Bitcoin Price Prediction Using Ensembles of Neural Networks

Edwin Sin
School of Physical and Mathematical Sciences
Nanyang Technological University
Singapore

Lipo Wang
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore

Abstract—**This paper explores the relationship between the features of Bitcoin and the next day change in the price of Bitcoin using an Artificial Neural Network ensemble approach called Genetic Algorithm based Selective Neural Network Ensemble, constructed using Multi-Layered Perceptron as the base model for each of the neural network in the ensemble.**

To better understand the practicality and its effectiveness in real-world application, the ensemble was used to predict the next day direction of the price of Bitcoin given a set of approximately 200 features of the cryptocurrency over a span of 2 years.

Over a span of 50 days, a trading strategy based on the ensemble was compared against a "previous day trend following" trading strategy through back-testing. The former trading strategy generated almost 85% returns, outperforming the "previous day trend following" trading strategy which produced an approximate 38% returns and a trading strategy that follows the single, best MLP model in the ensemble that generated approximately 53% in returns.

Keywords-Artificial Neural Networks, Ensemble, Bitcoin, Prediction, Genetic Algorithm

I. INTRODUCTION

A. Background

The digital cryptocurrency, Bitcoin, runs on an online decentralized network with no dependency on any government or legal entity as it relies heavily on peer-to-peer networking and cryptography to maintain its integrity [1]. This "trust-less" system facilitates ease of transactions of goods and services with low transaction costs associated [2].

As the market is relatively new, existing works related to forecasting in this market is fairly limited. One study showed that Google Trends data and volume of tweets related to Bitcoin on Twitter have positive correlation with Bitcoin's price and hence may be able to predict the fluctuations in price of Bitcoin [3]. In another study, Bayesian Regression, a binary classification algorithm, was used to predict price variation in Bitcoin and the prediction gave almost 200% returns in less than 60 days when used with a trading strategy [4]. The study concluded that there may be 'information' in Bitcoin's historical data that can help predict future price variations.

Lastly, a recent study on the price formation of Bitcoin revealed that macroeconomic and financial indicators tend to have a relatively small effect on Bitcoin prices in the short run. Instead, market forces that affect Bitcoin's supply and demand, in particular, the demand side variables such as transaction volume, showed a bigger effect on the price of Bitcoin [5].

B. Objective

In this paper, the objective is to understand how features of Bitcoin (such as transaction volume, cost per transaction) can affect the next day change in price level of Bitcoin through the use of an Artificial Neural Network (ANN) [6] ensemble approach called Genetic Algorithm based Selective Neural Network Ensemble (GASEN) [7]. The ensemble will be used to solve a binary classification problem: the next day change in the direction of the price of Bitcoin.

To better understand and evaluate its effectiveness, back-testing was done to see how a trading strategy based on the results of the ensemble can compare against a "previous day trend following" trading strategy as well as a trading strategy that follows the single, best MLP model in the ensemble.

II. LITERATURE REVIEW

A. Artificial Neural Network

ANN models are one of the most popular forecasting tools that has been widely used in various industries, such as engineering, finance and healthcare, due to its capability in performing complex tasks, such as classification, pattern recognitions and predictions, with great accuracies [8][9].

ANN success in these areas is largely attributed to some of its key defining features as these models are nonlinear models and are powerful universal functional approximators that can approximate any continuous functions to a desired level of accuracy [10]. As such, ANN models have been seen as a viable alternative to the traditional statistical models for forecasting problems [18-27]. In one study, a literature review of comparative studies between ANN and traditional statistical models has found that out of 96 studies, only 18% of the cases concluded that traditional methods have outperformed ANN while ANN models have either performed well or outperformed in 72% of the cases [11].

Hence, it is unsurprising that neural networks have been increasingly gaining more attention in the field of finance, especially for time series forecasting due to the highly non-linear and volatile nature of the financial market. The complex interactions between market-influencing forces and external random processes such as news or any sudden, disruptive changes presents a huge challenge for any fruitful financial forecasting using any quantitative models [12].

