Journal of Computers Vol. 32 No. 2, 2021, pp. 86-98
doi:10.3966/199115992021043202008

Water Evaluation Based on Multi-source K-Means Combination GA-BP

Liwen Chen, Rijing Zheng*, Jing Chen, M. Shoaib

College of Information Science and Engineering, Fujian University of Technology, Fuzhou, Fujian, 350118, China
zrj13950713626@163.com

Received 16 April 2020; Revised 27 September 2020; Accepted 21 October 2020

Abstract. The water quality parameters in the water environment are non-linear, random and dependent. The prediction accuracy and robustness of traditional water quality prediction models are generally not high. In order to optimize and improve the prediction accuracy of water quality prediction models, this paper proposes a multi-source K-means clustering combination GA-BP neural network is used to study the dynamic classification and prediction of water quality. Integrated learning uses multiple learning algorithms to obtain better prediction performance than traditional single learning algorithms. First, the water quality elements are classified according to similarity through multi-source K-means clustering, and then the weight of each element is calculated through the classification results. The GA-BP neural network is used to predict the changes of various elements of water quality. The application of the example of 36 feet Lake in Pingtan, China shows that the method is effective and feasible, and the accuracy of prediction is obviously improved which is helpful for analyzing the water quality of 36 feet Lake.

Keywords: multi-source K-means, GA-BP water-quality-classification, optimization combination

1 Introduction

With the development of artificial intelligence and Internet of Things technology, the evaluation of water quality based on advanced machine learning and data analysis technology has become a current research hotspot. The theories of water quality evaluation can be roughly summarized into five categories: water quality simulation method, mathematical statistics method, grey system theory method, chaos theory method and neural network model [1-3]. The above methods are used in different regions, and the results are diversity in the same reach. Compared with other mathematical models, machine learning has better applicability to non-linear and complex problems, so it can better reflect the inherent change and fluctuation characteristics of water quality parameters. Kim and Seo [4]combined clustering algorithm with artificial neural network to reduce the impact of unbalanced training data set; Wang Yanfeng et al. [5] used artificial neural network and comprehensive pollution index to compare the same example, and the results showed that artificial neural network is more adaptable used in water quality evaluation. Shi et al. [6] combined the wavelet analysis with the artificial neural network, and improved the prediction accuracy through wavelet de-noising on the water quality time series; Zhai Wei et al. [7] coupled gray prediction theory and artificial neural network methods to solve the problem of lagging response to water quality time series changes and improve the accuracy of water quality prediction. Due to the slow convergence speed of the neural network model, it is easy to produce over-fitting phenomenon in the training process [8], most of the researches use a variety of coupling methods to improve the water quality prediction accuracy. As an optimization solution method that imitates the biological genetic and evolution phenomena in the natural environment, genetic algorithm can effectively avoid the problems of

* Corresponding Author

falling into local minima and slow convergence speed by optimizing the initial weight parameters of artificial neural network.

This research attempts to apply the multi-source K-Means model to water quality evaluation, divide historical daily samples into several categories. Then classify and identify elements with sample characteristics. In addition, the outstanding optimization performance of genetic algorithm by establishing the objective function and constraint conditions, the genetic algorithm is used to optimize the connection weight of the input layer and the hidden layer of the BP neural network and the threshold value of the hidden layer, so that the model can improve the water quality. The accuracy of the comprehensive evaluation.

2 Classification Algorithm Principle

In order to choose a better algorithm with a single effect, the classification algorithms with higher usage rates—KNN algorithm, linear regression, decision tree, K-Means, and prediction algorithms—BP neural network 、XGBoost and LSTM were selected.

2.1 KNN

K-NN (k-nearest neighbor) was proposed by cover and Hart at 1968. Its basic idea is to calculate the distance between the tested sample and training sample by selecting the appropriate function [9]. Select the K training samples with the smallest distance to vote, the category with the highest number as the final prediction category of the tested sample.

Suppose two n-dimensional vectors:

$$a\left(x_{11},x_{12},\ldots,x_{1n}\right). \tag{1}$$

$$b\left(y_{21},y_{22},\ldots,y_{2n}\right). \tag{2}$$

$$d(a,b)=\sqrt{\sum_{k=1}^{n}(x_{1k}-y_{1k})^2}\ . \tag{3}$$

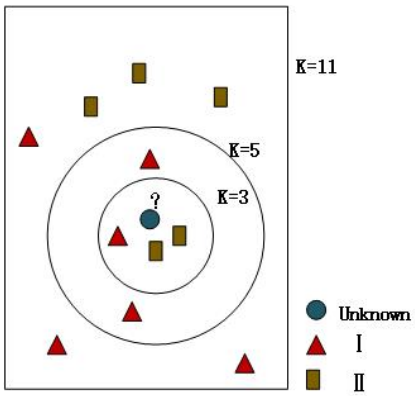


Fig. 1. Schematic diagram of KNN algorithm classification

It can be seen that in the test sample, when K = 3 is the closest to the unknown data, it is a box, so it is determined as a box. When K = 5, the closest is a triangle, so it is determined as a triangle.

