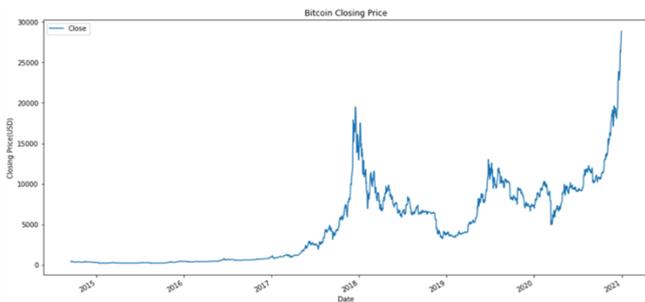


Fig 3. Close price History data of Bitcoin

2.2 Long Short-Term Memory (LSTM)

LSTM(Long Short-Term Memory)は、RNN(Recurrent Neural Network)の拡張として 1995 年に登場した、時系列データ(sequential data)に対するモデル、あるいは構造の 1 種である。LSTM は RNN のバリエーションで、同じようなリカレント構造または鎖状の構造を持つが、より多くの層を持つ。LSTM の鍵は、記憶を司るセル C_t と、セルへの情報の追加・削除を制御する 3 つのゲート(入力ゲート、出力ゲート、忘却ゲート)である。 Fig 4 は、LSTM アルゴリズムの構造を示している。

LSTM の前方学習過程は、以下の式で定式化できる。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$C_t = f_t * C_{t-1} + \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = O_t * \tanh(C_t) \quad (5)$$

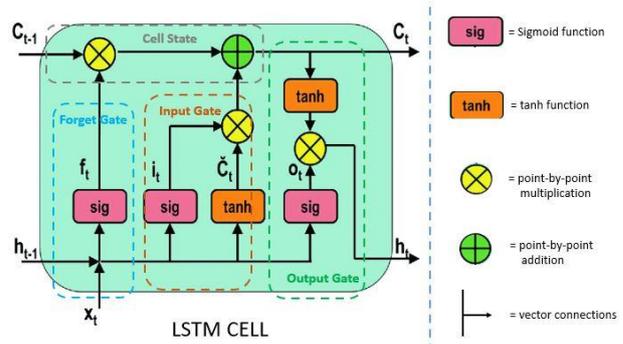


Fig 4. Basic structure of a LSTM unit [7]

LSTM のユニットは 4 つの部分から構成されています。最初の部分は忘却ゲート層と呼ばれ、Eq.(1)に従って、どの前の情報をセルから捨てるかを決定する。2番目の部分は新しい情報を作る部分で、Eq.(2)、Eq.(3)に従って入力ゲート層が含まれる。次のステップ Eq.(4)では、前の 2 つの部分から新しい情報を組み合わせることによって、前のセルを修正する。最後に、Eq.(5)にしたがって、現在の状態に対する出力を作成する。

2.3 精度評価指標

2.3.1 R^2 (決定係数)

回帰モデルの応答変数の分散のうち、予測変数で説明できる割合を示す指標である。 R^2 値が 1 に近いほど、モデルのデータセットへの適合性が高いことを意味する。これは、Eq.(6)で計算される。

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (6)$$

ここで、 \hat{y} は y の予測値で、 \bar{y} は y の平均値を表す。

2.3.2 Root Mean Squared Error (RMSE)

RMSE は Mean Squared Error の平方根である。残差の標準偏差を測定する。RMSE が 0 に近いほど見積もられる予測誤差が小さい、すなわち予測精度が高いことを表す。Eq.(7)で計算される。

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (7)$$

2.3.3 Mean Absolute Error (MAE)

MAE は Eq.(8)で計算され、RMSE と共に平均化された誤差の大きさを表す。RMSE 同様、MAE も 0 に近いほど予測精度が高いことだ。

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (8)$$

3. 実験

3.1 データセット作成方法

データ準備とは、データを収集、結合、整理、構造化するプロセスであり、その後、データの可視化、分析、機械学習アプリケーションによるデータマイニングとして考えることができる。解決したい問題に対して、正確なデータを投入することが重要である。この研究に使用されるデータセットは、ヤフーファイナンスのウェブサイト (<https://finance.yahoo.com>) から収集された日々の価格値で構成されている。全体のデータ収集期間は、2014年9月17日から2020年7月31日までである。このデータセットには、始値 (opening price)、高値 (high)、安値 (low)、終値 (Closing price) などを含む7つの属性と、公開されている発行済み株式の時価総額が含まれている。

