Seventh International Conference on Information Science and Technology
Da Nang, Vietnam; April 16-19, 2017

Predicting Stock Trend Using Multi-objective Diversified Echo State Network

Zhengfu Liu
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
cbf1@mail.ustc.edu.cn

Zhenyu Liu
Department of Science and Technology
China University of Political Science and Law
Beijing, China
lzhy@cupl.edu.cn

Yinlong Song, Zhichen Gong
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
{songyinl,zcgong}@mail.ustc.edu.cn

Huanhuan Chen
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
hchen@ustc.edu.cn

***Abstract*—The prediction of stock price movements is considered as a challenging task for financial time series analysis. The difficulty of predicting the trends lies in the dynamic temporality and noise in the stock data. The Echo State Network (ESN) is a popular time series prediction model that considers the temporality of the stock time series, but ESN often falls into the dilemma of over-fitting due to the existence of many unnecessary neurons in the hidden layer. To predict the stock price movements, in this paper, we propose Multi-objective Diversified Echo State Network (MODESN). MODESN assigns a diversity requirement to the ESN such that the network is able to maintain better generalization ability. Experiments are performed with Shanghai Composite index to evaluate the effectiveness of the proposed approach. Compared with classical Echo State Network, the proposed MODESN usually achieves higher accuracy and avoids over-fitting. The experimental results indicate that MODESN performs favorably compared with alternative methods.**

Keywords—stock trend prediction; Echo State Network; Multi-objective Diversified Echo State Network; Multi-objective learning

I. INTRODUCTION

Stock tendency prediction is an important and challenging problem that leads to many studies over the past decades.

Previous researchers usually employ neural networks (NN) to predict stock tendency. Oh and Kim [4] proposed a stock trading model, and the core component of which includes a two-phases Back Propagation neural network (BPNN). Wang et. al proposed Wavelet Denoising-based Back Propagation (WDBP) neural network to predict the Chinese stock trend [1]. WDBP decomposes original data into multiple layers by the wavelet transformation. Each layer has a low-frequency and a high-frequency signal component. Then a BPNN is established by the low-frequency signal of each layer for predicting the future value. But BPNN can only learn the relationship between current input and output, and it ignores the dynamic temporality of the stock data.

Recurrent neural network (RNN) learns a hidden representation for each input by considering the previous representation and the current input, which enables it to have implicit memory ability. Kamijo and Tanigawa [5] apply RNN to the recognition of stock price pattern. But typically, RNN relies heavily on hand-crafted gradient descent, so RNN requires high calculation cost and is easy to fall into the local optima. In order to solve this problem, Jaeger [7] proposed Echo State network (ESN), which consists of a high dimensional randomly determined reservoir as the hidden layer. Compared with the traditional RNN models, ESN does not have the gradient descent process. Hence, the computation burden is reduced greatly. Lin et. al [8] predicted stock price with an ESN model and achieved good results.

However, there are some problems when applying ESN in stock tendency prediction. Firstly, in order to fit the dynamics of the stock series well, there usually needs many neurons in the hidden layer of the ESN which may lead to over-fitting easily and increase the time complexity unnecessarily. Secondly, the structure and parameters in an ESN are typically randomly generated. Thus, the performance of ESN models with different parameters is highly randomized. In a typical ESN setting, random parameters are re-generated until an appropriate ESN is found by cross validation. Besides, it is not guaranteed that the randomly generated ESN is better than the earlier ones.

In order to address the above limitations of ESN, this paper proposes a Multi-objective Diversified Echo State Network (MODESN) for predicting stock price. MODESN defines the diversity as a regularization term to improve the generalization ability of the ESN and employs a multi-objective non-domination ranking and fitness blocking evolutionary algorithm to ensure that newly generated networks are better than the previous ones.

978-1-5090-5401-5/17/$31.00

When stock tendency is predicted, large amounts of features are likely to introduce more noise to the model. So it's necessary to reduce the number of features. To solve this problem, in this paper we present a method which combines MODESN and Japanese Candlestick.

