Short-term tidal current prediction based on GA-BP neural network

Xiangshuo Qiao¹, Fengyi Guo¹, and Runfeng Zhang²*

¹College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082 China
²School of Mechanical Engineering, Tianjin University, Tianjin, 300350 China
*Corresponding author: zhangrunfeng@tju.edu.cn

Abstract. Tidal current is a novel type of renewable energy for power supply. Accurate and stable tidal current prediction is an important research area in the field of tidal current energy development. In this paper, the GA-BP Neural Network is studied deeply, and the historical time series data and time factor are adopted to improve the input. Besides, the prediction model of tidal current components is established based on the experiment data. To validate the effectiveness of the presented method, other five popular prediction models are also introduced and compared. The simulation results show that the model established in this paper is superior to other five models and has better prediction ability in terms of two commonly used performance indices.

1. Introduction

Tidal current is the phenomenon of periodic horizontal movement of seawater under the action of celestial gravity. As a kind of renewable energy, tidal current energy has a broad development prospect, which is why we put forward higher and higher requirements for accurate prediction of tidal current.

The attempt of tidal current prediction can be traced back to 1892, when G. H. Darwin first proposed the harmonic model of tidal current at the Royal Society [1]. A. T. Doodson further developed the model and introduced the least square method to determine the parameters [2–3]. At the beginning of the 21st century, machine learning methods were widely used in various fields including tidal current prediction, which replaced the classical model gradually, and the prediction accuracy was improved by milestone. T. L. Lee and D. S. Jeng established an artificial neural network model for tidal current prediction using short-term records in 2002 [4]. Mahda J. Jahrom compared five estimators such as multi perceptron neural network (MLPNN), adaptive neuro-fuzzy inference system (ANFIS) and auto-regressive moving average (ARMAx), and found that the prediction accuracy depended greatly on the optimization of the prediction model [5]. Abdollah Kavousi-Fard proposed an accurate method based on support vector regression (SVR) and auto-regressive integrated moving average (ARIMA), and finally obtained a hybrid model to predict the tidal current with better performance than artificial neural network, ARIMA, genetic algorithm, and SVR [6]. The training process of SVM and ELM has been improved as while. In 2017, Guozhen Yang introduced univariate feature selection and f-test method [7], and Safari Nima proposed a non-parametric prediction interval (NPI)-based uncertainty model [8]. Kavousi-Fard Abdollah introduced a univariate prognostic approach with better predictive accuracy based on wavelet
transform and support vector regression (SVR) [9]. As to Ng and Betty, their paper proposed the technique of rotary empirical orthogonal functions (EOF), and brought a fast way to predict the tidal currents simultaneously at many locations was developed [10].

It is worth mentioning that similar trends exist in the prediction of other renewable energy resources. For instance, Wang et al. proposed a wind speed prediction method based on improved empirical mode decomposition (EMD) and GA-BP neural network, and a series of white noises were introduced into the raw data to improve the accuracy [11]. By combining kernel function with ELM, Feng et al. improved the prediction accuracy and adaptability of wind power prediction model as well [12].

2. Principle of GA-BP neural network

2.1. Back propagation neural network

Back Propagation Neural Network (BPNN) is a multi-layer feedforward neural network trained by the error backward propagation algorithm, which was proposed by scientists led by Rumelhart and McClelland in 1986. Currently, it is widely used in function approximation, classification application, pattern recognition and data compression, for its strong multidimensional function mapping ability.

This kind of network is composed of input layer, hidden layer and output layer, and it can learn the unknown mapping relation of a large number of input and output modes autonomously, and continuously modify the connection threshold value and weight value through reverse propagation, so as to minimize the network error. The topology of BP neural network is shown in Figure 1.

2.2. Genetic algorithm

Genetic Algorithm (GA), firstly initializes the population, and then selects the optimal fitness function. It continuously simulates the selection, crossover and mutation process of biological heredity in nature, and screens out individuals with good adaptability according to their health status. This process is repeated until the requirements are met, at which point the individuals in the population reach the optimum. The basic elements of GA-BPNN include population initialization, fitness function determination, and mutation genetic operation.
2.2.1. Population initialization
There are two ways of chromosome coding: binary coding and real coding. Traditional binary coding has the mapping error of continuous function discretization, while real coding is more suitable for multidimensional numerical problems, which makes the genetic algorithm closer to the problem space. In this paper, individual is coded with real numbers to obtain the initial population.

2.2.2. Fitness function determination
Fitness function is a criterion to distinguish the good and bad of individuals in a group according to the objective function. Specifically, individual fitness value \( F \) is defined as the sum of the absolute value of the error between the individual predicted output and the actual output, and its calculation formula is as follows:

\[
F = k \left( \sum_{i=1}^{n} \text{abs}(y_i - a_i) \right)
\]

where \( k \) is the coefficient, \( n \) is the number of output nodes, \( y_i \) is the \( i \)-th expected output, and \( a_i \) is the \( i \)-th predicted output.

