

## Predict Time Series with Multiple Artificial Neural Networks

Fei Li<sup>1</sup>, Jin Liu<sup>1</sup> and Lei Kong<sup>2,\*</sup>

<sup>1</sup>College of Information Engineering, Shanghai Maritime University, 201306  
Shanghai, China

<sup>2</sup>The third branch of Shanghai Administration Institute, 200233,  
Shanghai, China

{feili, jinliu}@shmtu.edu.cn, sdx\_kl@163.com

\* Corresponding Author: Lei Kong

### Abstract

*Time series prediction is a challenging research area with broad application prospects in machine learning. Accurate prediction on a time series' value can provide important information for the decision-makers. In the literature, many works were reported to extend different architecture of artificial neural networks to work with time series prediction. However, most of the work only considered the target time series itself, while neglecting the impact of the relevant time series. In this paper we proposed a novel method MANNP that makes use of multiple artificial neural networks to conduct the time series prediction. The proposed method creates time series model and forecast time series. To verify the effectiveness of the proposed method, we apply MANNP to a shipping price index time series prediction. The experimental results show that this method can improve accuracy of prediction when compared with traditional methods.*

**Keywords:** Time Series, Multiple Artificial Neural Network, Prediction

### 1. Introduction

Time series [1] refer to a series of observation values obtained in chronological order, and many areas are related to the time series. In the research field of natural science and social science, a large number of decision-making problems are inseparable from the forecast. Time series prediction [2] is expected to use several historical observations values of time series to predict the future value, and appears in many real-world problems, such as weather forecasting, stock index prediction, shipping index prediction.

A time series is a sequence of vectors,  $x(t)$ ,  $t = 0, 1, \dots$ , where  $t$  represents elapsed time. Theoretically,  $x$  may be a value which varies continuously with  $t$ , such as a temperature. In practice, for any given physical system, it will be sampled to give a series of discrete data points, equally spaced in time. To better illustrate the relevant concepts and ideas, we assume that every data point that is collected as the sample will be used for the prediction. Neural networks have been widely used as time series forecast. Work in neural networks has been concentrated on forecasting future developments of the time series from values of  $x$  up to the current time. Formally this can be stated as: find a function  $f: R^N \rightarrow R$  such as to obtain an estimate of  $x$  at time  $t+d$ , from the  $N$  time steps back from time  $t$ , so that:

$$x(t+d) = f(x(t), x(t-1), \dots, x(t-N+1)) \quad (1)$$

$$x(t+d) = f(y(t)) \quad (2)$$

Where  $y(t)$  is the  $N$ -ary vector of lagged  $x$  values. Normally  $d$  will be one, so that  $f$  will be forecasting the next value of  $x$ .

Time series prediction with the traditional methods is relatively well formed, but linear methods are often not very effective. Although some nonlinear time series prediction method can simulate complex relationships, most methods only consider the time series itself, does not weigh the impact of relevant time series. In order to solve this problem, this paper presents a multiple artificial neural networks prediction method, the method can significantly improve the accuracy of both single time series and multiple time series. The remaining of this paper is organized as follows. Section 2 presents the related work in the literature. Section 3 explains our proposed method. Section 4 presents the experiment results. The paper ends up with some conclusions in section 5.

## 2. Related Work

Traditional time series prediction methods can be divided into linear prediction methods and nonlinear prediction methods. Over the last decades, time series prediction with linear prediction has been broadly studied in statistics and several models have been proposed, *e.g.* Auto Regressive Model (AR), Auto Regressive Moving Average Model (ARMA) and Auto Regressive Integrated Moving Average Model (ARIMA). After that, time series prediction adopted regression methods to solve the linear problem. Since 1990s machine learning methods have been used in time series prediction. In the real world, most of the time series are nonlinear problem. Many machine learning methods have approached the time series prediction problem, *e.g.* Artificial Neural Networks (ANNs) and Support Vector Machines.