Japanese Candlestick is one of the most successful analysis methods used to visualize the stock price patterns. Generally, different patterns imply different stock trends. Japanese Candlestick assumes that the stock price of the next day is related to the price in the previous three days [10]. Some researchers have used candlestick as a feature in artificial neural network (ANN). Due to the characteristic of candlestick, all the features in three days have to be extracted when it is applied in ANN. For example, Jasemi et. al [9] proposed an algorithm to predict the stock market by using a supervised feed-forward neural network and Japanese Candlestick. They extracted 15 features to predict stock trend --- 5 features per day. Since MODESN is based on ESN which has implicit memory ability, only one-day data need to be dealt with when we extract features. That is to say, the combination of MODESN and Japanese Candlestick helps reduce the feature number so that we can predict stock market by employing only 5 features.

The major contributions of this paper are: (1) we define the diversity of ESN. We can improve the system's generalization ability and decrease the occurrence of over-fitting by optimizing the diversity of ESN. (2) We propose a multi-objective diversified ESN (MODESN) model which is based on multi-objective genetic algorithms (MOGA). By optimizing the accuracy and the diversity of ESN at the same time, MODESN will converge after sufficient iterations. (3) We reduce the feature number by combining the MODESN and the Japanese Candlestick to improve the prediction accuracy.

The rest of this paper is organized as follow. In part II, we review related works briefly on stock trend prediction algorithm. In part III, we introduce the proposed multi-objective Diversified ESN (MODESN). Part IV presents experimental results and part V concludes this paper.

II. RELATED WORKS AND STOCK PREDICTION ALGORITHM

A. *Japanese Candlestick*

Japanese Candlestick is a popular technical analysis approach in predicting stock trend. Japanese Candlestick consists of some candlestick patterns. Since different candlestick patterns indicating different stock trends[1], we can predict stock market by analyzing the candlestick patterns.

A candlestick pattern is made of one to three candle sticks [10]. And every candle stick is composed of the body (red and green), and an upper and a lower shadow (wick): the area between the open price and the close price is called the real body. The part above the body is upper shadow (wick) and the part under the body is called the lower shadow (wick). The shadow represents the highest and lowest stock price. The highest and lowest position of body are open price and close price respectively when the body is green, and the body is red, with the opening price at the bottom of the body and the closing price at the top [10]. Figure 1 illustrates the candle sticks.

Some researchers use Japanese Candlesticks to predict financial time series. We need the stock messages in last three days to predict the stock trend in Japaneses Candlesticks, so the traditional methods need to extract features from the previous three days. Milad Jasemi [9] presented a model to predict the stock market on the basis of a supervised feed-forward neural network and Japanese Candlestick. They extracted 15 features to predict stock trend---extracted 5 features per day. Goswami [12] proposed a model based on Candlestick and SOM-CRB to predict stock price fluctuation. They extracted 12 features during the last three days to predict the next day.

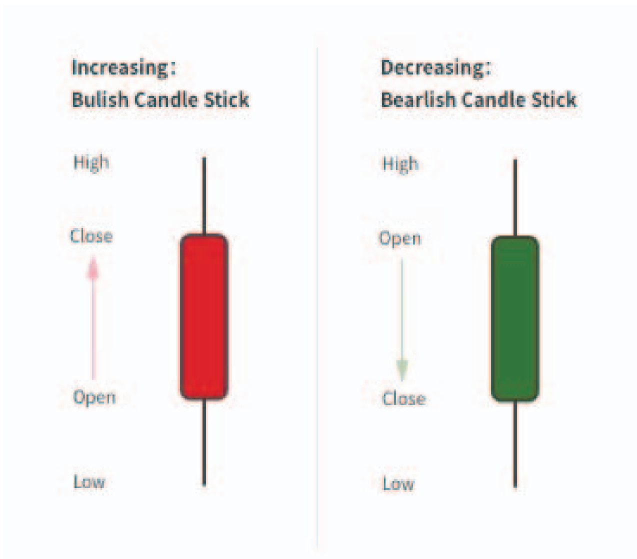

Figure 1 Illustration of the candlestick chart. The red body demonstrates the close price higher than the open price. The green body demonstrates the open price higher than the close price.

B. *Recurrent Neural Network*

Recurrent Neural Network (RNN) connections between units form a directed cycle, which creates an internal state of the network that allows it to exhibit dynamic temporal behavior[2] [13].

In traditional neural network models, each hidden layer is fully connected with previous layer, while there are no connections among the neurons in the same layer. This structure ignores the temporality of the time evolving data. In a RNN, neuron units in the hidden layer are also connected, which makes it to have the memory ability. The current output is related to the current input and the output of the front. Figure 2 shows the basic structure of a RNN. However, the recurrent training of the RNN network requires complex adjustment of the algorithm.

