Neural network approximation precision change analysis on cryptocurrency price prediction

A Misnik[1], S Krutalevich[1], S Prakapenka[1], P Borovykh[2] and M Vasiliev[2]

[1] State Institution of Higher Professional Education "Belarusian-Russian University", Mogilev, Belarus
[2] RoninAI Lab, New York, USA

Abstract. The neural networks are tool for approximation universal series of data, but their precision highly depends on a adequate set of inputs. Cryptocurrencies experience high levels of volatility due to absence of agreed upon pricing methodologies behind its valuation. In this article we analyze approaches of obtaining additional inputs for neural networks and explore their influence on its precision.

1. Introduction

In this study, we will estimate the price of Bitcoin (BTC), using market data and try to analyze improvement of the neural network's precision by adding subsequent factors such as social and time factor.

Study presented in this paper will include two neural networks architectures: Multilayer perceptron (MLP) and Long short-term memory (LSTM) neural networks. MLP is the most common architecture, providing acceptable accuracy in a large number of tasks, and LSTM is considered to be more accurate in the tasks of time series prediction

Results are tested and based on neural networks created by RoninAI Lab. RoninAI uses various neural networks for cryptocurrency rate prediction, lending hands-on data and analysis to this study.

2. Gathering market data

A number of sources mention the possibility of predicting price of an asset using market data. To test prediction accuracy of the market data, we started out by collecting historical minute-based market data from one of the oldest and biggest cryptocurrency exchange – Kraken. Market data inputs are listed in Table 1.

Table 1. Market data parameters.

Input name	Description
DATE AND TIME	Timestamp of given minute
OPEN	Rate at the 0 second of given minute
CLOSE	Rate at the 59 second of given minute
HIGH	Highest rate during given minute
LOW	Lowest rate during given minute
VOLUME FROM	Lowest volume of sales per second
VOLUME TO	Highest volume of sales per second

3. Training on market data

After training MLP and LSTM neural networks using market data as inputs and next-minute price of BTC as output, models returned results presented on Figure 1.

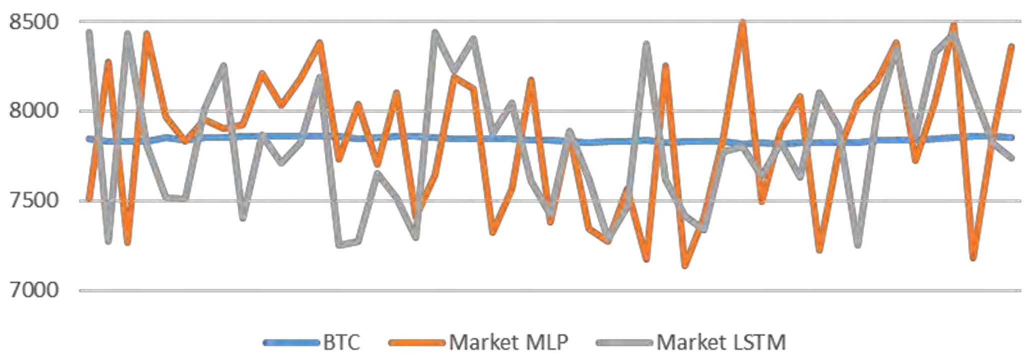

Figure 1. .Training results on market data.

The precision of predictions by both neural networks came out sufficiently low. LSTM showed slightly better accuracy.

4. Extending inputs with time factor

Expert time factor is a factor representing activity level of exchanges. Retraining neural networks with additional input of the time factor produced results shown on Figure 2.