### 2.1. Linear Model

One of the widely used linear prediction model is auto regressive (AR) model. In 1927 the British mathematician Yule [3] proposed this model to predict the variation of the market, marking appearance the time series analysis method. AR model is to use the known  $N$  data, to formulate the first  $N$  points in front or behind the data and to carry out the prediction of the data. The second one is autoregressive moving average (ARMA) model, in 1938 the Swedish statistician Herman Wold[4] completed a systematic study of the discrete stationary time series, proposed a famous decomposition method, and then presented ARMA model. There is another well-known time series prediction model, auto regressive integrated moving average (ARIMA) model, proposed by a US statistician Box Jenkins [5] in the 1970s. He proposed the time series analysis method theory and applications by a systematic and thorough discussion. This work is considered to be a leap in the history of the development of the time series analysis.

From the historical point of view, before the 1980s, prediction of time series usually used linear model. These models are linear and are not able to cope with certain nonstationary signals, and signals whose mathematical model is not linear. An obvious drawback is that these algorithms are linear, and are not able to cope with certain nonstationary signals and the signals with a distribution model that is nonlinear.

### 2.2. Nonlinear Model

As to the nonlinear time series prediction model, the first one is the radial basis function (RBF) neural network. In 1985, Powell proposed multi variable interpolation of radial basis function. In 1988, Broomhead and Lowe firstly applied RBF to the neural network design, which constituted a radial basis function neural network [6]. The basic idea of RBF is transformation of the input vector, which transforms the low dimensional input data into the high dimension space, and makes the weighted sum of the hidden layer as the output. Radial basis function network is a kind of local approximation network, which only uses a small number of neurons in the input space to determine the output of the network.

The second widely used model is back propagation (BP) neural networks [7], which was proposed by Rumelhart and McClelland in 1986. It's a multiple layer feedforward network trained with the error back propagation, and is one of the most widely used neural network models. When the network gets a pair of learning model, the neuron of activation values from the input layer propagate to the output layer through the hidden layer. The output layer of each neuron corresponds to input pattern network response. Subsequently, in order to reduce the actual output error and get the desired output, the output is tuned in accordance with the output layer through the middle layer. With the training of the back propagation of the error, the correct rate of the network output is also improved. A mapping is thus established between input and output, and then the purpose of prediction is achieved.

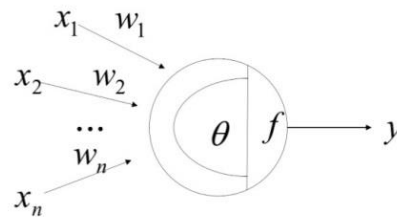
Support vector machine (SVM), which was first proposed by Corinna Cortes and Vapnik in 1995 [8] is also widely used in time series prediction. SVM showed many unique advantages in solving the small sample, nonlinear and high dimensional pattern recognition problems. In the practice, SVM not only was applied in machine learning problems such as fitting function, but also be applied to the prediction of the time series. The main idea of SVM is to map the input vector into a high dimensional feature vector space, and to construct the optimal classification surface in the feature space. Support vector machines has attracted attention of the academic researchers and has been widely used in various fields for its ability to approximate any complex system and advanced complete theoretical system. Moreover, Pisoni et al [9] used nonlinear autoregressive models (NARX) and artificial neural networks (ANNs) for environmental prediction. Faruk [10] proposed a hybrid neural network and ARIMA model for water quality time series prediction. Yu et al [11] proposed a new hyper-parameters selection approach for support vector machines to predict time series. Hrasko et al [12] used Restricted Boltzmann Machine and the Back propagation algorithm for time series prediction.

Neural networks (NN) are powerful when applied to the problems that require a model which is difficult to specify, but there is abundant data sample for the problem. Most of the studies reported above were simple applications of traditional time series approaches and ANNs. Most of these methods only considered the time series itself, did not weigh the impact of relevant time series. To address these problems, we proposed the MANNP model to solve the problem.

