Reproducing Bitcoin Price Prediction Using Ensembles of Neural Networks*

Joe Ulepic
Final Report
Machine Learning 4824
Virginia Tech, Blacksburg, VA
joeu@vt.edu

Abstract

The purpose of this paper is to analyze the findings of *Bitcoin Price Prediction Using Ensembles of Neural Networks* which uses multiple neural networks to predict the next day change in bitcoin price. It reported that its model had an accuracy of about 58% to 63%. Meaning that fifty-eight to sixty-three percent of the time it correctly determines if the price of bitcoin will go up or not. Through reproducing this model with newer data, we found our model to correctly classify the day change price in bitcoin with accuracy of about 54% to 64%. To obtain these results an ensemble of five neural networks were trained each with a different number of hidden nodes in them. Then the GASEN approach was used to find the set of optimal networks for making a prediction. We found that a single network is typically within a max 4% difference of the ensemble minimizing the benefit of it. While some real-world applications can benefit from the GASEN method, we conclude that day change in bitcoin price is not one of them.

1 Methodology

1.1 Problem Statement

The gaining popularity of Bitcoin cannot be denied. This year alone the asset that prides itself on being the blue chip cryptocurrency, has seen gains of about 151% in 2020 alone. As many institutional investors begin flocking to bitcoin, and its value continues to gain, people are remaining optimistic that it will see long term success.

With any asset class and investment decision, however, a firm must be careful in how it comes to conclusions. They cannot adopt views that seem popular at the time, without strong evidence to support them. Firms avoid bad decisions by welcoming debate and involving a diversity of opinion. Similar to this practical operation, the paper *Bitcoin Price Prediction Using Ensembles of Neural Networks*, outlines the idea that multiple neural networks are better at making decisions than only one. As if each neural network is in the room proposing their decision based on their experience, or training.

While it is simple to think of a neural network as an expert investor, the paper by Sin and Wang does not go further than citing previous literature on this idea which concluded that in some cases the optimal subset of networks is more effective than averaging all of them [1]. This paper investigates the effectiveness of a subset of neural networks, in relation to a single neural network, to test whether effective for daily bitcoin price prediction.

*E. Sin and L. Wang, "Bitcoin price prediction using ensembles of neural networks," 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, 2017, pp. 666-671, doi: 10.1109/FSKD.2017.8393351.

1.2 Data

The data to produce this model is taken from Blockchain.com, a comprehensive bitcoin wallet and data provider. The dataset is composed of 20 features for each day over a period of the last three years. Combined, the final dataset was 1090x20 where the last 50 days were used as the validation set and the rest as the training set. Some of the notable features in the dataset include hash rate and difficulty. These technical indicators represent the processing power in the network and the how hard it is to mine a block, respectively. Along with the block sizes, and total transactions, these features were time dependent and were adjusted to their deltas or change in value. Finally, each feature was standardized to a mean of 0 and standard deviation of 1 since the value of each input varied greatly.

1.3 Ensemble

After collecting and pre-processing the data, five feed forward neural networks were created with two hidden layers. They sizes of the hidden layers vary so that each neural network is slightly different. The first layer of each network starts at 5 and increases by 5 consecutively to 25. The second layer is the whole number half of the first nodes. So, if the first layer has 15 nodes, the second has 7.

Each network is trained three times with three different algorithms, the Levenberg Marquardt algorithm, back propagation, and the Quasi Newton method. Each is used to see if a certain training algorithm produces better results for the ensemble than a single network.

The optimal networks from the ensemble are generated using the GASEN method. GASEN works by estimating the correlation between the individual neural networks C, thus being able to estimate their generalization error, E. Then the generalization error is applied to the standard generic algorithm [2].

$$C_{ij}^{V} = \frac{\sum_{x \in V} (f_i(x) - d(x))(f_j(x) - d(x))}{|V|} \qquad E_w^V = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j C_{ij}^V = w^T C^V w$$

From here the ensemble is chosen by computing lambda, or the mean of the generalized weights returned by the genetic algorithm. If a networks weights are above lambda it is selected for the ensemble. Lastly, predictions are made with the median of predictions from the individual networks in the ensemble.

2 Results

The accuracy of the GASEN ensemble was 64% when trained with the Levenberg Marquardt algorithm and Quasi Newton method and 62% for back propagation. Interestingly the GASEN method always selected the last three neural networks for each training algorithm. These individual networks with hidden layer nodes represented by [15,7], [20,10], and [25,12]. Stepping aside from the ensemble, it important to point out that all the individual networks except for two resulted in accuracy greater than 60%.

Algorithm	Levenberg Marquart	Back Propagation	Quasi Newton
Ensemble Accuracy	64%	62%	64%
[5,2] Accuracy	64%	62%	64%
[10,5] Accuracy	62%	62%	60%
[15,7] Accuracy	50%	62%	52%
[20,10] Accuracy	60%	64%	64%
[25,12] Accuracy	64%	62%	64%

3 Conclusion

While the paper *Bitcoin Price Prediction Using Ensembles of Neural Networks*, is correct in determining the classification rate for the model, it overstates the importance of the ensemble. It is important to note however, that utilizing the ensemble ensures that you get the best results any single network can produce. This is evident because the Quasi Newton method and Levenberg Marquart ensembles produced the same as the most accurate individual networks, 64%, while the backpropagation ensemble returned only 2% below at 62%. Since the individual networks are within a margin of 4% its clear that the ensembles do not provide sufficient improvement to individual networks and can conclude that their not effective for this problem. To further explore this topic, problems with less data could be studied to see if individual networks vary more than in this example, thus able to benefit from an ensemble.

4 Bibliography

[1] 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.

[2] D. E. Goldberg. Genetic Algorithm in Search, Optimization and Machine Learning. AddisonWesley, Reading, 1989.