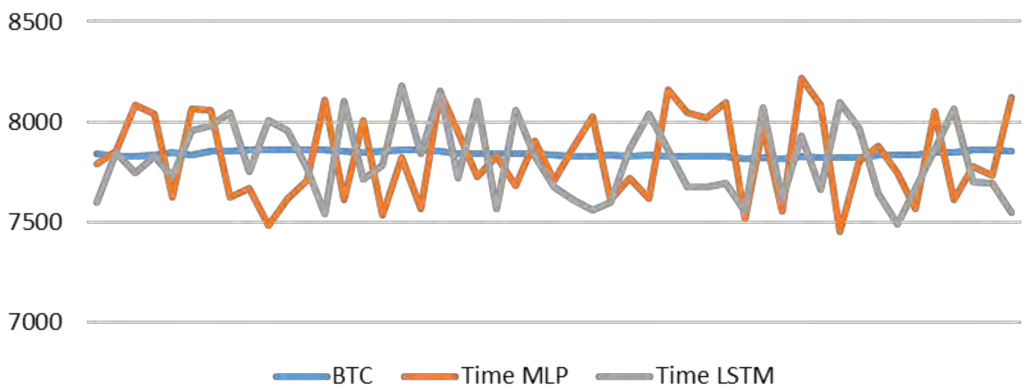

Figure 2. . Training results on market data and time factor.

Despite sufficiently low level of precision, predictions produced by both MLP and LSTM turned out to be much better compared to predictions based on market data alone.

5. Extending inputs with social factor

It is estimated that the majority of market players in the cryptocurrency space are retail investors making their investment and trading decisions based on the condition of visual charts. An example of such chart is shown on Figure 3.

We wanted test the importance of visual charts on predictive power of neural networks.

Let's consider an algorithm that allows us to identify some complex indicators of a numerical series, based on which we can predict trader's subjective assessment on appearance of chart.

In general, any series of data can be considered as a sum of linear and harmonic components.

The purpose of further research is to investigate the algorithm for isolating these components and their normalization.

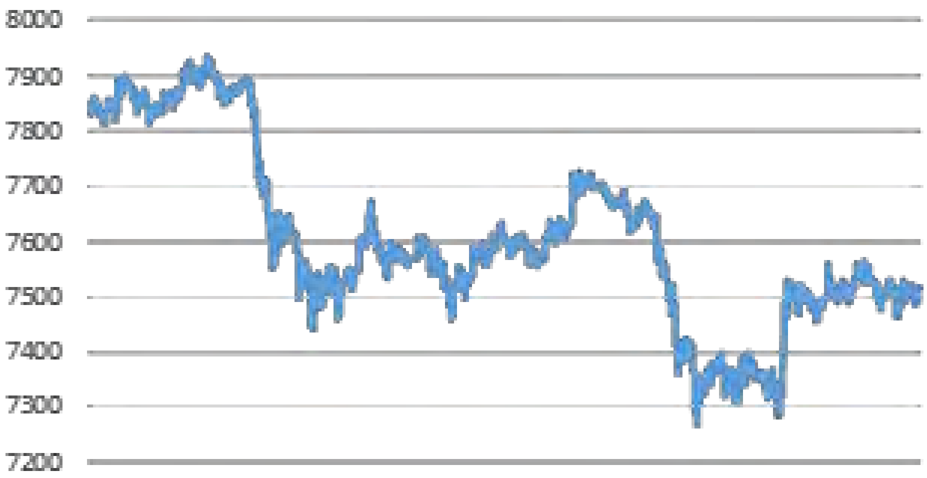

Figure 3. . BTC price chart.

The proposed algorithm assumes the following stages.

1. Definition of the minimal element E_{min} of the series E.
2. Carrying out the subtraction operation $E' = Ei - E_{min}$.
3. Approximation of the series E' by a polynomial

$$E_l = a(0) + a(1) * n, \quad (1)$$

where a(0), a(1) are the approximation coefficients; n-discrete values of the time axis.

4. Calculation of the coefficient of relative change in the linear component over the period T by the formula:

$$E_L = a(1) * T * 100 / E_{min}, \quad (2)$$

The value of E_L is relative and does not depend on the absolute value of the series. If the quantity $E_L > 0$, the linear component increases.

5. To estimate the harmonic component of the series, we perform a Fourier transform (FT) for the series E'.
6. Determine the moduli of the oscillation amplitudes in the frequency domain A(w). Carry out filtering of frequencies according to the amplitude values.

To analyze the efficiency of the proposed algorithm, we used MATLAB environment.

Step 1. The linear component is constant. Harmonic component is absent. The results are shown on Figure 4. The value of the coefficient EL displayed on the second chart. In this case, EL = 0.