### **3. Multiple Artificial Neural Network Prediction (MANNP) Model**

#### **3.1. Artificial Neural Network**

Artificial neural network [13] has the ability of nonlinear adaptive information processing. It is not necessary to know the mathematical equations of expression mapping relationship, and can be trained to store a large number of mapping schema of input and output. ANN is usually a complex network system, which is composed of a large number of simple neurons. After training and testing with collected sample data, a neural network model for analyzing the data can be established. A neuron is shown in Figure 3.1,  $x = (x_1, x_2, \dots, x_n)$  is the input of the neuron,  $w = (w_1, w_2, \dots, w_n)$  is the connection weights of the neuron;  $\theta$  is the threshold of the neuron;  $y$  is the output of the neuron;  $f$  is the activation function.



**Figure 3.1. Neuron Model**

The output formula is

$$y = f \left( \sum_{i=1}^n w_i x_i - \theta \right) \quad (3)$$

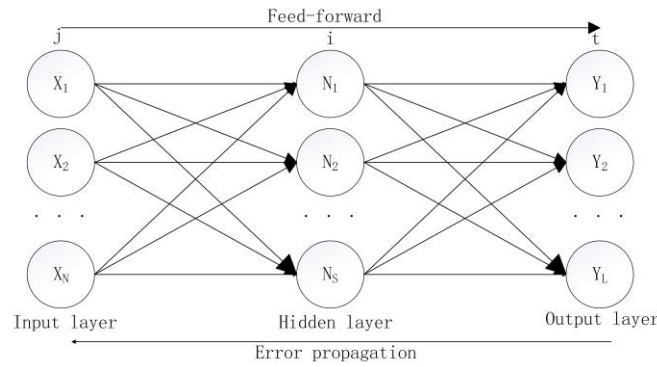
The common activation function is

$$f(x) = \frac{1}{1 + e^{-x}} \quad x \in (-\infty, +\infty) \quad (4)$$

Artificial neural network (ANNs) are a class of typical intelligent learning paradigm, widely used in practical application domains including: pattern classification, function approximation, optimization, prediction and automatic control and many others. In this paper, backward propagation neural network is used for modeling the time series prediction.

Topological structure of BP neural network in consists of input layer, output layer and a number of hidden layers, each layer contains a number of neurons, and each neuron is connected with a different layer of neurons. There is no connection on neurons of the same layer. The learning algorithm of BP neural network is forward and backward propagation. Forward propagation allows the input information to be propagated to the output layer, and activation functions work on the corresponding weights. When the error between the output and the desired value is bigger than the given precision value, it will make the error propagate back. In the process of error propagation, the weights and thresholds of each layer are modified. So with the repeated iterative operation, finally the error of the transmission signal will reach to the given precision.

The BP network architecture consists of one hidden layer of neurons with nonlinear transfer functions and an output layer of linear neurons with linear transfer functions. A schematic diagram of back-propagation network is given in Figure 3.2, where  $X_j$  ( $j=1,2,\dots,N$ ) represents the variables;  $N_i$  ( $i=1,2,\dots,S$ ) represents the outputs of neurons in hidden layer and  $Y_j$  ( $t=1,2,\dots,L$ ) represents the outputs of neural network. A neural network must be trained to determine the values of weights that will produce the correct outputs.



**Figure 3.2. BP Neural Network Model**

### 3.2. Description of MANNP Model

After 1980s, there has been a resurgence in the field of time series prediction, that it becomes clear that this kind of prediction work is a suitable application for a neuronal network. The ANN approach for time series prediction is non-parametric, in the sense that it is not necessary to know any information regarding the process that generates the signal. Neural networks based prediction has been explored from their beginning and development, because of the approximation and generalization property of ANN.

