Bitcoin Price Forecasting based on CNN-LSTM Hybrid Neural Network Model

Zhu Chen #1, Yifan Kou #2, Yitao Shi #3

Abstract

In this study, we trained a CNN-LSTM hybrid neural network model to predict the closing price of Bitcoin (BTC) using previous 5-days trading information. We first applied a GoogLeNet structured 1D Convolutional Neural Network (CNN) to aggregate feature patterns embedded in the original time series data from T-5 to T-1. Then, a Long Short-Term Memory (LSTM) model takes the concatenated feature vectors from the GoogLeNet to predict the Bitcoin closing price in T. Comparing the model performance with other single structure neural networks, such as LSTM, GRU, and CNN, the CNN-LSTM hybrid model effectively improve the accuracy of price prediction as measured by the percentage prediction error of the closing price. The study can potentially help investors' decision-making in the cryptocurrency market.

Introduction

Recent years, the potential for high returns in the Bitcoin market attracts great attention of financial investors. However, the Bitcoin price changes are on a greater scale than that of the traditional financial assets such as stock and bonds, which entails a high risk. Therefore, the study of effective price forecasting methods is of great practical importance to investors, researchers, and policymakers around the world.Although several online platforms make available technical analysis tools that allow the bitcoin speculators to identify trends and market sentiment, most tools are set to predict BitCoin price movement direction rather than the actual price, and hence only provides limited insights.

As such, our study focuses on predicting the future BitCoin closing price, hoping to provide BitCoin investor stronger reference during decision making process. Considering BitCoin's nature as a financial asset, we hope to identify the model that predict the future closing price with the smallest percentage error. We defined our loss function as Mean Absolute Percentage Error (MAPE) and set multiple regression model as our baseline model. Several neural network models including the Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), GoogLeNet Structured Convolutional Neural Network (CNN), and a CNN-LSTM hybrid model are trained to forecast the future BitCoin closing price.

Related Literature

To date, scholars home and abroad has conducted many insightful and in-depth discussion on the trend or price prediction on BitCoin.

Although different conclusion has been drawn, in most of the studies, ensemble or hybrid approache achieves better result than single structured model.Guo, Lei, Ye and Fang proposed an Multiple-scaled Residual Block(MRC)-LSTM model that beats the single structured algorithm(Guo, Lei, Ye and Fang 2021). Cocco, Tonelli, and Marchesi compared the performance of the single stage frameworks, formed by an NN (BNN, FFNN, or LSTMNN), highlighting that the two stages hybrid frameworks perform better than the one stage frameworks(Cocco, Luisanna and Tonelli, 2019).

Mallqui and Fernades employs different RNN, Tree, SVM and ensemble algorithms and found that the RNN-Tree classier performs the best in predicting Bitcoin price trend(Mallqui and Fernandes, 2019). Yet there are also studies claim that it is unnecessary to any hybrid approach or even neural network.Uras, Marchesi, and Tonelli suggests linear regression and LSTM can achieve similar result.(Uras, Marchesi, and Tonelli 2020). And Li, Dai studied the Bitcoin price forecasting using Back Propagation (BP) neural network, CNN, LSTM and a CNN-LSTM hybrid model, where the CNN to provide a result with lowest MAPE loss (Li, Dai 2019).

Dataset and Features

Raw Data.

We extracted the raw data-set from CoinMarketCap.com, which is one of the most referenced and trusted source for cryptocurrencies' trading infromation. The raw data on BitCoin price follows a chronological order with 3042 data points. The record starts from July 2, which is the first-ever record of BitCoin price available on CoinMarketCap.com, and ends on October 29, 2021, which is the starting point of this study. The raw data contain 7 variables containing market trading information and a summary table is shown in Table 1.

Table 1: Raw Data Description

Variables	Description
Open	Earliest price record in USD of BTC during the time stamp range
Close	Latest price record in USD of BTC during the time stamp range
High	Highest price record in USD of BTC during the time stamp range
Low	Lowest price record in USD of BTC during the time stamp range
Volume	Trade volume during the time stamp range
MarketCap	Market capital record as of the end of the time stamp
Date	Ten digits integer that represents time stamp

Feature Selection.

Market data

We used the data extracted from CoinMarket.com. as shown above. Within the total 7 features, the Date feature is taken as the timestamp, and all other features except for the Market Capital feature are selected as our model input. The Market Capital feature is excluded because it can be directly calculated using the closing price and volume and hence provides redundant information. Figure 1 presents a visualization of BitCoin's historical Closing Price and a summary table on the 5 market trading features that we selected - Open, High, Low, Close, and Volume.