B. Ensemble Learning

A neural network ensemble consists of a set of individually trained ANN. The results from each of these individual classifier is then combined to produce a single result, representative of the ensemble. An ensemble of neural network has been shown to have better accuracy and robustness than a single ANN model associated with a significant reduction in generalization error [13].

The ensemble methodology works on the same premise as real world decision making in that having different opinions on a matter is better than one for decision making. This can be achieved in an ensemble by weighing multiple individual classifiers and combining the outputs from each of the classifier to produce a collective output altogether [14].

Most proposed ensemble approaches work by taking all the available ANN to constitute an ensemble. However, studies showed that it may be better to include some, but not all of the ANN to be part of the ensemble. The relationship between the correlation of the individual ANN and the generalization ability of the ANN ensemble showed that the size of an ensemble can be reduced without compromising generalization ability by taking into account only a subset of the ensemble [15].

GASEN was hence proposed as it selectively choose an appropriate subset of the ensemble to form the final ANN ensemble. Its ability to generate a smaller ensemble with low generalization error and lower computation cost compared to the usual ensemble approach such as enumerating and averaging all makes it a strong ensemble approach that can possibly be used to perform complex forecasting tasks [7].

III. METHODOLOGY

A. Data

The dataset used in this paper were retrieved directly from Blockchain.info, one of the most popular Bitcoin wallet, as well as Bitcoinity.org, a comprehensive, data driven platform that covers the key features of Bitcoins on each of the Bitcoin exchange.

From Blockchain.info, a total of 24 time-series data were found. These data include some of the defining features of bitcoin, such as market price, trade volume and transaction fees.

From Bitcoinity.org, a total of 190 time-series data were found as the website contains data of 19 defining features of Bitcoin for 10 of the most popular Bitcoin exchanges such as BitFinex, Huobi or BTCChina. Some of the features for each of these exchanges includes the prices, volume and bid-ask spread for each of the exchange.

Tables 1 and 2 in Appendix outlines the list of features used in our dataset as well as the ones removed. Table 3 outlines the specific exchanges that were tracked on Bitcoinity.

780 days' worth of historical data were retrieved from these websites. 730 days (2 years) of data will constitute the training dataset which will be used to train the ANN ensemble while the last 50 days will constitute the back-testing dataset, which will be used for back-testing to test the trading strategy. The range of the data used for training is from 2nd May 2015 to 30th April 2017 while the range of the data used for back-testing will be from 1st May 2017 to 20th June 2017.

The dataset was pre-processed to remove features that contain empty elements. Out of 222 features that were extracted, 8 were removed. The features that were removed are listed in the Appendix. Then, the inputs were standardized by mapping the mean and standard deviation to 0 and 1 respectively for each row due to the large variability in the input data. This helps in training the ensemble more efficiently and accurately.

The final training dataset has a size of 730x190 while the back-testing dataset has a size of 50x190.

B. Ensemble Model

The ANN ensemble was created with a set of 5 Multi-Layered Perceptron (MLP), all with the same specifications but different number of nodes in the layers. As shown in Fig 1, each MLP model has an input layer with 190 nodes, 2 hidden layers with varying nodes and 1 output layer with 1 node. The number of nodes in the first hidden layer were preset to be in multiples of 5. 1st MLP will have 5, 2nd will have 10 and the 5th will have 25. The number of nodes in the second hidden layer for all MLP will be the floored half of the nodes in the first hidden layer.