MANNP model is a multiple artificial neural network prediction model. Artificial Neural networks have capabilities to process nonlinear adaptive information, and are able to store large amounts of input and output mappings. With the repeatedly back-propagated error, the correct response will be raised in the output layer. A mapping is thus established between input and output, and then the purpose of time series value prediction is achieved. MANNP used multiple ANNs to create the model for time series prediction. The following is the steps to create MANNP model:

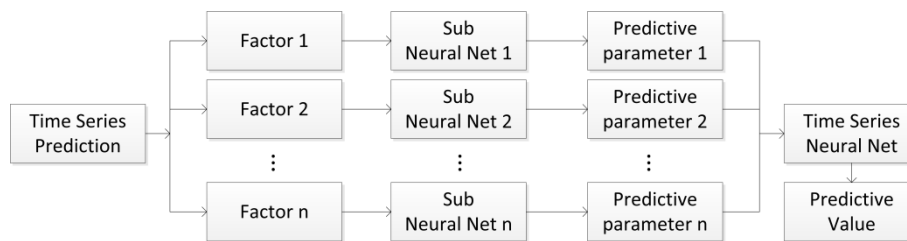
Step1: Analyze the time series prediction, and find out the influence factors of this time series, then research on influence factors of the target time series.

Step2: Collect relevant factor time series data from the Internet, analyze and preprocess the historical data.

Step3: Train neural network of the influence factors respectively, and save their models, and then train this time series neural network, save this model.

Step4: Use the predicted values of influence factors as the input values of the neural network for target time series, then use the multiple neural network to generate the predicted value.

The MANNP model is shown in Figure 3.3.



**Figure 3.3. MANNP Model**

The following pseudo code describes the methodology used in training NNs.

```

For each data series {
    For each data transformation {
    
```

```

    Linearly scale in-sample data to (-1, 1), retaining parameters
    For NN architectures with hidden nodes of (1, 3, 5, 7, 9...) {
        Train n NNs from random starting parameter weights
        Keep the best of the n based on SSE
    }
    Linearly scale the out-of-sample data using parameters from in-sample (from
    above)
    Using the best network architecture and parameter set, perform forecast on out-
    of-sample data
    Save the each model.
}
}

```

The following pseudo code describes MANP model predict object time series.

```

For each data predict {
    N factor load each trained model
    Predict the predictive parameter
    Save the each predictive parameter
}
Predict object time series {
    Load the object trained model
    Using n factors predictive parameters as input data
    Predict the predictive value
    Output the value
}

```

For training purposes, a dynamic back-propagation algorithm is required to compute the gradients, which is more computationally intensive than static back-propagation and takes more time. In addition, the error surfaces for dynamic networks can be more complex than those for static networks.

The connection weights and bias values of ANN are initially chosen as random numbers and then fixed by the results of training process. Many alternative training methods are available, we use the back-propagation [14] to train our models. When we train the models, we assume there are  $P$  input-output pairs of vectors for ANN. The goal of any training algorithm is to minimize the global (mean squared sum) error  $E$  between the real network output  $o^{(p)}$  and the desired output  $d^{(p)}$ . The formula is defined as follows:

$$E = \frac{1}{2P} \sum_p \sum_k (d_k^{(p)} - o_k^{(p)})^2 \quad (5)$$

Where  $p$  is the index of the  $P$  training pair of vectors,  $k$  is the index of elements in output vector,  $d_k^{(p)}$  and  $o_k^{(p)}$  are  $k$  th element of  $p$  th desired output vector and real output vector respectively. Additionally, we used the mean square error (MSE) to evaluate the learning effects of BP ANN. The training continues until the MSE falls below some threshold or tolerance level.

### 3.3. Multi-step Forecasting

Obviously, for real-world application, the main objective in time series forecasting is to have the ability to predict in time, up to a certain time prediction of horizontal  $K$ . For forecasting, the so called multi-step prediction task can be achieved by two methods. One is directly prediction, which is to explicitly train a model to predict several steps. Another

is the iterative method, which is doing repeated one-step predictions up to the desired horizontal point [15].

The multi-step forecasting (predicting more than one day in advance) is iterative process where the output of system is fed again as input, in other words, it takes the previous forecasting data instead of the true values to predict the next values when there are no previous true values. And it is commonly used for short horizontal forecast [16].