2.2 Linear Regression Algorithm

Linear regression is a basic machine learning regression algorithm. It is a regression analysis that uses linear regression equations to model the relationship between one or more independent [10] and dependent variables.

Input training data set T, output samples corresponding to sample feature vectors x and label y.

The model of the linear regression algorithm is:

$$T = \{(x_1, y_1), (x_2, y_2) ... (x_n, y_n)\} . \tag{4}$$

$$y = w * x + b . \tag{5}$$

$x \in R^n$ is input feature vector, $y \in R$ output labels, $w \in R^n$ $b \in R$ model parameters.

The loss function calculation equation is:

$$L(w,b) = \frac{1}{2N} \sum_{i=1}^{N} [y_i - x_i]^2 . \tag{6}$$

2.3 Decision Tree

Decision Tree is a tree-structured classification and regression algorithm consisting of nodes and directed edges, as shown in Fig 2. Rectangle nodes are internal nodes, which represent certain features or attributes: circular nodes are leaf nodes, which represent a certain category of samples [11]. In the process of decision tree generation, node splitting needs to be divided according to the optimal characteristics of the training data on this node.

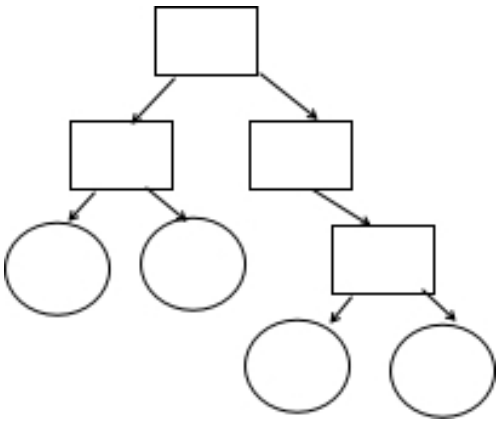


Fig. 2. Decision tree model

3 Prediction Algorithm Principle

3.1 BP Neural Network

The artificial neural network is a model that simulates neuronal signal transmission in the human brain. The artificial neural network is a network structure, and the connection layers are in parallel relationship. Each layer contains basic operating units, called nodes [12]. As shown in Fig. 3, the network structure includes three layers: an input layer, a hidden layer, and an output layer. The input data is stored in the input layer node, and the data is transmitted to the output layer through the hidden layer. The nodes of each layer are connected to each node of the adjacent layer. The input data of a node in each layer is the non-linear conversion of the output data of each node in the previous layer. The neural network containing a hidden layer can represent arbitrary inputs. Output mapping [13]. The input data of each node in the hidden layer is transformed by the transfer function to obtain the output data of the node. The transfer function usually uses the Sigmoid function.

BP neural network is a kind of feed-forward neural network, which adopts the back-propagation algorithm, namely BP (Back-Propagation) algorithm.

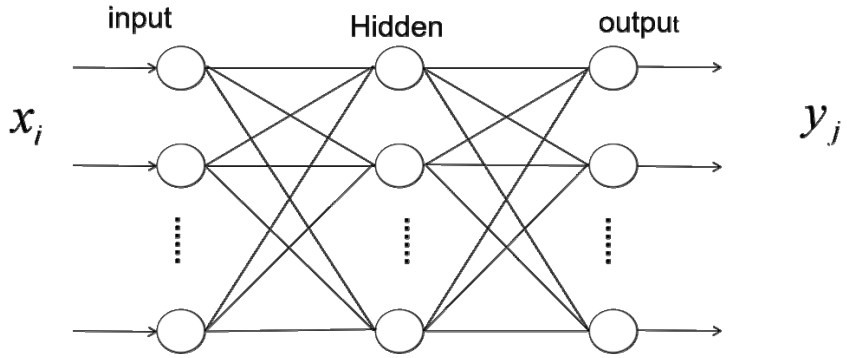


Fig. 3. BP neural network structure

If BP neural network input $x_1, x_2, ...x_n$, the input of the i-th node of the hidden layer is

$$nu_i = \sum_{j-1}^{M} \omega_{ij} x_j + \partial_i . \tag{7}$$

ω_{ij} Weight between the i-th node in the hidden layer and the j-th node in the input layer ; ∂_{ij} threshold representing the i-th node of the hidden layer.

The output of the i-th node of the hidden layer is

$$y_i = \varphi(nu_i) = \psi(\sum_{j=1}^{M} \omega_{ij} x_j + \partial_i) . \tag{8}$$

$\psi(x)$ excitation function representing the hidden layer $\varphi(x) = 1/(1+e^{-x})$., The input of the Kth node of the output layer is

$$nu_k = \sum_{i=1}^{q} \omega_{ki} y_i + \alpha_k . \tag{9}$$

α_k Shows the threshold of the K-th node of the output layer.

The output of the K node of the output layer is:

$$\alpha_k = \psi(nu_k) = \psi(\sum_{i=1}^{q} \omega_{ki} y_i + \alpha_k) . \tag{10}$$

$\psi(x)$ excitation function representing the output layer $\psi(x) = \frac{1}{1+e^{-x}}$.