[1] https://en.wikipedia.org/wiki/Candlestick_chart

[2] https://en.wikipedia.org/wiki/Recurrent_neural_network

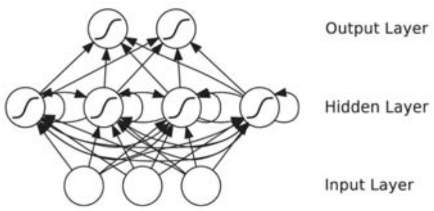

Figure 2 Basic structure of RNN. It demonstrates a RNN with three layers, i.e. input layer, hidden layer and the output layer.

C. *Multi-objective evolutionary algorithm (MOEAs)*

Evolutionary algorithms (EAs) are inspired by insights of biologically selecting process [6]. EAs are one of the most promising approaches to handle multi-objective optimization problem and have demonstrated their encouraging results [2][11][13]. MOEAs search solutions for multiple possibly conflicting objectives simultaneously. A set of candidate solutions are considered in a run, which is also called a population [13]. It tries to find a compromise between two or more conflicting objectives. Assume the objective functions are f_{i}, where i is in {1,2,...,n}. To compare different solutions, the idea of non-domination is used.

The definition of the non-domination is:

- For any two solutions S1 and S2, S1 dominate S2 if for all objective function f_{i}, i in{1,2...n}, S1 is better than S2.
- If S1 do not dominate S2, then S1 non-dominate S2.

The set of solutions that is non-dominated by any others are called Pareto optimal solutions. All the Pareto optimal solutions constitute the Pareto front. Adapting to the different inclinations of decision makers, each solution in this set is optimal in the sense of non-domination. The principle of MOEA is to approximate the non-dominated front uniformly. However, MOEAs do not promise to return the global optimal solution and alternatively a trade-off between different objectives.

Evolutionary multi-objective optimization seems to be straightforward to deal with multiple objectives [6][11] and MOEAs perform a stochastic search for a satisfactory solution in the space of chromosome, which consists of the free parameters of the underlying model.

NSGA-II is a kind of non-domination based MOEAs that is popularly employed in the domain of EC. In this paper, we employ NSGA-II to optimize the objectives in our experiment.

III. MULTI-OBJECTIVE DIVERSIFIED ECHO STATE NETWORK

In this section, we propose that diversifying the state matrix in the ESN to decrease the probability of occurrence of over-fitting and ensure that recently generated ESN is better than the previous.

A. *Echo State Network*

ESN is a special kind of RNN. It distinguishes itself from a typical RNN by the randomly generated hidden layer [7]. The neurons in the hidden layer are recurrently connected to provide nonlinear modeling ability. On the top of the nonlinear hidden component, the output function can be efficiently trained by linear regression. The formula of ESN can be justified by:

$$X_{t+1} = f(w_0 * input_{t+1} + w_1 * X_t) \quad (1)$$

$$output_{t+1} = X_{t+1} * w_2 \quad (2)$$

where X_t represents the value of the hidden layer at the time t, and X is also called the state matrix; w_0 is the connection weight matrix of the input layer and the hidden layer, w_1 is the connection weight matrix of the hidden layer and hidden layer and w_2 is the connection matrix of the hidden layer to the output layer. $input_t$ and $output_t$ represent the input and output at the time point t. Equation (1) is the state transition function of ESN. *f* is a nonlinear function, which is usually set as tanh or sigmoid function. Equation (2) is the state to output mapping, which is the only part to be trained for ESN.

Figure 3 demonstrates the structure of ESN. Different from traditional RNN, there are many neurons in the hidden layer in ESN. w_0 and w_1 are generated randomly and will not be changed. Then we can get the state matrix X by the formula (1). Finally, in according to the formula (2), we can get w_2 by linear regression.