Step 2. The linear component grows. Harmonic component is absent. The results are shown in Figure 5. The coefficient EL = 20%. The spectrum modules have values in the low-frequency range (1-2 Hz.). It should be remembered that the main frequency band for the module of the real sequence lies in the interval $0 < k <= N / 2-1$.

Therefore, frequencies above 15 Hz in our example should be ignored. If it is necessary to analyze higher frequencies, we will increase N - the number of sampling points.

Step 3. The linear component decreases. Harmonic component is absent. The results are shown in Figure 6.

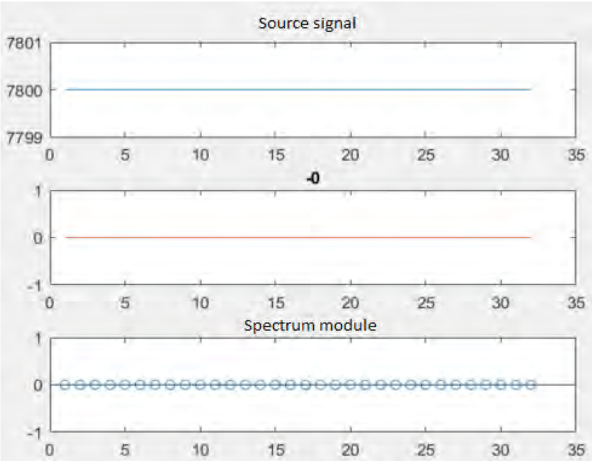

Figure 4. Analysis step 1.

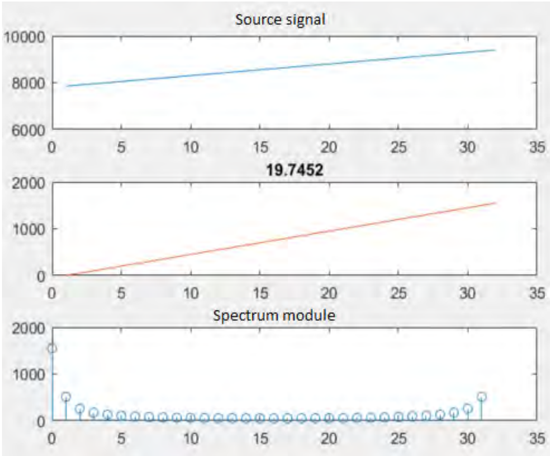

Figure 5. Analysis step 2.

Step 4. The linear component is absent. Harmonic component is present at a frequency of 3 Hz - amplitude 20. The results are shown in Figure 7.

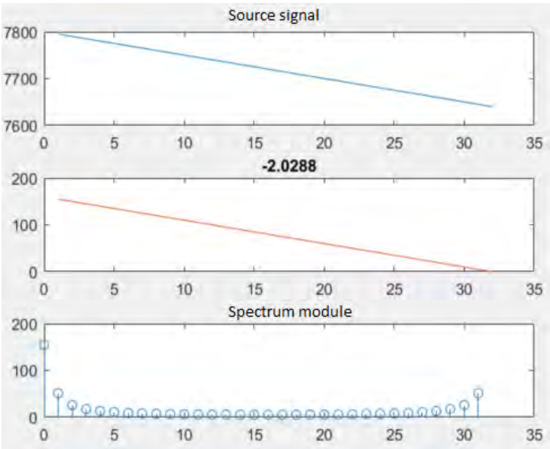

Figure 6. Analysis step 3.

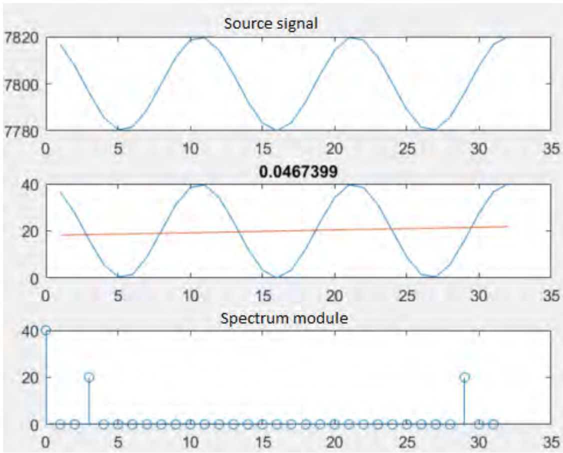

Figure 7. Analysis step 4.