The BP neural network algorithm uses gradient descent learning. For P training samples, the output error is:

$$E = \frac{1}{2} \sum_{P=1}^{P} \sum_{P=1}^{L} (T_k^{\phi} - o_k)^2 . \tag{11}$$

T_k expected output, o_k actual output

According to the error gradient descent rule, the correction amount of the output layer weights is sequentially corrected as $\Delta\omega_{ki}$, output layer threshold correction $\Delta\alpha_k$, Correction of hidden layer weights $\Delta\omega_{ij}$, Output layer threshold correction $\Delta\theta_i$.

$$\Delta\omega_{ki} = -C\frac{\partial E}{\partial\omega_{ki}} \quad \Delta\partial_k = -C\frac{\partial E}{\partial_k} \quad \Delta\omega_{ki} = -C\frac{\partial E}{\partial\omega_{ij}} \quad \Delta\omega_{ki} = -C\frac{\partial E}{\partial\omega_{ij}} . \tag{12}$$

3.2 XGBoost

XGBoost is an improved GDBT algorithm proposed by Friedman et al in 2001. It is an iterative decision tree algorithm, which uses both first and second derivatives. Each iteration uses the idea of gradient descent [14]. Based on all the trees generated in the previous step, it proceeds in the direction of minimizing the given objective function. Generate a new tree, and finally achieve a more accurate classification effect.

3.3 LSTM

LSTM neural network unit has a tuple (Cell), which is also a memory unit. The state of c_t at time t. By operating on the Input Gate, Forget Gate and Output Gate, can read and modify the memory unit of the LSTM neural network. This operation is generally implemented using sigmoid or tanh functions. The working process of the LSTM unit is: at each moment, the LSTM unit receives input of two types of external information x_t, the current state and the hidden state of the LSTM at the previous h_{t-1}. In addition, each door also receives an internal information input, which is the state of the memory unit $c_{t\text{-}1}$. After receiving the input information, each gate will operate on the input from different sources, and a logic function determines whether it is activated [15].

4 Multi-source Binary K-Means Combined GA-BP Neural Network

4.1 Multi-source Binary K-Means

The clustering algorithm belongs to unsupervised learning, and does not require a clear label to perform good classification. Cluster analysis groups the objects based on the information of data describing the objects and their relationships. Its purpose is to make objects in a group similar to each other, but different in othoer. The greater the similarity within the group and the greater the gap between the groups, the better the clustering [16]. The clustering algorithm adopts K-Means to achieve classification. This is a centroid-based technology. First, K centroids are randomly selected, and the selection of K value is the parameter given by itself, that is the desired cluster (classification) Number. After the centroid is determined, the distance from each point to the centroid is calculated, and the closer points are classified into a cluster. Then repeat the above steps, change the centroid and update until the centroid no longer changes, or the change is small. It is considered that the algorithm has reached convergence and the classification is completed. The main steps are as follows:

(1) Select the specific clustering center, that is, the value of K.

(2) Calculate the Euclidean distance of each sample data point to the K class center points in turn, and classify each sample into the class with the class center point closest to it.

(3) Recalculate the center of each class generated in step 2 and adjust the center point.

(4) Repeat steps 2 and 3 until the centroid is no longer moving that is, the algorithm has reached convergence.

(5) Calculate the contour coefficient and the statistics of each group to determine whether the classification is reasonable.

Calculate the distance from each point to the cluster experience. Its purpose is to make:

$$SSE = \sum_{i=1}^{k} \sum_{p} \in C_i dist(p, c_i)^2 . \quad \textbf{(13)}$$

SSE is the sum of the squares of the deviations of all the objects in the data set, P is the point of the cluster and space. c_i is the cente of C_i. For each object in each cluster, the square of the distance between each data point and the center point of its cluster class is required and then summed to make the resulting cluster more compact and different clusters more independent.

Qualitative analysis is performed to determine whether the classification is satisfactory according to the numerical difference of each group of classifications, and it can also be determined by calculating the contour coefficient. The contour coefficient reflects the degree of separation between different cluster classes and the degree of aggregation of the same cluster class.

Step 1: calculate the degree of aggregation of the same cluster class

$$A(O)=\frac{\sum_{O'\in C_i,o\neq o'} dist(o,o')}{|C_i|-1}. \tag{14}$$

Step 2: Calculate the degree of dispersion between different clusters

$$B(o)=\min_{C_j:1\le j\le k,k\neq i}\left\{\frac{\sum' o\in C_i dist(o,o')}{|C_j|}\right\}. \tag{15}$$

Step 3: calculate the profile coefficient

$$S(o)=\frac{A(o)-B(o)}{\max\{A(o),B(o)\}}. \tag{16}$$

D is the entire data set, C_i divided cluster class, o a data point is an object. A is the degree of aggregation of the same cluster class. The smaller the value, the more compact of the cluster. It is the degree of separation of different clusters. The larger the B value, the greater the degree of separation between clusters. The effect of classification S closer to 1 is that the degree of separation between clusters is large and the degree of data aggregation within each cluster is also large.