Time series is dynamic, not only the sequence of the day and the previous day's changes have a relationship, but the changes in the previous period of time is also relevant. Therefore, in this paper, the sliding window simulation method is used to use the previous time series as the input of the neural network. In the prediction, the  $m$  actual observation value are the input of the network, and the output value is the predicted values for the next moment. According to the particular time series modelling situation, our model will carry out the  $K$  step prediction,  $K$  is the moving window for time series prediction.

## 4. Experiment

### 4.1. Experiment Settings

In the experiments, we used time series of a shipping price index as the prediction target. For the impact factors of shipping price index, we choose five factors for this time series, which are international oil price index(WIF sequence), consumer price index(CPI sequence), producer price index(PPI sequence), NASDAQ sequence, and the exchange rate sequence, which are well known relevant to predict shipping index sequence.

We gathered historical data from January 1, 2014 to February 20, 2015, a total of 416 data points which is composed with six data value. The data is separated into two sets, the former 400 data points are used as the train set and the latter 16 data points are used as test set. Firstly, five time series were given the prediction value, then they were used as input values to shipping index prediction BP neural network, and finally predict the shipping index time series value.

We used the neural network, in order to eliminate the influence of unit of data for training and predicting, so we need to normalize the input of neural networks, and anti-normalize the output of neural networks. In this paper, we used the following formula to normalize the data processing:

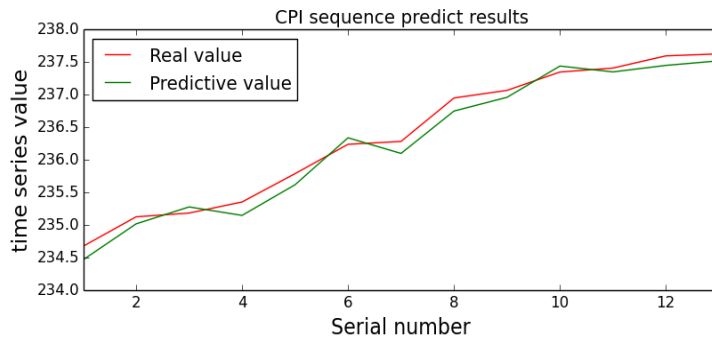
$$\tilde{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (6)$$

And we used following formula to anti-normalize the data processing:

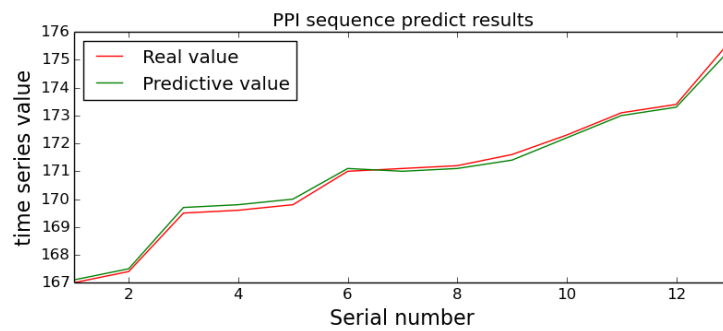
$$\tilde{y} = y*(x_{max} - x_{min}) + x_{min} \quad (7)$$

### 4.2. Experiment Result

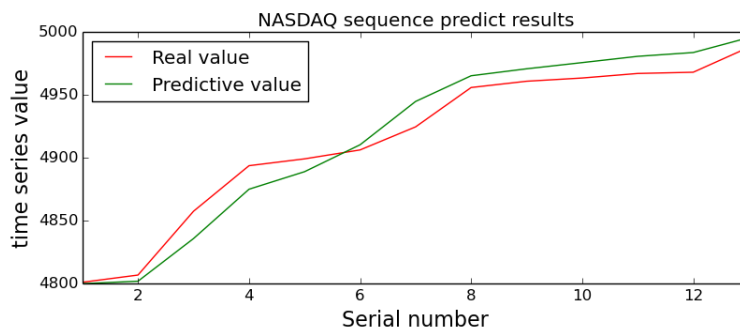
After the experiment results are obtained, the aforementioned five time series prediction are shown in Figure 4.1, 4.2, 4.3, 4.4, and 4.5 respectively.