Step 5. The linear component increases. Harmonic component is present at a frequency of 3 Hz - an amplitude of 400.

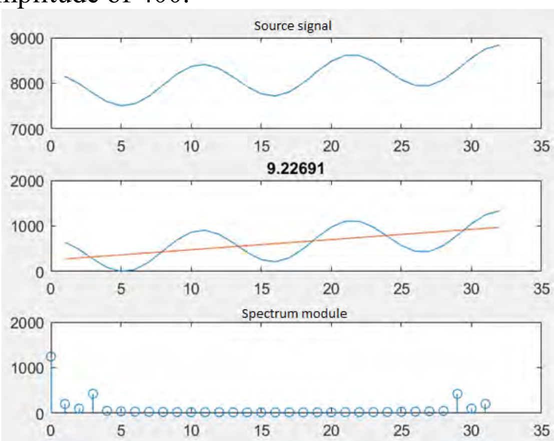

Figure 8. Analysis step 5.

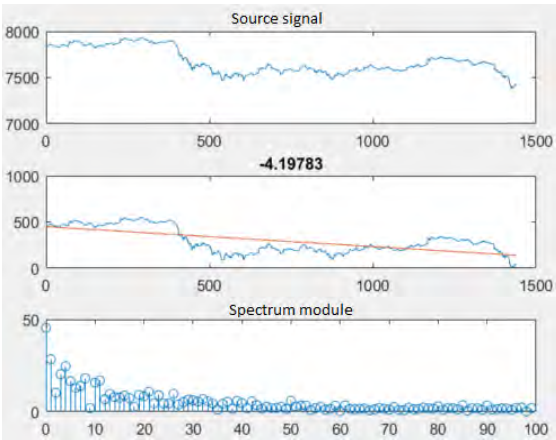

Figure 9. Analysis - application.

We apply the proposed algorithm to the data presented in Figure 3. The result is shown in Figure 9. The figure shows that the linear component predicts a downtrend with a rate of 4.1% per period. Modules of amplitudes of harmonic components have influence in frequencies up to 10 Hz.

The proposed algorithm distinguishes the linear and harmonic components of the numerical series.

The proposed coefficient E_L - can act as a measure of the trend of changes in the values of a numerical series.

Moduli of the amplitude of the oscillations in the frequency domain after the Fourier transformation can act as a measure of the estimation of the vibrational component of a series of data.

In our case we've extended inputs of neural networks with 4 new parameters: coefficient E_L, and 3 harmonic component ranges: 0 to 10Hz, 10 to 30Hz, 30Hz and higher.

After calculating aforementioned parameters for each row of our dataset MLP and LSTM networks have been retrained. The results of training are shown on Figure 10.

The precision of predictions by both MLP and LSTM networks appeared to be much better compared to previous results.

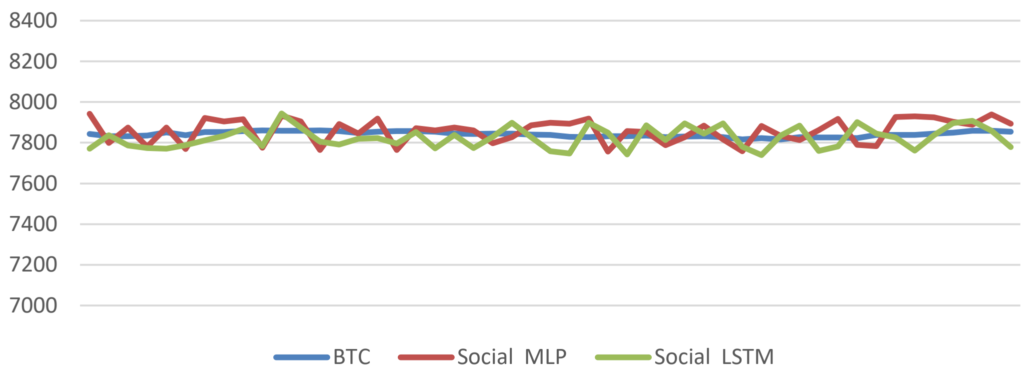

Figure 10. Training results on market data, time factor and social factors.