Due to the complexity of the ensemble and the computational cost for back-testing, all of the MLP were trained with a supervised learning algorithm called Levenberg-Marquardt (LM) algorithm. The algorithm showed considerable improvement in terms of speed and accuracy when compared with heuristic algorithms for the convergence of ANN for the case of predicting S&P500 price index, such as the standard Gradient Descent, Resilient Backpropagation or Fletcher-Reeves Conjugate Gradient algorithm [16]. The LM algorithm was also considered one of the best training algorithm for classification tasks during a performance comparison between other algorithms [17]. Mean-squared error was used to evaluate the quality of the neural network and the hyperbolic tangent sigmoid function (1) was used as an activation function for all the nodes in the hidden layers and output layer. The number of epochs for each MLP was set to be 30.

$$y = \frac{2}{1 + e^{-2x}} - 1 \tag{1}$$

Lastly, the training dataset with 730 entries was divided into three subsets, training, validation and testing, by randomly determining the indices for each data point. The training set accounts for 70% of the data set while both the validation and test set were 15% each. After training the ensemble, the back-testing dataset will be used to test the network to predict the next day change in Bitcoin's price.

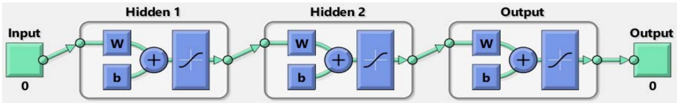

Figure 1 Model of MLP

To forecast the next day change in direction of the price of Bitcoin, each individual MLP is used to solve a binary classification problem, producing a target output of either 0 or 1. 0 implies that the price will drop and 1 implies that the price will go up the next day.

For each of the MLP, the neural network was trained 5 times and the best iteration out of 5, characterized by the highest accuracy during training on the training set, will be chosen to be part of the ensemble.

After training the individual ANN, GASEN was implemented by choosing an appropriate subset of the ensemble from the 5 trained MLPs [7].

GASEN is achieved in 4 steps:

Step 1: Assign each MLP a weight which reflects its importance in the ensemble

Step 2: Optimize the weights of each MLP using Genetic Algorithm by minimizing the generalization error of the ensemble.

Step 3: Select all the MLP models whose corresponding weight is larger than a preset threshold, λ, to constitute the ensemble.

Step 4: Final output of the ensemble is generated via simple averaging from the MLP ensemble.

We set the value λ to be the mean of the optimized weights after Genetic Algorithm rather than a fixed value as initially proposed in the GASEN approach [7]. This ensures that the final ensemble always contain the significant MLP classifiers on every iteration as the ensemble is re-trained before it predicts the next day price direction and a fixed λ may potentially omit certain classifiers.

C. Back-Testing

After training the model, back-testing was done using the next 50 days' dataset to evaluate the performance of the ensemble. A simple trading strategy based on the results of GASEN was compared against a "previous day trend following" trading strategy as a baseline. The "previous day trend following" strategy trades by following the direction of the change in price level of Bitcoin the day before.

Assuming that both the system starts off with USD$10,000 initially and that there are no transaction fees involved in every trade. If the strategies predict that prices were to go up the next day, the system will buy the maximum number of Bitcoins it can at the current price level or not do anything if is unable to buy any more Bitcoins. If the strategies predict that prices were to go down the following day, the system will sell all the Bitcoins at the current price level or not do anything if there is none left.

After each day over the next 50 days in the dataset, the next day's dataset will be appended to the training dataset and the ANN ensemble will be re-trained to take into account the new data and predict the price movement for the day after.

IV. RESULTS

The accuracy of each MLP on the training dataset in the ensemble is different from one another and typically ranges from about 58% to 63%. The size of the final ensemble ranges about 2 to 4 after taking into consideration the computed threshold, λ, for each of the day during back-testing.

The accuracy obtained from the back-testing was 64%, with 32 correct predictions out of 50 days. To better understand the predictions, the total asset value for each of the trading strategy was recorded by monitoring the total value of cash and Bitcoins each day.

Over the span of 50 days, the trading strategy based on GASEN ended with a total asset value of $18,484.43 while the "previous day trend following" strategy had a total asset value of $13,800.64 assuming that both of the strategies started with $10,000 initially. Fig 2 reflects how the total asset value over time changed with this strategy.