**Figure 4.1. Multi-Step Prediction Results Of CPI Sequence**

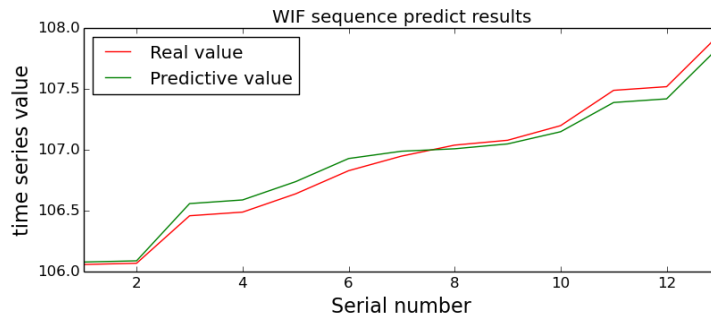


**Figure 4.2. Multi-Step Prediction Results of PPI Sequence**

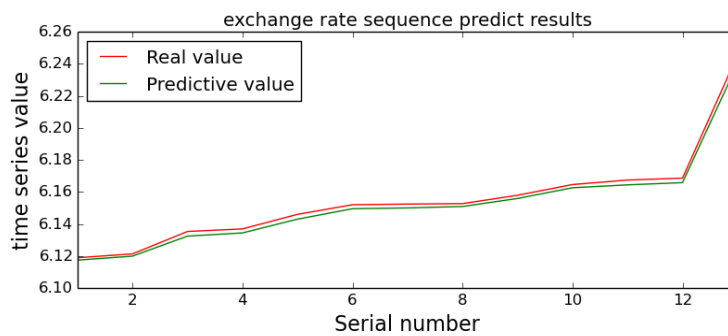


**Figure 4.3. Multi-Step Prediction Results of NASDAQ Sequence**





**Figure 4.4. Multi-Step Prediction Results of WIF Sequence**



**Figure 4.5. Multi-Step Prediction Results of Exchange Rate Sequence**

We also conducted a comparison experiment, one result is obtained by using MANNP model to predict the shipping price index time series, while another result is obtained by using traditional single ANN model. Experiment results are shown in Table 1. The first column is data number, the second column is the real value, the third column is the prediction value produced by MANNP, and the fourth column shows the prediction value by Single ANN.

**Table 1. MANNP and Single ANN Prediction Results.**

No.	Real Value	MANNP Prediction	Single ANN Prediction
1	2670	2665	2705
2	2430	2400	2505
3	2900	2880	2800
4	2600	2610	2706
5	2460	2470	2610
6	2100	2095	2200
7	2200	2300	2000
8	2460	2515	2630
9	2340	2310	2245
10	2840	2912	2760
11	2800	2750	2890
12	2700	2760	2915
13	2850	2830	2700
14	2915	2940	2845
15	2800	2760	2950
16	2940	2970	3015

The corresponding diagram is shown in Figure 4.6. The red curve represents the real value, the green one represents predictive value of MANNP, the purple one represents the predictive value of single neural networks. The diagram shows that the MANNP model prediction result is more accurate than the single ANN model prediction result. Predictive value by MANNP and the real value are very close. It shows the accuracy of the prediction created by our model.

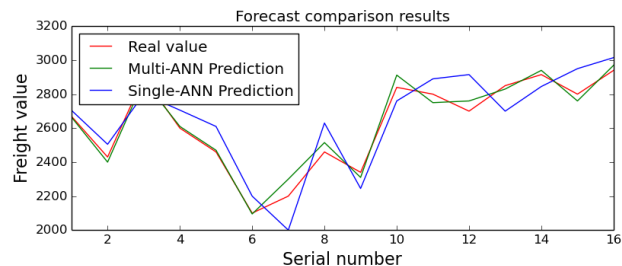


Figure 4.6 MANNP Model Prediction Results

## 5. Conclusion

In this paper, we proposed a novel time series prediction method, MANNP model. This model can successfully predict the time series by providing it with data of the relevant factors. It can be concluded that, prediction made by MANNP can be used as an effective time series analysis and prediction tools. We will keep working on the improvements of MANNP. In the future, this method can be extended to other areas such as financial market or weather forecasts for uncertainty modeling.