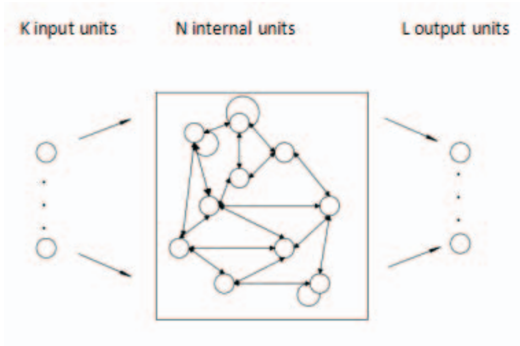

Figure 3 Structure of ESN

B. *Multi-objective Diversified Echo State Network (MODESN)*

The cores of the MODESN are the definition of the diversity of state matrix which is optimized to decrease the occurrence of over-fitting and the utilization of multi-objective evolutionary algorithm i.e. NSGA-II to ensure that recently generated network is better than the previous ones.

In order to decrease the occurrence of over-fitting, we need to reduce the number of neurons in the hidden layer. Therefore, we need to let the different neurons in the hidden layer carry different information as much as possible. In ESN, for any one of the nodes in the hidden layer, the input at the time t is related to a value. The inputs at all the time are related to a time series. That is to say, letting each neuron encode different information as much as possible is equivalent to letting all the time series different as much as possible.

Xie et. al [3] proposed Diversifying Restricted Boltzmann Machine (DRBM) to deal with the problem that Restricted Boltzmann Machine (RBM) ignores the problem of the long-tail region in document modeling. The main idea of DRBM is to make the connect vector of the input layer and each neuron in hidden layer as different as possible. That is to say, DRBM improves the neuron's ability of expressing message in the hidden layer by optimizing the connect matrix of input layer and hidden layer.

However, we cannot use this idea directly in our model. Since, in the ESN, the value of the hidden layer in t+1 time is not only related to the input layer but also the hidden layer in time step t. If we optimize the connection matrix of the input layer and the hidden layer or the connect matrix of hidden layer and hidden layer or both, we cannot ensure that all time series are related to hidden layer as different as possible. Therefore, different from the DEBM, MODESN optimizes the state matrix X directly. MODESN defines the diversity of state matrix by which we can measure how different is the time series related to neurons in the hidden layer, and ensures recently generated network is better than the previous.

Firstly, we define the distance $\mathrm{d}(a_i, a_j)$ between time series a_i and a_{j} to be

$$\mathrm{d}(a_i, a_j) = \sum_k |a_{ik} - a_{jk}|,$$

where a_{ik} represents the k-th element of the time series a_{i}. Then we define the diversity metric of state matrix X as

$$\Omega(X) = \Psi(X) - \Pi(X),$$

where

$$\Psi(X) = \frac{1}{K^2} \sum_{i=1}^{K} \sum_{j=1}^{K} d(a_i, a_j)$$

is the mean of the distances between all pairs of hidden neuron time series, and

$$\Pi(X) = \frac{1}{\mathrm{K}^2} \sum_{i=1}^{K} \sum_{j=1}^{K} | d(a_i, a_j) - \Psi(X) |$$

is the standard deviation of these distances. Obviously, the larger $\Omega(X)$ is, the more different these time series are. Therefore, $\Omega(X)$ responses the generalization ability. Therefore, letting each neuron represents different message as much as possible is equivalent to letting $\Omega(X)$ become larger as much as possible.

In MODESN, $\Omega(X)$ and accuracy (R) in the data set are the optimal objects. We get the final model by the NSGA-II.

The gene of our approach is composed of the weight matrix w_0 which connects the input layer and the hidden layer and w_1 which connects the hidden layer and hidden layer.

We stop the evolution by two criterions: (1) the iteration exceeds the maximum number. (2) the improvement of the objective values are below a preset threshold.

IV. EXPERIMENT

In this section, we present the experimental setup and results. Compared with the typical ESN, our results show that the MODESN is able to decrease the occurrence of over-fitting, and improves the generalization ability and ensure that it's faster to find a better network.

A. Experiment Setup

The data used in this paper is the data of Shanghai Composite Index from 2000/1/4 to 2016/1/21. We divide these data into two partitions--- the training set (2000/1/4-2012/6/1) and the test set (2012/6/4-2016/1/21).