4.2 GA-BP Neural Network

Genetic Algorithm (GA) is a search heuristic algorithm for solving optimization problems. It has scalability and can be easily combined with other algorithms to produce a hybrid algorithm that combines the advantages of both parties. Therefore, for the defects of the BP neural network, GA is used to optimize the initial value and threshold of the BP neural network, so that the optimized BP neural network can be better trained and predicted [17]. The basic idea is: the individual represents the initial weight and threshold of the network, and the prediction error of the BP neural network initialized by the individual value is used as the fitness value of the individual. Then the optimal individual is found through selection, crossover, and mutation operations, that is the optimal BP neural network initial Weights and thresholds.

The main steps of optimizing the BP neural network by GA are as follows: the structure design of the BP neural network →GA parameter setting→Prediction of BP Neural Network. The specific algorithm flow is shown in Fig. 4.

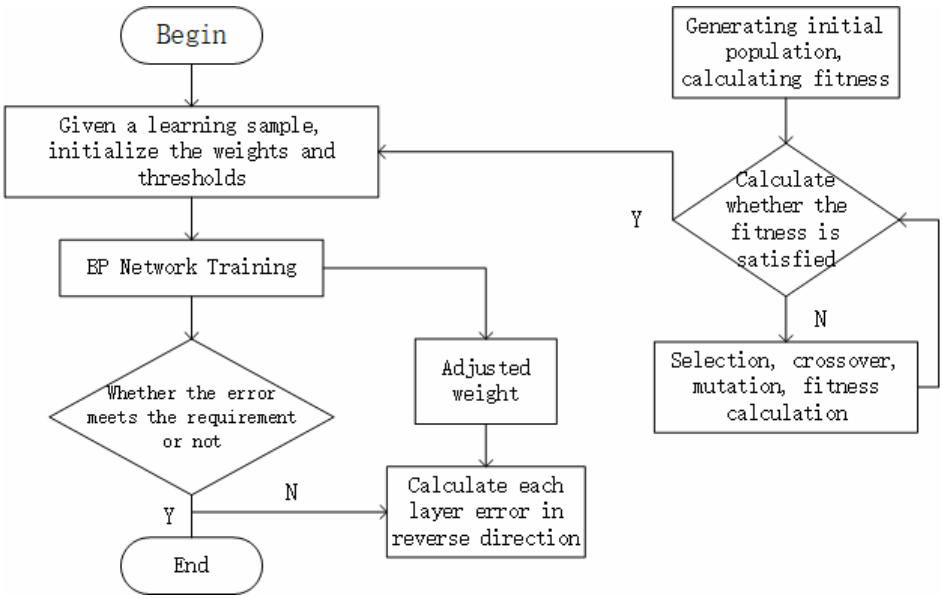


Fig. 4. GA-BP flow chart

Step 1: Population initialization. Individual coding uses the method of real number coding, and each individual is a real number string. The real number string consists of the weight and threshold of the input layer to the hidden layer and the weight and threshold of the hidden layer to the output layer. Set evolution algebra and maximum evolution algebra, randomly generate the initial population.

Step 2: Obtain the body evaluation function. The initial weights and thresholds of the BP neural network are obtained from the values of the initial population in Step 1. Use the prediction error value obtained after BP network training as the individual fitness function:

$$F = k \sum_{i=1}^{m} | y_i - o_i | . \tag{17}$$

F is the individual fitness function, m is the number of nodes in the output layer, i is the number of nodes, and y_i is the expected output of the i-th neural network; o_i is the predicted output for the first node of the network; K is a constant.

Step 3: Select the operation. According to the individual fitness, the roulette method was used to select excellent individuals to inherit to the next population. The specific strategy chosen is to use the fitness ratio as the benchmark, then the selection probability of each j is as follows:

$$p_j = \frac{f_j}{\sum_{n=1}^{p} f_j} \quad n = 1, 2, 3...., p . \tag{18}$$

$$f_j = \frac{k}{F_j}. \tag{19}$$

p_j is the selection probability of the j, f_j is the fitness of the j, F_j is the fitness function of the j-th individual.

Step 4: Crossover calculation. Because the individual uses the original number coding method in Step 1, the cross method also uses the original number cross method. For selected pairs of individuals, swap some of the chromosomes between them, thereby generating new individuals:

$$\begin{aligned} a_{xi} &= a_{xi}(1-b) + a_{yi}b \\ a_{yi} &= a_{yi}(1-b) + a_{xi}b \end{aligned} . \tag{20}$$

a_{xi} a_{yi} position of the x and y chromosomes, b is the random number between [0, 1].