The dataset was also tested against the single, best MLP model in the ensemble. The accuracy as reported from back-testing was 60%, with 30 correct predictions out of 50 days. As shown on Fig 2, the trading strategy that relied on the predictions from the single, best MLP model in the ensemble did not perform as well as the ensemble, generating a total asset value of $15,336.58 at the end.

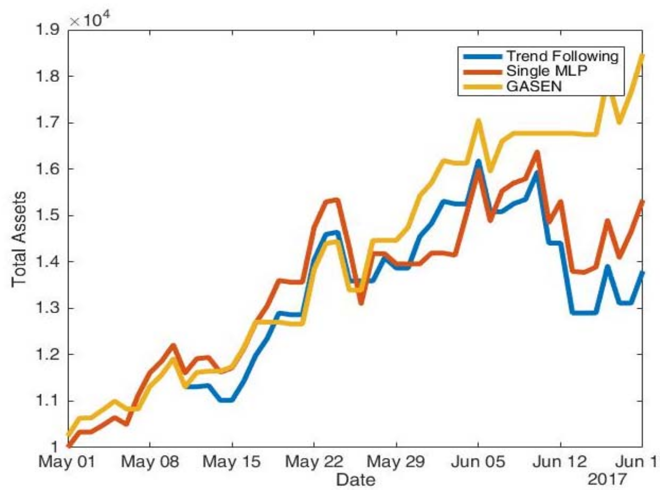

Figure 2 Total Asset Value over Time

V. CONCLUSION

In conclusion, the ensemble method, GASEN, was able to perform well for the classification task with consistent accuracy of around 58% to 63%. With a simple trading strategy, the ensemble was able to obtain promising result in making profit of close to 85% in return. The performance of this prediction indicates that the features of Bitcoin may contain 'information' useful for predicting the next-day change in price level of Bitcoin.

In the future, further studies can be carried out in several directions. First, systematic feature selection should be done, in order to find the optimal subset of input features that leads to the best Bitcoin price prediction accuracy, since irrelevant input features could have introduced noise and reduced the prediction accuracy [22],[28]-[46]. Some important input features could have been missing in the present work. Second, we concentrated on the MLP only. There are a number of other excellent regression techniques that should be investigated, notably, random forests [47], wavelet neural networks [25]-[27], and fuzzy neural networks [48]-[55].

REFERENCES

[1] Satoshi N., "Bitcoin: A peer-to-peer electronic cash system," November 2008.

[2] Grinberg R., "Bitcoin: An innovative alternative digital currency," Hastings Science and Technology Law Journal, vol. 4, pp. 160, December 2010.

[3] Matta M., Lunesu I. and Marchesi M, "Bitcoin spread prediction using social and web search media," UWAP Workshops, June 2015.

[4] Shah D. and K. Z., "Bayesian regression and Bitcoin," Proceedings of 2014 Allerton Conference on Communication, Control, and Computing, October 2014.

[5] Ciaian P., "The economics of Bitcoin price formation," Applied Economic, vol. 48, pp. 1799-1815, November 2015.

[6] L.P. Wang and X.J. Fu, Data Mining with Computational Intelligence, in Springer, 2005.

[7] Z.-H. Zhou., J.-X. Wu., Y.J. Wu and S.-F. Chen, "Genetic Algorithm based Selective Neural Network Ensemble," Proceedings of the 17th International Joint Conference on Artificial Intelligence, vol. 2, pp. 797-802, August 2001.

[8] S. Gupta and L.P. Wang, "Stock Forecasting with Feedforward Neural Networks and Gradual Data Sub-Sampling," Australian Journal of Intelligent Information Processing Systems, vol. 11, pp. 14-17, 2010.

[9] T.X. Zheng, K. Fataliyev, and L. P. Wang, "Wavelet neural networks for stock trading and prediction," SPIE Defense, Security, and Sensing, vol. 8750, April 2013.

[10] Tarsauliya A., Kant S., Rahul K. and Shukla A., "Analysis of artificial neural network for financial time series forecasting" International Journal of Computer Applications, vol. 9, November 2010.