## Acknowledgements

This work was supported by Shanghai Municipal Sci&Tech Commission “Science and Technology Innovation Action Plan” project (14511107400), and by Shanghai Maritime University research fund project (No. 20130469), and by State Oceanic Administration China research fund project (201305026).

## References

- [1] N. Davey, S. P. Hunt and R. J. Frank, “Time Series Prediction and Neural Networks”, *Journal of Intelligent & Robotic Systems*, vol. 31, no. 1-3, (1999), pp. 91-103.
- [2] A. S. Weigend and N. A. Gershenfeld, “Time Series Prediction: Forecasting the Future and Understanding the Past”, *Santa Fe Institute Studies in the Sciences of Complexity*, vol. 10, no. 94, (1994).
- [3] N. A. Gershenfeld, A. S. Weigend and N. A. Gershenfeld, “The Future of Time Series: Learning and Understanding”, *Working Papers*, (1993).
- [4] H. O. A. Wold and P. A. Whittle, “Study in the Analysis of Stationary Time-Series”, *Biometrika*, vol. 15, no. 1, (1954), pp. 125-139.
- [5] G. E. P. Box and G. M. Jenkins, “Time Series Analysis Forecasting and Control”, *Journal of Time*, vol. 3, no. 2, (1970), pp. 199-201.
- [6] D. S. Broomhead and D. Lowe, “Multi-variable functional interpolation and adaptive networks”, *Proceedings Ijcai*, vol. 2, no. 3, (1988), pp. 321-355.
- [7] R. J. Erb, “Introduction to Backpropagation Neural Network Computation”, *Pharmaceutical Research*, vol. 10, no. 2, (1993), pp. 165-170.
- [8] C. Cortes and V. Vapnik, “Support-Vector Networks[C]// *Machine Learning*”, (1995), pp. 273-297.
- [9] E. Pisoni, M. Farina and C. Carnevale, “Forecasting peak air pollution levels using NARX models”, *Engineering Applications of Artificial Intelligence*, vol. 22, no. (s 4-5), (2009), pp. 593-602.
- [10] D. Ö. Faruk, “A hybrid neural network and ARIMA model for water quality time series prediction”, *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, (2010), pp. 586-594.

- [11] Y. Yu, J. Song and Z. Ren, "A new hyper-parameters selection approach for support vector machines to predict time series[M]//Pervasive Computing and the Networked World", Springer Berlin Heidelberg, (2013), pp. 775-787.
- [12] R. Hrasko, A. G. C. Pacheco and R. A. Krohling, "Time Series Prediction Using Restricted Boltzmann Machines and Backpropagation ☆", Procedia Computer Science, vol. 55, (2015), pp. 990-999.
- [13] R. J. Erb, "Introduction to Backpropagation Neural Network Computation", Pharmaceutical Research, vol. 10, no. 2, (1993), pp. 165-170.
- [14] A. More and M. C. Deo, "Forecasting wind with neural networks", Marine Structures, (2003), vol. 16, no. 1, pp. 35-49.
- [15] I. Aizenberg, L. Sheremetov and L. Villa-Vargas, "Multilayer Neural Network with Multi-Valued Neurons in Time Series Forecasting of Oil Production", Neurocomputing, vol. 8495, (2015), pp. 61-70.
- [16] E. Cadenas and W. River, "Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks", Renewable Energy, vol. 34, no. 1, (2009):274-278.

## Authors

**Fei Li**(1989-), he was born in Shanxi, master candidate, specialized in data mining and machine learning research .

**Jin Liu**(1975-), he was born in Sichuan, associate professor, specialized in web data mining, machine learning and software engineering.

**Lei Kong**(1975-), he was born in Anhui, senior engineer, specialized in web data mining, and software engineering.