Figure 1. Data Summary

Input Data Summary

	Open	High	Low	Close	Volume
count	3,042	3,042	3,042	3,042	3,042
mean	8,328	8,566	8,076	8,348	11,987,620,000
std	13,539	13,945	13,101	13,573	19,202,890,000
min	69	75	66	68	0
25%	453	458	444	453	38,415,820
50%	3,493	3,603	3,444	3,512	1,752,580,000
75%	9,200	9,358	9,032	9,206	18,884,710,000
max	66,002	66,930	63,611	65,993	350,967,900,000

The line chart in Figure 1 visualizes the BitCoin's historical Closing Price, which is our prediction target in interest for timestamp T. The closing price was relatively stable at the early stage but fluctuated significantly from the start of 2018. The BitCoin closing price reached an all-time high in 2021 when the closing price exceeded 65,000 USD in February and April.The table in Figure 1 shows the summary of the 5 input variables we used. After selecting the feature, the 5 features are consistently applied to all models throughout the study to keep different models comparable to each other.
Given the fact that only 3042 data points are available, despite the target closing price presents a different shape as shown in the line chart, we included all 3042 data points for the study. We divided our data into training dataset, validation dataset, and testing dataset using the ratio of 6:2:2 along the chronological order. The 2 dash lines in the Figure 1 Line chart shows the separation of the 3 datasets.
Macro-economic factors
This study doesn't incorporate macro-economic variables as both the empirical result and the literature suggest that macroeconomic data doesn't help enhancing the price prediction.
Initially the SP 500 index is selected as a touchstone for incorporating macro-macroeconomic data to our model as it is the most widely used proxy to indicate the financial market dynamics. However, the testing error become larger when SP 500 data is incorporated. To solve this counter intuitive result, we reviewed several scholar research and according to Kapar's paper,published in WILEY's The World Economy Journal, ever since 2018, the "financial variables such as SP 500 index and the gold spot price are no longer statistically significant" (Kapar and Olmo, 2020).
Although such empirical result and research might be counter-intuitive at first, it is actually understandable if we recall that financial assets are usually highly efficiently priced — Investors react promptly to the macro-economic environment changes. This is to say, the macro-economic factors are possibly already "priced" into the asset and can be reflected by the market trading data, such as the price change within a trading day, investors' trading volume, the momentum in price during the time series, etc.

Data Preprocessing.

Missing Data
As BitCoin is a digital financial assets with markets opening 24/7 all year around we have full record over the period of study and missing data handling is not required for this project.
Normalization
We applied the Sklearn MinMaxscaler to normalize the date for neural network models. Data normalization can effectively prevent the influence of different dimensions scale and accelerate the converging speed in training process.

$$X' = \frac{X - Xmin}{Xmax - Xmin}$$

Methodology

Loss Function.

We defined Mean Absolute Percentage Error (MAPE) as the loss function for this study.

$$\text{MAPE} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \left| \frac{y - \hat{y}}{y} \right|$$

MAPE is selected among other metrics for model evaluation mainly for 2 folds:
1) MAPE is a very popularly used loss function in financial asset pricing practices because compared to other loss functions such as MSE, or MAE it's more interpretable and has better practical meaning.
2) Relative errors do not depend on the scale of the dependent variable, this measure enables comparing forecast accuracy between differently scaled time-series data.

Baseline Model.

Multiple Linear Regression
We applied multiple linear regression model — the classic model for price prediction as our base line model. The model takes the close price from T-5 to T-1, and T-1 volume, open, high, low price as features.
The model reports a 11.22% MAPE which indicate its poor prediction power. By examining the correlation table among features, we consider that the model performs poorly potentially because it suffers from multicollinearity problem since high correlation among features are observed.

Single-structured Neural Network Model.