[11] Khashei M. and Bijari M., "An artificial neural network (p,d,q) model for timeseries forecasting," Expert Systems with Applications, vol. 37, pp. 479-489, January 2010.

[12] West D., Dellana S. and Qian J.X., "Neural network ensemble strategies for financial decision applications," Computers & Operations Research, vol. 32, pp. 2543-2559, October 2005.

[13] Paliwal M. and Kumar U., "Neural networks and statistical techniques: A review of applications," Expert Systems with Applications, vol. 36, pp. 2-17, January 2009.

[14] Lior R., "Ensemble-based classifiers," Artificial Intelligence Review, vol. 33, pp. 1-39, February 2010.

[15] Z.-H. Zhou., J.-X. Wu., W.T., "Ensembling Neural Networks: Many could be better than all," Artificial Intelligence, vol. 137, pp. 239-263, May 2002.

[16] Lahmiri S., "A comparative study of backpropagation algorithms in financial prediction," International Journal of Computer Science, Engineering and Applicatns, vol. 1, pp. 1-39, August 2011.

[17] Baptista D., Rodrigues S. and Morgado-Dias F., "Performance comparison of ANN training algorithms for classification," 2013 IEEE 8th International Symposium on Intelligent Signal Processing, vol. 1, pp. 115-120, September 2013.

[18] L. P. Wang, F. F. Chan, Y. Wang, and Q. Chang, "Predicting Public Housing Prices Using Delayed Neural Networks," TENCON 2016 - 2016 IEEE Region 10 Conference, Singapore, 2016.

[19] W.T. Lim, L. P. Wang, Y. Wang, and Q. Chang, "Housing price prediction using neural networks," 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016, pp.518-522. 2016.

[20] A. Mitra and L.P. Wang, "Improving Artificial Neural Network based Stock Forecasting using Fourier De-Noising and Hodrick-Prescott Filter", Proc. 2015 Int. Conf. Information, Communications and Signal Processing (ICICS 2015), 2015.

[21] Y. Fang, K. Fataliyev, L. P. Wang, X. Fu and Y. Wang, "Improving the Genetic-Algorithm-Optimized Wavelet Neural Network Approach to Stock Market Prediction," 2014 International Joint Conference on Neural Networks (IJCNN 2014), pp. 3038-3042, 2014.

[22] Y. He, K. Fataliyev, and L. P. Wang, "Feature Selection for Stock Market Analysis," 20th International Conference on Neural Information Processing (ICONIP2013), LNCS 8227, pp. 737-744, 2013.

[23] G. Dong, K. Fataliyev, L. P. Wang, "One-Step and Multi-Step Ahead Stock Prediction Using Backpropagation Neural Networks", 9th International Conference on Information, Communications and Signal Processing (ICICS 2013), 2013.

[24] M. Zhu and L. P. Wang, "Intelligent trading using support vector regression and multilayer perceptrons optimized with genetic algorithms," The 2010 International Joint Conference on Neural Networks (IJCNN), pp.1-5, 2010.

[25] S. Gupta and L.P. Wang, "Neural networks and wavelet de-noising for stock trading and prediction", in Time Series Analysis, Witold Pedrycz and Shyi Ming Chen (eds.), 2012

[26] K.K. Teo, L.P. Wang, Z.P. Lin, "Wavelet packet multi-layer perceptron for chaotic time series prediction: effects of weight initialization," Computational Science -- ICCS 2001, Proceedings Pt 2, Volume: 2074, pp. 310-317, 2001.

[27] L.P. Wang, K.K. Teo, and Z.P. Lin, "Predicting time series with wavelet packet neural networks", 2001 IEEE International Joint Conference on Neural Networks (IJCNN 2001), Vols 1-4, pp.1593-1597, 2001.

[28] D. Y. Eroglu, K. Kilic, "A novel Hybrid Genetic Local Search Algorithm for feature selection and weighting with an application in strategic decision making in innovation management," Information Sciences, vol.405, pp.18-32, 2017.