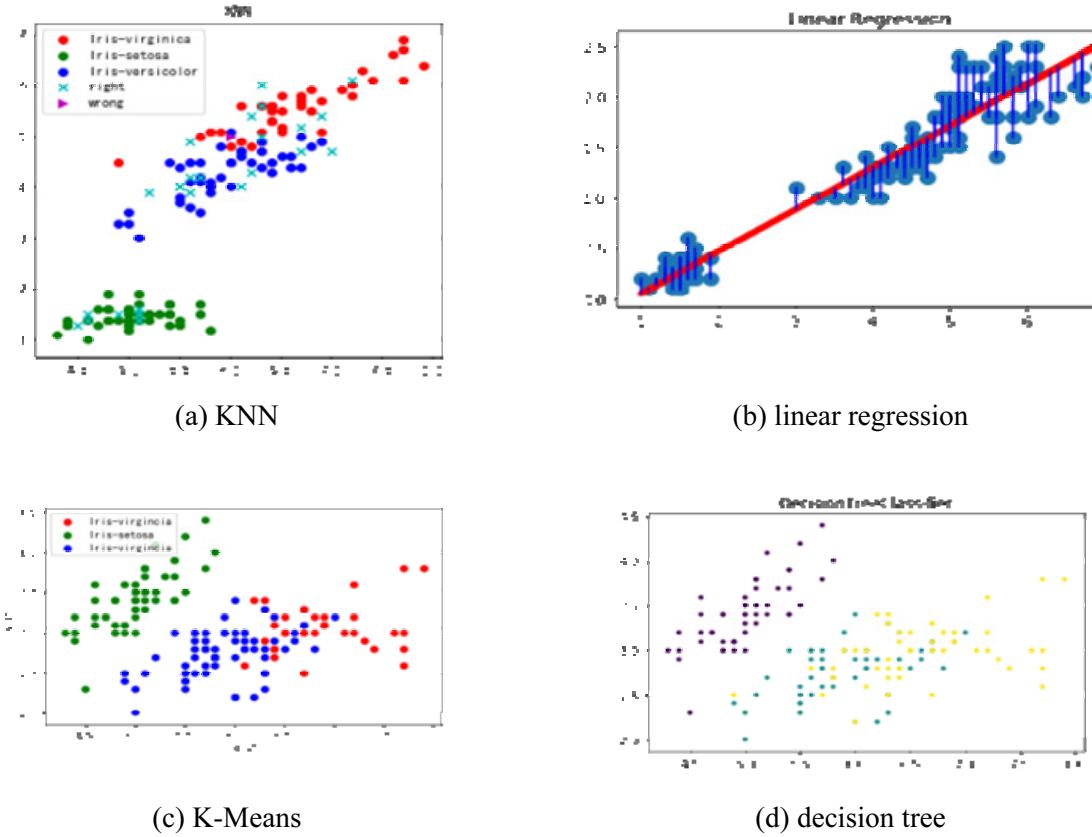
Step 5: mutation operation. For selected individuals, change the value of one or some genes to other alleles with a certain probability

Step 6: Replace the randomly selected initial value during the training with the optimal individual obtained by GA optimization, and use it as the new weight and threshold value during the training of the BP network. Then a new round of training of the BP network can get the optimal results.

5 Algorithm Test

Results obtained by KNN, linear regression, decision tree, K-Means classification.

In (a), the K value is 3, and the accuracy in the classification result is about 95%. In (b), the selected data and its corresponding category are calculated using the linear regression adaptive function to calculate the corresponding coefficients. The results show that they all fluctuate between 1, but the classification results fluctuate greatly; In (c) K-Means updates the cluster class by continuously generating the center point of the cluster. Iteration to the center point no longer changes, and the classification result It is clear and the accuracy rate can reach about 95%; In (d) the decision tree model is trained based on the divided data set. The experimental results show that the dispersion of each category is large and the classification result is not clear.



(a) KNN (b) linear regression

(c) K-Means (d) decision tree

Fig. 5. Results from different classifications

From the analysis of the above it is found that because the K-Means algorithm itself is to find potential natural grouping structures in the data, it is efficient and scalable for data sets with obvious attributes such as water qua0At the same time, it is more suitable for water quality evaluation in Pingtan, China.

The results obtained by XG boost, LSTM, GA-BP etc.

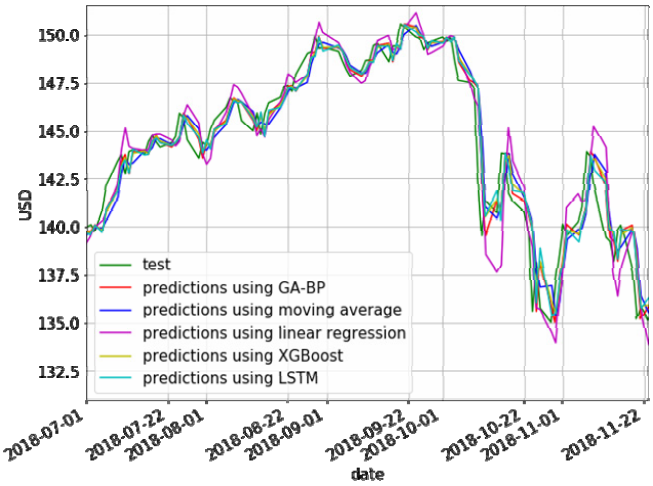


Fig. 6. Comparison of predictive analysis

From the two experimental results in Fig. 9 and Table 1, it is shown that using the root mean square error (RMSE) and the average absolute percentage error (MAPE) for measurement, and it can be seen from the graph that the linear method has the worst prediction effect. Also, It is difficult to judge visually which kind of effect is the best. From the index results, we can see that GA-BP gives the lowest RMSE and MAPE, so the effect is better.