We extract five features based on Japanese Candlestick from datasets as inputs. The five features are (C-O)/O, (H-O)/O, (L-O)/O, (ma5-ma10)/ma10 and (C- CB)/CB, where C, O, H and L represent the close price, the open price, the high price and the low price of the stock respectively. Ma5 and ma10 represent the average close price for the past 5 days and 10 days, respectively. Generally speaking, when ma5 is larger than ma10, technical analysis considers the stock price is rising in the recent periods. When ma5 is less than ma10, the stock price is falling in the recent periods. Therefore, (ma5-ma10)/ma10 reflects the stock trend in the recent periods. The CB represents the close price of the day before. Finally, we set the label 1 when the next day's close price is higher than the day's close price, and set -1 when lower and set 0 when equal.

We design three groups of experiments. In part B we first randomly generate 30000 ESNs, and then select the ESNs which get the top 10 highest accuracy in the training set. Finally, we integrate these ESNs by majority voting to predict the stock trend in the test set. In part C we get the MODESNs in which $\Omega(X)$ and the accuracy on the training set are simultaneously optimized by NSGA-II. And then, we integrate these MODESNs by majority voting to predict the stock tend in the test set. In part D we get the MODESNs which optimizes $\Omega(X)$ and the accuracy on the training set by NSGA-II. The difference between part C and D is that we set $\Omega(X)$ to the constant 1 in the experiment D. We turn the model into a single objective optimization model which only optimizes the accuracy.

B. Random Generated ESN (RGESN)

We set the number of neurons in the hidden layer as 70 and randomly generate 30000 ESNs. And then, we sort these ESNs in accordance with the accuracy and select 10 ESNs which get the highest accuracy in the training set. Finally, we integrate these ESNs by voting to predict the stock trend in the data set. Table I shows the accuracy of the top 10 ESNs in the training set and test set. And the final integrated model's accuracy in the training set and the test set are 76.300% and 73.200%, respectively.

TABLE I. ACCURACY OF THE TOP 10 ESNS IN THE DATASETS

Accuracy in the training set(%)	Accuracy in the test set(%)
75.192	73.502
75.000	73.863
74.115	72.316
76.846	74.502
74.692	73.000
74.500	73.263
74.615	73.043
76.000	73.770
75.577	72.683
74.884	72.229

C. Multi-objective diversity ESN(MODESN)

Firstly, we randomly generate 200 ESNs which has 70 neurons in the hidden layer and calculate these ESNs' diversity $\Omega(X)$ and accuracy R. And then, we create a new population which contains 100 ESNs by selection, crossover and mutation. We get 18 ESNs in the 300^{th} generation. Finally we integrate these 18 ESNs by voting to predict in the test set. Table II presents the performance of these 18 ESNs on the data set. And the final integrated model's accuracy in the training set and the test set are 77.702% and 76.100%, respectively.

TABLE II. PERFORMANCE OF THESE 18 ESNS

Accuracy on the training set(%)	Diversity	Accuracy on the test set(%)
77.500	1574.3	76.502
75.00	1718.0	75.863
77.115	1703.8	75.316
78.846	1561.7	77.502
77.692	1570.4	75.863
77.500	1701.9	75.863
74.615	1718.2	75.043
75.000	1715.2	74.770
75.577	1714.5	76.683
77.884	1566.5	77.229
77.884	1569.1	78.322
75.961	1712.4	76.409
75.961	1712.5	75.863
78.076	1564.6	76.136
76.923	1711.7	76.136
76.346	1712.1	75.043
77.692	1572.3	77.229
76.923	1705.5	76.409

D. Single objective ESN (SOESN)

The difference of the experiment B and the experiment C is that we set the $\Omega(X)$ to constant 1 which turns the NGSA-II to a single objective optimization algorithm. That is to say, to select, to cross and to mutate during the 300 generation is to optimize the right rate R. Finally, we get the ESN which achieves the highest right rate on the training set. And the final ESN's accuracy in the training set and the test set are 78.800% and 74.200%, respectively.

E. Analysis

According to the performance of RGESN, MODESN, SOESN in the data set, it is observed that in the training set, the SOESN model has the highest accuracy. We use genetic algorithm to optimize accuracy. The highest accuracy on the training set can be obtained, but SOESN is not the highest in the test set. This also demonstrates that there is some degree of over-fitting when we use SOESN. Figure 4 demonstrates the difference of the accuracy in the training set and the test set. *In* fact, when we use MODESN, it can be found that the difference of the accuracy in the training set and test set is the smallest, and the accuracy in the test set is the highest. And the only difference of MODESN and SOESN is that MODESN not only optimizes the accuracy but also optimizes the value of diversity. This demonstrates that optimizing the value of diversity can avoid over-fitting.