2.2.3. Genetic operation
In genetic algorithms, Roulette Wheel Selection is used for selection operations. The probability \( P_i \) of individual \( i \) being selected is proportional to the value of its fitness, that is:

\[
P_i = \frac{F_i}{\sum_{j=1}^{N} F_j}
\]

where \( F_i \) is the fitness value of individual \( i \); \( N \) is the number of individuals in the population.

Crossover is the transfer of good gene combinations from one generation to the next. A gene position of a pair of individuals is randomly selected as the crossover position to form a new excellent individual. The mathematical expression of this process is as follows:

\[
\begin{align*}
X_i' &= X_i m + X_i (1 - m) \\
X_j' &= X_j m + X_j (1 - m)
\end{align*}
\]

where \( X_i' \) and \( X_j' \) are two individuals after crossover operation, \( X_i \) and \( X_j \) are two individuals before crossover operation, and \( m \) is a random number evenly distributed between \([0, 1]\).

The mutation operation is to change the gene value of one or more genomes as its allele with a certain probability for individuals in the population. The following is a possible solution:

\[
X_j = X_{\min} + \beta(X_{\max} - X_{\min})
\]

where \( X_{\max} \) and \( X_{\min} \) are the maximum and minimum values of the initial individual respectively, \( X_j \) is the new gene after mutation, and \( \beta \) is a random number evenly distributed between \([0, 1]\). Particularly, we can also use the following formula for mutation operation:

\[
\eta(g) = \lambda \left( 1 - \frac{g}{G_{\max}} \right)^2
\]

\[
X_j = \begin{cases} 
X_j + (X_j - X_{\max}) \eta(g), & r > 0.5 \\
X_j + (X_{\min} - X_j) \eta(g), & r \leq 0.5
\end{cases}
\]

where \( X_{\max} \) and \( X_{\min} \) have the same meaning as in (4), \( \lambda \) is a random number, \( r \) is a random number evenly distributed between \([0, 1]\), \( g \) is the number of iterations and \( G_{\max} \) is the maximum number of iterations.

2.3. GA-BP neural network
Genetic Algorithm (GA) with its strong macro search ability and good global optimization performance can achieve global optimization ideally. Therefore, in the specific search process, using
Genetic Algorithm to optimize Back Propagation Neural Network (GA-BPNN) can avoid the trap of local optimal solution, which just makes up for the shortcomings of BPNN.

GA-BP Neural Network mainly consists of two parts: BP neural network and genetic algorithm. This algorithm is used to optimize the initial weights and thresholds value of BPNN, whose basic idea is to use individual on behalf of the network's initial weights and thresholds and to use the prediction error of BPNN on behalf of the fitness value of individuals. The model looks for the best individual through genetic operation such as selection, crossover and mutation. Then, that process will be repeated until the requirements of the prediction error are met or the maximum number of iterations is achieved. Finally, the optimal prediction results are obtained. The flow of GA-BP Neural Network is shown in Figure 2.

![Flow chart of GA-BP Neural Network algorithm.](image)

3. Simulation studies and analysis

3.1. Data selection and processing
Since the instantaneous velocity of tidal current is random and unstable, the average value in every ten minutes is taken as the historical data in this paper. The time factor should also be considered because of the non-negligible influence of it on the flow speed. The process of data selection and processing can be described as follows:

Step 1: The raw data is given in terms of Length Set $V$ and Angle Set $\theta$. We take the east-west direction (EWD) and the north-south direction (NSD) as the directions of the two components, as follows:

$$X = V \cos \theta$$

$$Y = V \sin \theta$$

where $X$ is the EWD component and $Y$ is the NSD component.
Step 2: Due to the relative independence of the two components, this paper adopts the same method for data processing of the two components. The following is an example of this process with EWD data. Define the input as:

\[ I(t,T) = [X(t-T-1) \ldots X(t-1) \ t'] \]

where \( t' \) is the time factor after normalization and \( T \) is the time tag. Then \( f_p(I(t,T)) \) is used to predict the observation value \( X(t) \) at time \( t \).

Step 3: Finally, the data will be segmented to obtain the training set and the test set.

3.2. Data sources

The measured tidal current data from Zhoushan port area are used to train and test the model. Specifically, the historical data comes from the values of four consecutive days and 24 consecutive hours a day in March 2020. The sampling interval of the data was 10 min, with a total of 566 sampling points.