Table 1. Comparison of results of forecast indicators

	method	RMSE	MAPE (%)
0	GA-BP	1.1	0.56
1	Moving Average	1.3	0.64
2	Linear REgression	1.4	0.71
3	XGBoost	1.2	0.58
4	LSTM	0	0.58

6 Experiment and Analysis

6.1 Water Quality Evaluation Index

Water quality evaluation is based on the water quality standards and various index values of the water samples, determine the water quality grade by calculating with a certain mathematical model. The water quality evaluation is based on the "Surface Water Environmental Quality Standards" (GB 3838-2002) as the grading standard, some of the indicators are shown in Table 2. This standard divides the surface water quality standards into 5 categories according to the environmental functions and protection objectives of the surface waters.

Table 2. Some surface water indicators

Index	I	II	III	IV	V
PH			6~9		
dissolved oxygen/ ($mg.L^{-1}$)	>7.5	6	5	3	2
potassium permanganate/ ($mg.L^{-1}$)	2	4	6	10	15
ammonia nitrogen/ ($mg.L^{-1}$)	0.015	0.5	1.0	1.5	2.0
total phosphorus/ ($mg.L^{-1}$)	0.02	0.1	0.2	0.3	0.4

6.2 Data Sources

The 36-foot Lake in Pingtan, China is located in the eastern waters of Fujian Province. It is the largest natural freshwater lake in Fujian Province that has a total area of 2.1 million square meters, a water storage capacity of 12.9 million cubic meters, a maximum water depth of 16 meters. It is located at 25° 15'~25°45' north latitude and 119°32'~120° longitude east..

As shown in Fig. 7, on the 36-foot lake in Pingtan combining the zoning method, the system (grid) and the professional judgment method to set the dot layout. The collection frequency is once every four hours. The collection parameters include water temperature, pH, dissolved oxygen, conductivity, Turbidity, ammonia nitrogen, potassium permanganate, total phosphorus, total nitrogen, chlorophyll. They are collected by different sensors, transmitted to the detection platform through the communication module. The water quality indicator detection sensor is shown in Table 2. Collection from January to January 2015 The data on August 31, 2019 is used as the data source for the research.

6.3 Data Application and Inspection

After the original data is predicted and processed, it is substituted into the K-Means model and iterated repeatedly, calculate the distance from each point to the cluster experience. The purpose is to make the resulting clusters more compact and independent. The result of classification is shown in Fig. 6. Through the GA-BP neural network model, all the data is divided into 90% training and 10% testing. The prediction results of 10 elements are shown in Fig. 7.

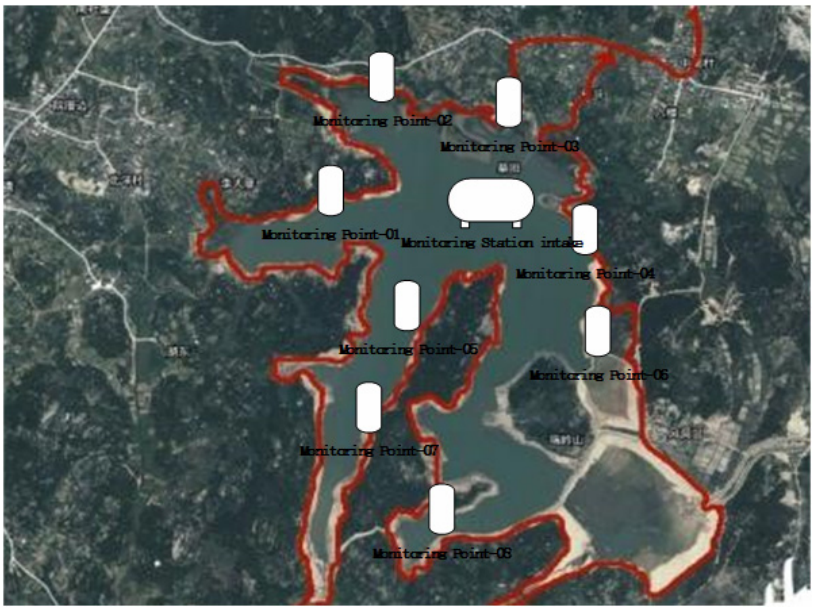


Fig. 7. Distribution of monitoring points

Table 3. Detection of sensor parameters for each index

water-quality index	detecting instrument
turbidity (NUT)	RMD-C6
conductivity (us/cm)	RMD-ISEP10
PH	RMD-H8
dissolved oxygen (mg/L)	RMD-OB-3
water temperature (℃)	RMD-H8
ammonia nitrogen (mg/L)	RMD-ISNH4