[29] Chao-Ton Su and Yu-Hsiang Hsiao, "Multiclass MTS for Simultaneous Feature Selection and Classification," IEEE Trans. Knowledge and Data Engineering, Vol. 21, pp.192-205, 2009.

[30] L.P. Wang, Y.L. Wang, C. Qing, "Feature selection methods for big data bioinformatics: a survey from the search perspective," Methods, vol.111, pp.21-31, 2016

[31] Y. Liu, J.-W. Bi, Zh.-P. Fan, "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms," Expert Systems with Applications, vol.80, pp.323-339, 2017.

[32] L.P. Wang, Nina Zhou, and Feng Chu, "A general wrapper approach to selection of class-dependent features," IEEE Trans. Neural Networks, vol.19, no.7, pp.1267-1278, 2008.

[33] Oh, I.-S., Lee, J.-S., Moon, B.-R.: Hybrid genetic algorithms for feature selection. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 26, pp.1424-1437, 2004.

[34] L.P. Wang, Feng Chu, and Wei Xie, "Accurate cancer classification using expressions of very few genes," IEEE-ACM Trans. Bioinformatics and Computational Biology, vol.4, pp.40-53, 2007.

[35] Bing Liu, Chunru Wan, and L.P. Wang, "An efficient semi-unsupervised gene selection method via spectral biclustering", IEEE Trans. Nano-Bioscience, vol.5, no.2, pp.110-114, June, 2006.

[36] Liu, H.; Dougherty, E.R.; Dy, J.G.; Torkkola, K.; Tuv, E.; Peng, H.; Ding, C.; Long, F.; Berens, M.; Parsons, L.; Zhao, Z.; Yu, L.; Forman, G.; Evolving feature selection, IEEE Intelligent Systems, vol.20, pp.64-76, 2005

[37] Xiuju Fu and L.P. Wang, "Data dimensionality reduction with application to simplifying RBF network structure and improving

classification performance", IEEE Trans. System, Man, Cybern, Part B-Cybernetics, vol.33, no.3, pp. 399-409, 2003.

[38] José-María Montero, Román Mínguez, Gema Fernández-Avilés, "Housing price prediction: parametric versus semi-parametric spatial hedonic models", Journal of Geographical Systems, pp 1-29, 2017.

[39] X.J. Fu and L.P. Wang, "A GA-based novel RBF classifier with class-dependent features", CEC'02: Proceedings of the 2002 Congress on Evolutionary Computation, Vols 1 and 2, pp. 1890-1894, 2002.

[40] X.J. Fu and L. P. Wang, "Rule Extraction from an RBF Classifier Based on Class-Dependent Features", Cec'02: Proceedings of the 2002 Congress on Evolutionary Computation, Vols 1 and 2, pp. 1916-1921, 2002.

[41] Z. Lan, O. Sourina, L.P. Wang, Y. Liu, "Real-time EEG-based Emotion Monitoring Using Stable Features", Visual Computer, vol.32, Issue 3, pp 347-358, March 2016.

[42] Nina Zhou and L. P. Wang, "A modified T-test feature selection method and its application on the Hapmap genotype," Genomics, Proteomics & Bioinformatics, vol.5, pp. 242-249, 2007.

[43] F. Fazayeli, L. P. Wang, and J. Mandziuk, "Feature Selection Based on the Rough Set Theory and Expectation-Maximization Clustering Algorithm", Rough Sets and Current Trends in Computing, LNCS vol. 5306, pp. 272-282, 2008.

[44] Nina Zhou and L. P. Wang, "Class-dependent feature selection for face recognition", Advances in Neuro-Information Processing, Pt Ii, vol.5507, pp.551-558, MAR 2009. (Proc. 15th International Conference on Neural Information Processing (ICONIP'08), Auckland, New Zealand, 2008.)

[45] Nina Zhou and L. P. Wang, "A Novel Support Vector Machine with Class-dependent Features for Biomedical Data", IEEE International Conference on Systems, Man, Cybernetics (SMC 2006, October 8-11, 2006, Taipei), pp. 1666- 1670, 2006.