3.3. Simulated analysis

The first 520 data were taken as the training set of the prediction model, and the remaining 46 data were taken as the test set. With time delay \( T=8 \), the prediction results of BPNN, GA-BPNN, Support Vector Machine based on Whale Optimization Algorithm (WOA-SVM), Extreme Learning Machine (ELM), Wavelet Neural Network (WNN) and Long Short-Term Memory (LSTM) were compared. The simulation results of each model are shown in Figures 3 and 4.

![Fig. 3. Comparison of predicted and true values in the east-west direction.](image)

![Fig. 4. Comparison of predicted and true values in the north-south direction.](image)

From two figures, it can be observed that the tidal current curve with sharp peak and valley characteristics is not a smooth transition curve, which is caused by the randomness and volatility of
the tidal current. Due to some geographical reasons, the overall flow direction of the observation point is closer to the NSD, and the smaller component value in the EWD leads to more fluctuations.

All the five prediction models can reflect the trend of tidal current with time to some extent, and the predicted trend is broadly in line with the real. In order to compare the accuracy of different models in the prediction of tidal current, it is necessary to establish performance indexes and analyse the error quantitatively.

3.4. Performance indicators
In order to better evaluate the prediction accuracy of each model, root mean square error (RMSE) and mean absolute error (MAE) are selected as the evaluation criteria. The two indicators are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (f(I(t,T)) - X(t))^2} \quad (10)$$
$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |f(I(t,T)) - X(t)| \quad (11)$$

where $N$ is the total number of predicted data points. In general, we expect smaller RMSE and MAE to achieve higher prediction accuracy. Tables 1 and 2 list the average RMSE and MAE of six common tidal current prediction models running independently for 20 times respectively.

When comparing the prediction results of GA-BPNN model with other models, it is shown that GA-BPNN realizes the highest prediction accuracy with average MAE values of 0.041 in EWD and 0.050 in NSD, and RMSE values of 0.052 in EWD and 0.066 in NSD, respectively.

It should be noted that although there are significant numerical range differences between the two directions of the tidal current data, the prediction errors of most models tend to be consistent. Compared with the other five models, LSTM has the worst stability, with RMSE reaching 0.206 in NSD, which is a very unsatisfactory value.

| Table 1. RMSE of different models. |
|-----------------------------------|
|       | BPNN | GA-BPNN | WOA-SVM | ELM | WNN | LSTM |
| EWD   | 0.068| 0.052 | 0.069 | 0.065 | 0.077 | 0.083 |
| NSD   | 0.074| 0.066 | 0.078 | 0.077 | 0.080 | 0.206 |

| Table 2. MAE of different models. |
|-----------------------------------|
|       | BPNN | GA-BPNN | WOA-SVM | ELM | WNN | LSTM |
| EWD   | 0.053| 0.041 | 0.055 | 0.052 | 0.062 | 0.069 |
| NSD   | 0.064| 0.050 | 0.060 | 0.061 | 0.066 | 0.175 |

4. Conclusion
In this paper, GA is used to optimize the traditional BP neural network, and the prediction model of tidal current component is established. In order to verify the validity of the model, this paper introduces two performance indicators and compares with other prediction algorithms. Simulation shows that GA-BPNN has higher prediction accuracy than BPNN, WOA-SVM, ELM, WNN and LSTM. Therefore, this model can provide some reference value for practical application.

For the future work, it is a meaningful topic to establish integrated learning model of tidal current prediction instead of single learning model. The obvious hysteresis effect of GA-BPNN needs to be solved as well.

References
[1] G. H. Darwin, Proc. Roy. Soc. Lond. A 52, 345-376(1892)
[2] A. T. Doodson, I. H. R 33, 85-126(1957)
[3] A. T. Doodson, Proc. Roy. Soc. Lond. A 100, 305-329(1921)
[4] T. L. Lee, D. S. Jeng, Ocean Eng 29, 1003-1022(2002)
[5] Mahda J. Jahromi, Ali I. Maswood, Tseng K. J., P. E. S. G. M, 1-8(2010)
[6] A. Kavousi-Fard, IEEE T. Geosci. Remote 55, 112-118(2017)
[7] Guozhen Yang, Haifeng Wang, Hui Qian, Jianming Fang, IOP Conference Series: Materials Science and Engineering 199, 12024(2017)
[8] N. Safari, S. M. Mazhari, B. Khorramdel, C. Y. Chung, IEEE T. Sustain. Energ 10, 748-758(2019)
[9] A. Kavousi-Fard, W. Su, IEEE T. Geosci. Remote 55, 3108-3114(2017)
[10] Ng, Betty, J. Atmos. Ocean. Tech 10, 868-879(1993)
[11] Shouxiang Wang, Na Zhang, Lei Wu, Yamin Wang, Renew. Energ 94, 629-636(2016)
[12] Junjie Feng, Xueqiang Fan, Zhifei Zhang, Pengda Yan, Xing Guo, Xiaohui Zhang, Electric Age 11, 29-31(2018)