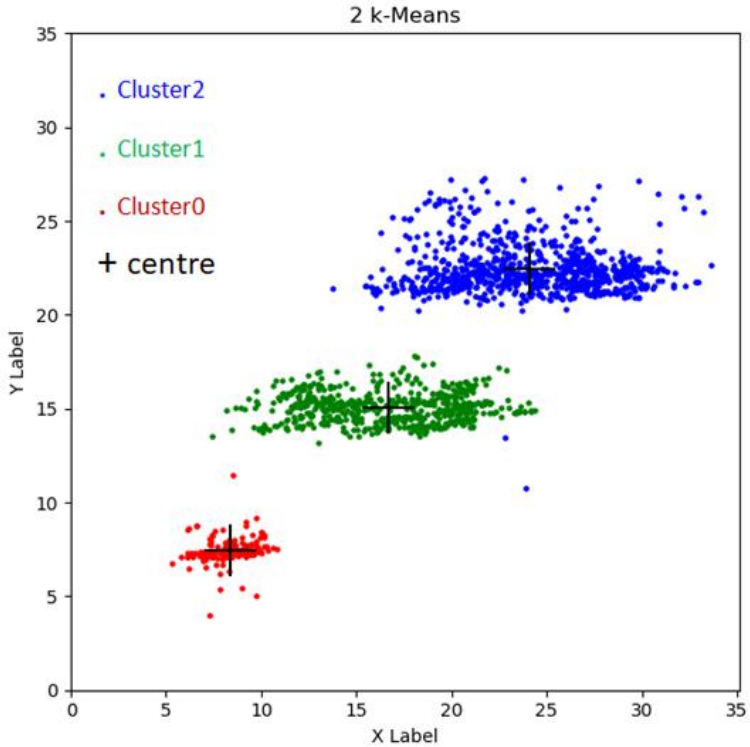


Fig. 8. Classification results

Cluster	tempera ture (℃)	pH	dissolved oxygen (mg/L)	conductivit y(μS/m)	turbidity (NTU)	Ammonia nitrogen (mg/L)	Permanganate index (mg/L)	phosphorus (mg/L)	nitrogen (mg/L)	Chlorophyll (μg/L)
0	1.97	7.53	8.33	3.26	4.51	5.15	2.45	1.97	7.71	2.03
1	2.22	7.48	8.04	1.98	2.41	5.19	2.27	1.93	5.29	1.81
2	2.36	7.44	8.36	8.27	5.29	4.26	2.64	2.32	6.82	1.02