[46] Mustafa Göçken, Mehmet Özçalici, Asli BoruAyşe, Tuğba Dosdoğru, "Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection," Neural Computing and Applications, pp.1-16, 2017.

[47] L. Breiman, "Random Forests", Machine Learning, vol.45, pp.5-32, 2001.

[48] J. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," IEEE Trans. Syst., Man, Cybern., vol.23, pp.665-685, 1993.

[49] Y. Frayman and L.P. Wang, "Data mining using dynamically constructed recurrent fuzzy neural networks," Research and Development in Knowledge Discovery and Data Mining, vol. 1394, pp. 122-131, 1998.

[50] J. Qiao, W. Li, X.-J. Zeng, H. Han, "Identification of fuzzy neural networks by forward recursive input-output clustering and accurate similarity analysis," Applied Soft Computing, vol.49, pp.524-543, 2016.

[51] G. P. Choi, K. H. Yoo, J. H. Back, M. G. Na, "Estimation of LOCA Break Size Using Cascaded Fuzzy Neural Networks," Nuclear Engineering and Technology, vol.49, pp.495-503, 2017.

[52] Y. Frayman and L.P. Wang, "A Dynamically-constructed fuzzy neural controller for direct model reference adaptive control of multi-input-multi-output nonlinear processes", Soft Computing, vol.6, pp.244-253, 2002.

[53] X. Ma, Y. Jin, Q. Dong, "A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting," Applied Soft Computing, vol.54, pp.296-312, 2017.

[54] L. P. Wang and Y. Frayman, "A Dynamically generated fuzzy neural network and its application to torsional vibration control of tandem cold rolling mill spindles", Engineering Applications of Artificial Intelligence, vol. 15, pp. 541-550, 2002.

[55] S.-Y. Chen, C.-Y. Lee, "Compensatory fuzzy neural network control with dynamic parameters estimation for linear voice coil actuator," Precision Engineering, vol.48, pp.191-202, 2017.

APPENDIX

We give the details on the sources of data, features used in the simulations, and features excluded due to missing data.

Bit-X
Bitfinex
BitStamp
BTCChina
Huobi
Kraken
LakeBTC
OKCoin
Others (Smaller size exchanges like itBit, BTC-e, etc.)

Table 1. Exchanges tracked on Bitcoinity.org

Average Block Size
Blocks Size
Cost per Transaction
Cost Per Transaction by Percentage
Difficulty
Estimated Transaction Volume
Estimated Transaction Volume in USD
Hashrate
Market Capitalization
Market Price
Median Confirmation TIme
Miner's Revenue
Number of Orphaned Blocks
Number of Transactions
Number of Transactions excluding Chains longer than 100
Number of Transactions excluding Popular
Number of Transactions per Block
Number of Transactions in Total
Number of Unique Addresses
Output Volume
Total number of Bitcoins
Transaction Fees
Transaction Fees in USD
UTXO Count

Table 2. Features retrieved from Blockchain.info

Block Size Votes on each Exchange
Confirmation Time in each Exchange
Block Version of each Exchange
Block Size of each Exchange
Difficulty in each Exchange
Hash Rate of each Exchange
Transaction Count in each Exchange
Arbitrage in each Exchange
Bid-Ask Sum in each Exchange
Book Value in each Exchange
Market's Capitalization in USD for each Exchange
Market's Capitalization for each Exchange
Price in each Exchange
Price-Volume of each Exchange
Rank of each Exchange
Spread of each Exchange
Trades per Minute in each Exchange
Volatility of each Exchange
Volume of each Exchange

Table 3. Features retrieved from Bitcoinity.org

Bip 9 Segwit
Bitcoin Unlimited's Share
Mempool Count
Mempool Growth
Mempool Size
Mempool State by Fee Level
Number of Wallets on Blockchain.info
NYA Support

Table 4. Features removed from Blockchain.info