3.2 比較モデル

本研究では、LSTM モデルを評価するために、2つのモデルを使用した。FB Prophets と ARIMAX である。

FB Prophets はオープンソースの時系列予測アルゴリズムで、統計や時系列予測の専門知識がなくても簡単に使えるように Facebook が設計したものである。このツールは、直感的なパラメータを提供し、簡単にチューニングすることができる。

ARIMAX は、特定の変数を追加した ARIMA モデルである。ARIMAX とは、Autoregressive Integrated Moving Average with Exogenous Variables の略で、外生変数を用いた自己回帰和分移動平均のことである。

3.3 結果

BTC の履歴データを用いて、LSTM、ARIMAX、FB Prophets アルゴリズムで得られた結果である。各モデルについて、結果は Table 1 の通りである。最も低い RMSE と MAE、1 に近い R2 を与えるモデルが最良のモデルであると考えられる。この基準に基づけば、特に R2 に着目すると、ビットコインに適用されたすべてのモデルは良いモデルと言えるが、3つのモデルの中では LSTM が最も優れていることが分かった。LSTM モデルの RMSE と MAE は最も低くなっている。このように、LSTM は ARIMAX や FB Prophet と比較して、長期的な依存関係を予測する能力が高いと言える。

Table 1. RSME, MAE and R2 score of LSTM, ARIMAX, FB Prophet Model

Model	RMSE	MAE	R2
LSTM	2073.2762	40.2498	0.9578
ARIMAX	2492.5643	1791.1647	0.9396
FB PROPHET	2469.6159	1796.7453	0.9407

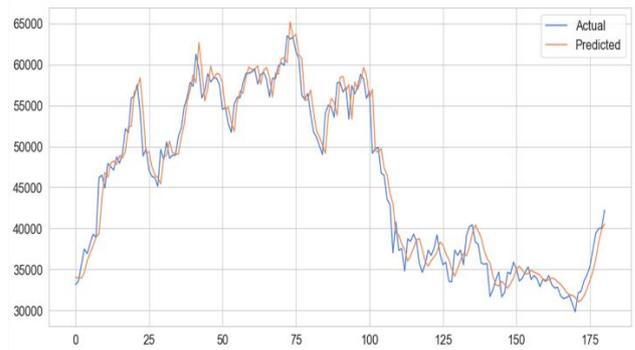


Fig 5. LSTM Model Forecast

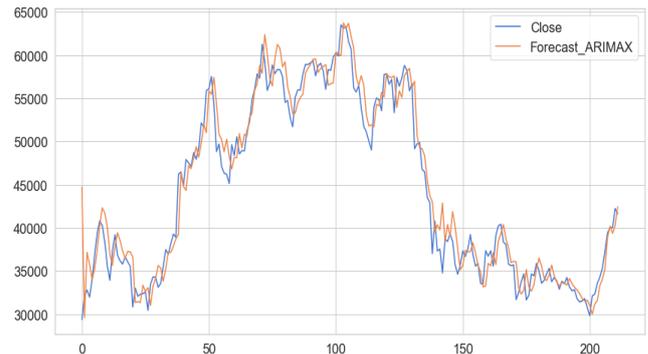


Fig 6. ARIMAX Model Forecast

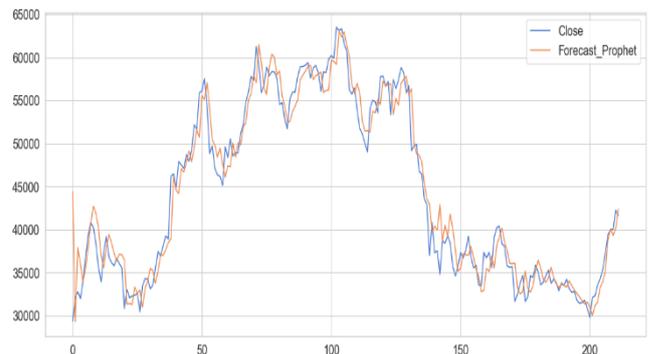


Fig 7. FB Prophet Model Forecast

Fig 5 は、LSTM モデルを用いた BTC の実際の価格と予測値の比較を示している。このグラフは、テストセットにおいて、予測値と実際の価格の差がほとんどないことを示しており、このモデルが最も優れていると考えられる。BTC に対する LSTM モデルの予測モデルの MAE は 40.2498、RMSE は 2073.2762、R2 は 0.9578 であった。データを統計的に分析すると、予測価格は、平均値が 42,721.617 米ドル、最大値が 63,923.227 米ドル、最小値が 32,046.422 米ドルであるのに対し、実際の価格は、平均値が 47,348.606631 米ドル、最大値が 67,566.828 米ドル、最小値が 29,001.7207 米ドルとなった。また、実際の価格と予測価格の平均値の差は 4,626.9896 米ドルであった。