Hoping to enhance the prediction performance, Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) with more sophisticated structure and better capacities in capturing non-linearity patterns. For RNN models, considering that Vanilla RNN is poor at long-term memory, we explored GRU and LSTM models. And for RNN models, 1d Conv layers are structurally combined to extract data in the time series input.
GRU
Our GRU model takes the three-dimensional input in the shape of [batch size, time steps, number of features]=[64, 5, 5]. Previous five-day information with open, high, low, close, and volume are selected as inputs to predict the close price at T. The GRU model has 128 hidden dimensions and connects to 1 fully connected layers for prediction.
We finalized our model using epoch = 300, learning rate = 1e-5, Adam optimizer, and the GRU model reports a 5.2851% MAPE loss.
LSTM
Our LSTM model takes the input in the shape of [batch size, time steps, number of features]=[64, 5, 5]. Similar to the GRU model, previous five-day information with open, high, low, close, and volume are selected as inputs to predict the close price at next timestamp. We developed a stacked LSTM model with 4 LSTM layers each with 64 hidden neurons.
The model is finalized using epoch = 200, learning rate = 5e-3, Adam optimizer, and the LSTM model reports a 4.0828% MAPE loss.
CNN
The CNN model consists two parts: 1 vanilla CNN model, and a CNN model in GoogleNet structure. The vanilla CNN layer has 16 Conv1D filters which expands the original five dimensional inputs to 16 dimensional features. The data is passed to the GoogLeNet structure. The GoogLeNet is consisted of 3 convolutional layer branches with convolutional kernels of size 1, 3, and 5 to slide the sequences, and a jump connection to further aggregate historical information efficiently. By using CNN with different kernal sizes, we are able to simultaneously summarize the information with 1, 3 and 5-day period to the fully connected layers. ReLu function has been used throughout all CNN layers as activation function.
The model is finalized using epoch = 300, learning rate = 1e-4, Adam optimizer, and the CNN model reports a 4.6337% MAPE loss.

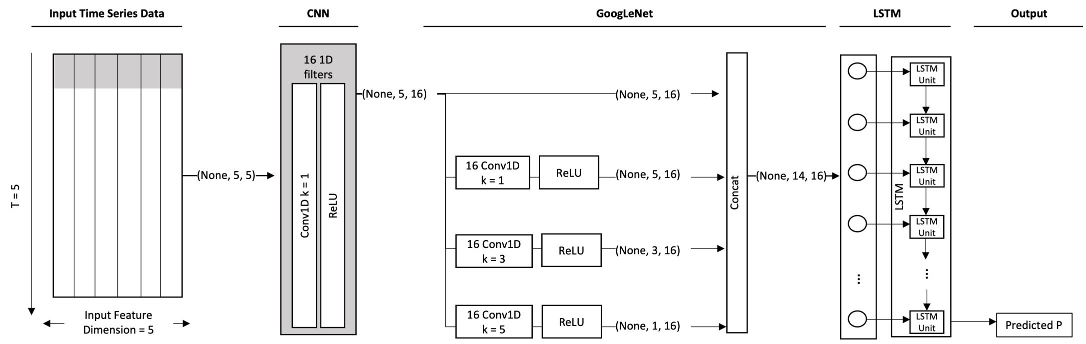
Hybrid Neural Network Model.

Although all neural networks, GRU, LSTM, and CNN outperforms the base line model - the Multiple Linear Regression,the models' MAPE are still a bit too high to provide useful insights for BitCoin investors.

CNN-LSTM Hybrid Model

Inspired by the ideal of model ensemble and hybrid approach mentioned by previous researches, and based on the current trained model, we proposed a CNN-LSTM hybrid neural network model for further prediction accuracy improvements. The CNN-LSTM Model Structure is summarized in Figure 2.

Figure 2. CNN-LSTM Hybrid Neural Network



The origional time series data with length 5 (from T-4 to T-1) and feature dimension 5 are first passed to a CNN layer with 16 1D Conv filters where 5 dimensions are expanded to 16 dimensional feature space. Then it is passed to the GoogLeNet with 4 branches, with each branch containing: 1) a jump connection from the previous CNN layer ; 2) 16 Conv1D Conv layer with k = 1; 3) 16 Conv1D Conv layer with k = 3; 4)16 Conv1D Conv layer with k = 5. The motivation of designing the first branch is to reinforce the learning in the previous CNN layer, while the other Conv 1D branches with different kernal sizes are meant to simultaneously extract the feature trends under different time period 1 day, 3 days and 5 days.

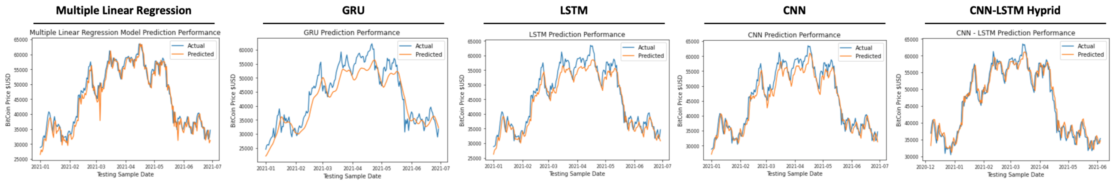
The LSTM is followed by a concatenation layer where the 4 branches from GoogLeNet are concatenated into a (None, 14, 16) data shape. The LSTM containing four hidden layers, each containing 64 neurons is used to further extract the sequential patterns to predict the BitCoin's closing price in T.