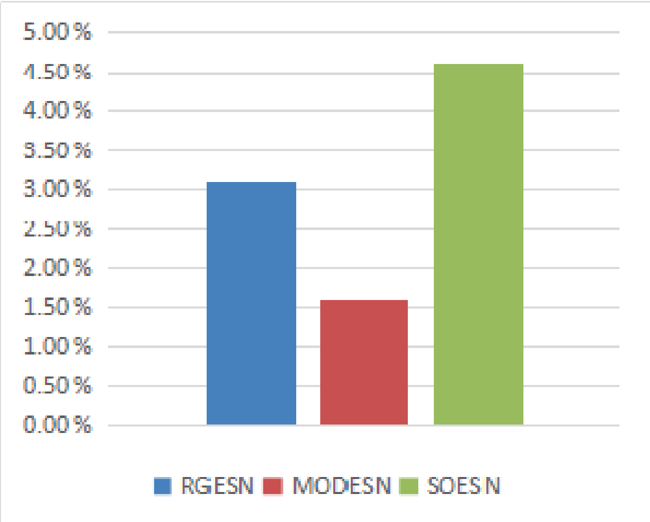

Figure 4 Compared with the training set, the decreased accuracy in the test set

V. CONCLUSION

In this paper, we define the diversity of ESN and propose a Multi-objective Diversified ESN (MODESN) for predicting the stock price trend. A multi-objective evolutionary algorithm is presented to optimize the diversity and prediction accuracy combined with Japanese Candlestick method. Compared with classical ESN, the accuracy of MODESN and SOESN are improved in both training and test set. And in the test set, MODSEN achieves a higher accuracy, which indicates that MODESN can not only ensure the higher accuracy but also can avoid over-fitting.

Acknowledgements: This work is supported by the National Natural Science Foundation of China under Grants 61673363, 91546116, 61511130083.

REFERENCES

[1] J. Z. Wang, J. J. Wang, Z. G. Zhang, and S. P. Guo, "Forecasting stock indices with back propagation neural network", *Expert Systems with Applications*, vol. 38, pp. 14346-14355, 2011.

[2] J. Clune, J. B. Mouret, and H. Lipson, "The evolutionary origins of modularity," Proceedings of the Royal Society B Biological Sciences, vol. 280, pp. 23-24, 2013.

[3] P. Xie, Y. Deng, and E. Xing, "Diversifying restricted boltzmann machine for document modeling", A*CM SIGKDD International Conference*, pp. 1315-1324, 2015.

[4] K. J. Oh, and K. J. Kim, "Analyzing stock market tick data using piecewise nonlinear model", *Expert Systems with Applications*, vol. 22, pp. 249-255, 2002.

[5] K. Kamijo, and T. Tanigawa, "Stock price pattern recognition-a recurrent neural network approach", *IJCNN International Joint Conference on Neural Networks*, vol. 1, pp. 215-221, 1990.

[6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II", *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 182-197, 2002.

[7] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication", *Science*, vol. 304, pp. 78–80, 2004.

[8] X. Lin, Z. Yang, and Y. Song, "Short-term stock price prediction based on echo state networks", *Expert systems with applications*, vol. 36, pp. 7313 – 7317, 2009.

[9] M. Jasemi, A. M. Kimiagari, and A. Memariani, "A modern neural network model to do stock market timing on the basis of the ancient investment technique of japanese candlestick", *Expert Systems with Applications*, vol. 38, pp. 3884–3890, 2011.

[10] S. Nison, "Japanese candlestick charting techniques: a contemporary guide to the ancient investment techniques of the Far East", *Penguin*, 2001.

[11] H. Chen, and X. Yao, "Multiobjective neural network ensembles based on regularized negative correlation learning", *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1738–1751, 2010.

[12] M. M. Goswami, C. K. Bhensdadia, and A. Ganatra, "Candlestick analysis based short term prediction of stock price fluctuation using som-cbr", *IEEE International Conference on Advanced Computing*, pp. 1448-1452, 2009.

[13] L. Li, X. Yao, R. Stolkin, M. Gong, and S. He, "An evolutionary multiobjective approach to sparse reconstruction", *IEEE Transactions on Evolutionary Computation*, vol. 18, pp. 827–845, 2014.