Fig 6 は、ARIMAX モデルを用いた BTC の実際の価格と予測値の比較を示している。BTC に対する ARIMAX モデルの予測モデルの MAE は 1791.1647、RMSE は 2492.5643、

R2は0.9396であった。データを統計的に分析すると、予測価格について、平均値が47,628.006894米ドル、最大値が67,807.6520米ドル、最小値が28,625.0941米ドルであるのに対し、実際の価格は、平均値が47,348.606631米ドル、最大値が67,566.828米ドル、最小値が29,001.7207米ドルとなった。実際の価格と予測価格の平均値の差は、279.400米ドルであった。

Fig 7は、FB Prophets モデルを用いたBTCの実際の価格と予測値の比較を示している、BTCに対するFB Prophets モデルの予測モデルのMAEは1,796.7453、RMSEは2,469.6159、R2は0.9407であった。データを統計的に分析すると、予測価格は、平均値が47,272.2841米ドル、最大値が67,101.5718米ドル、最小値が28,311.1107米ドルであるのに対し、実際の価格は、平均値が47,348.606631米ドル、最大値が67,566.828米ドル、最小値が29,001.7207米ドルとなった。実際の価格と予測価格の平均値の差は、76.322531米ドルであった。

4. まとめ

2008年にビットコインが登場して以来、暗号通貨の分野では圧倒的な存在感を示している。世界中に何百万人ものユーザーがいるだけでなく、暗号通貨の価格を予測することは様々な研究が行われており、そこから投資へのヒントなどを求める動きが盛んである。

本研究では、ヤフーフाइナンスのビットコイン価格を用いて、30タイムステップ、合計6年以上の情報を取得し、次のタイムステップの終値を予測した。その際、LSTM・ARIMAX・FB Prophetの3つの手法を実施した。RMSEとMAEで評価した結果、LSTMが全ての手法の中で最も優れた価格予測性能を持っていることが分かった。また、価格の動きとモデルの精度には関連性があった。変動期よりも通常期の方が予測の精度が高いことが明らかになった。現在直面している大きな課題は、予測モデルの効率化である。本研究では、LSTMにおいて解析を行うと結果が出るまでに1日程度かかることもあった。そのため、現在の段階ではデータセットにいくつかの特徴を含めるだけで、十分な情報が得られると仮定しているが、実際にはビットコインはもっと複雑で、他の市場や他の要因との相関関係があるかもしれない。今後の課題としては、より効率的な代替アルゴリズム探索や現在のデータセットに市場インデックス

などの追加情報を含めることで、パフォーマンスを向上させることである。

参考文献

1. [Emerging Tech Bitcoin Crypto. pdf \(ussc.gov\)](#)
Satoshi Nakamoto. “Bitcoin: A Peer-to-Peer Electronic Cash System”
2. https://www.researchgate.net/publication/328407542_Big-Crypto_Big_Data_Blockchain_and_Cryptocurrency
Hossein Hassani; Xu Huang; Emmanuel Sirimal Silva. “Big-Crypto: Big Data, Blockchain and Cryptocurrency”. 2018/10.
3. https://www.researchgate.net/publication/326837070_Comparative_Performance_of_Machine_Learning_Algorithms_for_Cryptocurrency_Forecasting
Nor Azizah Hitam, Amelia Ritahani Ismail. “Comparative Performance of Machine Learning Algorithms for Cryptocurrency Forecasting”. 2018/7
4. https://www.researchgate.net/publication/339092042_A_LS_TM-Method_for_Bitcoin_Price_Prediction_A_Case_Study_Yahoo_Finance_Stock_Market
Ferdiansyah; Siti Hajar Othman; Raja Zahilah Raja Mohd Radzi; Deris Stiawan. “A LSTM-Method for Bitcoin Price Prediction: A Case Study Yahoo Finance Stock Market.” 2–3 October 2019.
5. https://www.researchgate.net/publication/349367498_Predicting_bitcoin_price_using_lstm_And_Compare_its_predictability_with_arima_model_1
Anshul Saxena. “Predicting bitcoin price using LSTM And Compare its predictability with ARIMA model”, 1 February 2018.
6. <https://research.facebook.com/blog/2017/02/prophet-forecasting-at-scale/> - Ben Letham, Sean J. Talor. “Prophet: forecasting at scale”, 23 February 2017.
7. https://pluralsight2.imgix.net/guides/8a8ac7c1-8bac-4e89-ace8-9e28813ab635_3.JPG