We finalized our Hybrid model using epoch = 400, learning rate = 5e-4, and the model report a MAPE of 3.3296%, which is the lowest MAPE among all models we built.

Results and Discussion

For model comparison, we performed price prediction using all 5 models on the same 2021H1 data set (from January 1st 2021 to June 30th 2021). Figure 3 visualizes the actual and predicted price line for each of the model.

Figure 3. Model Comparison: Actual and Prediction Results on BitCoin Closing Price 2021 H1



The corresponding MAPE is summarized in Table 2, where we can observe that the hybrid model has the best performance with the lowest MAPE. Our empirical experiment has testified that the hybrid model indeed outperforms the single structured models, since it potentially aggregates the advantages of both CNN and LSTM. The hybrid model can first extract the trends and hidden interactions of the data with different period through GoogleNet structured CNN models and later uses LSTM with a memory hidden cell to further mines information in the time series.
As the hybrid model shows its promising predictive ability over other models, we are confident that it has huge potential in providing useful insights for investors to make investment decision with further improvements.

Table 2: Model Prediction Performance Summary

Models	MAPE(%)
Multiple Linear Regression	11.2203
GRU	5.2851
LSTM	4.0828
CNN	4.6337
Hybrid NN	3.3296

Yet, we also see the 2 shortcomings of this model:
1. Lack of analysis on market sentiment: Although macro-economic data shown to be statistically insignificant as discussed in the feature selection section, market sentiment has been proved to be a vital factor for Cryptocurrency price. However, in current study scale, we were not able to include sentiment analysis.
2. Limited generalizability: As we apply the model to Ethereum, which ranked 2nd for its market capital, the model performed poorly, indicating the model has limited generalizability.

Conclusion and Future Work

Conclusion
In this study, several machine learning models are comparatively implemented to forecast Bitcoin price, including Linear Regression, GRU, LSTM, CNN, and a CNN-LSTM hybrid model.The CNN-LSTM hybrid model is testified to outperform the baseline lienear regression model and all other single-structured neural networks.
Future work
Our future work aims to further enhance the model's prediction accuracy. Firstly, we hope to train our model using larger available samples. As BitCoin is a financial asset with limited history, the available data for current study is still limited which potentially limits the model performance. Secondly, we hope to conduct more thorough optimization with a better model performance tracking and evaluation framework to further strengthen the visibility and persuasiveness of hyper parameter selection. Lastly, our future research work also aims to find and incorporate reasonable market sentiment proxy data to the model to further include market reaction on cryptocurrency-related news and government policies to improve prediction accuracy. We believe by adding training data, strengthening hyper-parameter fine tuning and incorporating market sentiment data could contribute to a better model performance.

References

[1] A. Greaves and B. Au, "Using the bitcoin transaction graph to predict the price of bitcoin," 2015.

[2] D. C. Mallqui and R. A. Fernandes, "Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques," *Applied Soft Computing*, vol. 75, pp. 596–606, 2019.

[3] F. Ferdiansyah, S. H. Othman, R. Zahilah Raja Md Radzi, D. Stiawan, Y. Sazaki, and U. Ependi, "A lstm-method for bitcoin price prediction: A case study yahoo finance stock market," pp. 206–210, 2019.

[4] J. Sen, S. Mehtab, and A. Dutta, "Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models," 8 2021.

[5] B. Kapar and J. Olmo, "Analysis of Bitcoin prices using market and sentiment variables," 7 2020.

[6] L. Cocco, R. Tonelli, and M. Marchesi, "An agent-based artificial market model for studying the bitcoin trading," *IEEE Access*, vol. 7, pp. 42 908–42 920, 2019.

[7] M. Rauchs, A. Blandin, K. Klein, G. C. Pieters, M. Reca-natini, and B. Zheng, "2nd global cryptoasset benchmarking study," 2018.

[8] Q. Guo, S. Lei, Q. Ye, and Z. Fang, "Mrc-lstm: A hybrid approach of multi-scale residual cnn and lstm to predict bitcoin price," 05 2021.

[9] R. Guerraoui, "Genuine atomic multicast in asynchronous distributed systems," *Theoretical Computer Science*, vol. 254, pp. 297–316, 2001.

[10] S. McNally, J. Roche, and S. Caton, "Predicting the price of bitcoin using machine learning," pp. 339–343, 2018.

[11] W. D. Yan Li, "Bitcoin price forecasting method based on cnn-lstm hybrid neural network model," 2019.