6. Extending inputs with social networks factor

Recently social networks became a powerful phenomena influencing the opinions of people on various issues.

To make our predictions more accurate we attempted to use data obtained from social networks.

We have extended our inputs with parameters extracted from social networks on BTC-related topics. Results of neural networks training are shown on Figure 11.

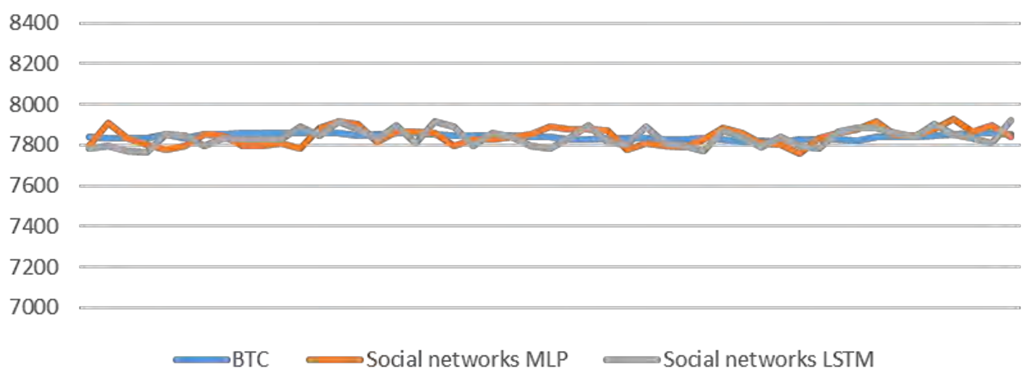

Figure 11. Training results on market data, time factor, social factors and social networks factors.

The precision of predictions by both MLP and LSTM networks became even closer to real price of BTC.

7. Results

As shown, the extension of market data inputs with additional relevant parameters can significantly increase the precision of predictions.

The method, described in chapter V can be used on different datasets, to provide additional input parameters.

In our experiments LSTM networks have shown slightly better precision, then MLP.

8. References

[1] S. Haykin, Neural Networks and Learning Machines (3rd Edition), Prentice Hall, 2009

[2] Tax, N.; Verenich, I.; La Rosa, M.; Dumas, M. (2017). "Predictive Business Process Monitoring with LSTM neural networks". Proceedings of the International Conference on Advanced Information Systems Engineering (CAiSE): pp 477–492. arXiv:1612.02130 Freely accessible. doi:10.1007/978-3-319-59536-8_30.

[3] A. Misnik Neural network modeling of heat supply systems // "Neurocomputers: development, application" - 2012. №2. - pp 62-73.

[4] V. Borisov, A. Misnik Combined neural network modeling for the operational management of complex systems / / Information technology. - 2012. - No. 7. – pp 69-72.

[5] Wierstra, Daan; Schmidhuber, J.; Gomez, F. J. (2005). "Evolino: Hybrid Neuroevolution/Optimal Linear Search for Sequence Learning". Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI), Edinburgh: pp 853–858.

[6] Stein, Elias; Shakarchi, Rami (2003), Fourier Analysis: An introduction, Princeton University Press, ISBN 978-0-691-11384-5.

[7] Rahman, Matiur (2011), Applications of Fourier Transforms to Generalized Functions, WIT Press, ISBN 978-1-84564-564-9.

[8] Abarbanel, Henry (Nov 25, 1997). Analysis of Observed Chaotic Data. New York: Springer. ISBN 978-0387983721.

[9] Tang, Jiliang; Chang, Shiyu; Aggarwal, Charu; Liu, Huan (2015). "Negative Link Prediction in Social Media" (PDF). In Proceedings ofACM International Conference on Web Search and Data Mining.