Fig. 9. Means of the central points of each category

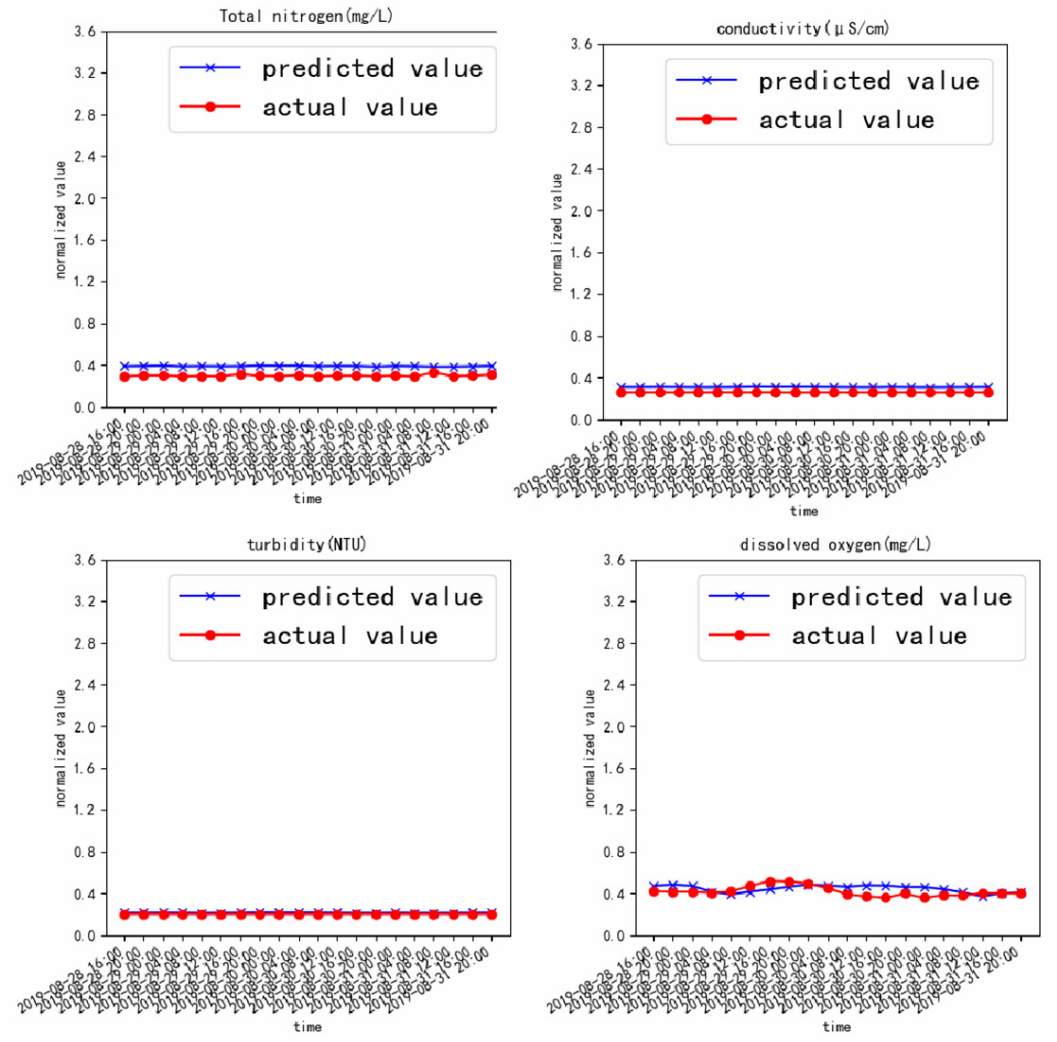


Fig. 10. Forecast trend map

From the Fig. 10, the average conductivity number edited with 0 is 3.26, the conductivity average number 1 is 1.98, and the conductivity average number 2 is 8.27. Conductivity has a clear difference in each category, turbidity and total nitrogen classification effect is also better, other elements in the time period of collection change are not much, the overall analysis of the three categories have significant differences. By calculating the outline coefficient of the classification result is 0.9, which is closer to 1, so there is a clear distinction between the averages of each group in the classification results. The differences between the values in the groups are small, and the classification results are relatively reasonable. At the same time, the traditional classification only uses a single element as a reference. Boundary, multi-source K-Means combines multiple parameters, the classification results are more reasonable, providing a reference for water quality investigations.

According to the classification results of K-Means, the electrical conductivity, turbidity, and total nitrogen have changed significantly in recent years, while the changes in water temperature, pH, dissolved oxygen, ammonia nitrogen, potassium permanganate, total phosphorus, and chlorophyll is small, so the classification results can be referred to in the forecast and analysis of Pingtan water quality.

Through k-means cluster analysis, the changes in conductivity, turbidity, and total nitrogen have been more obvious in recent years. In order to improve the utilization rate of the model, these three types of data are verified in the GA-BP neural network model. Take the concentration data of all pollutants in the previous 1d and the data of a certain pollutant in the previous 5d as the predicted input changes, and output the forecast of a certain pollutant. The predicted concentration level and change trend are in good agreement with the changes in the measured value.

It can be concluded from the prediction results and water quality grades that the water quality of the 36 foot Lake Basin of Pingtan has remained at a high level for many years.

7 Conclusion

In view of the existing problems in the current water quality detection and evaluation method, a Water Quality Classification Prediction Algorithm based on the multi-source di-K-Means combination GA-BP, the neural network model is proposed, and the historical attribute specific information on water quality is classified by using the dichotomy K-Means. The concentration of various parameters of water quality is predicted by the GA-BP neural network model. The algorithm can make full use of the interrelation between the characteristics of water quality, and further improve the accuracy of water quality prediction by relying on the concentration relationship of element turbidity, conductivity and total nitrogen in water quality problems.

Acknowledgements

This work was supported in part by the Fujian Science and Technology Department Grant 2018H01010223, Quanhzou Science and Technology Bureau Grant 2017G012, Fujian University of Technology Grant GY-Z17011, Department of Education of Fujian Province Grant JT180339.

References

[1] Zhong Qian. river restoration schemes based on experimental design and water quality simulation model are preferred. China University of Petroleum (Beijing), 2017.

[2] Zheng Guochen, Zhi Liling, Guan Di, Liu Yu, Sun Yutong, He Xu, Ma Xiaobing. Evaluation of water quality of Xinlicheng reservoir based on grey analysis method (01)(2019) 14-16+86.

[3] Sun Lian, Xin Kunlun, Yan Hei, Tao Tao. Location of pollution source in water supply network based on BP neural network and user complaint information. Water supply and drainage 55(08)(2019) 136-141.

[4] KIM S E, SEO I W. Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers. Journal of Hydro-environment Research (9)(2015) 325-339. DOI: 10.16241/j.cnki.1001-5914.2018. 09.025.

[5] Wang Wanfen, Tan Lugui, Liu Xiaosheng. Environmental quality assessment of Pi River based on fuzzy comprehensive evaluation method. Resource Development and Marketing (05)(2008) 410-412 416.

[6] SHI B, WANG P, JIANG J P, et al. Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomal. Science of the total environment (610-611)(2018) 1390-1399. DOI: 10.1016/j.scitotenv.2017.08.232.

[7] ZHAI W, MAO J, MENG Y D, et al. Study of the prediction of water quality based on artificial neural network with grey theory. South-to-North Water Transfers and Water Science & Technology 18(1)(2020) 138-143.

[8] M Skiba, M Mrówczyńska, A Bazan-KrzywoszańskaModeling the economic dependence between town development policy and increasing energy effectiveness with neural networks. Case study: The town of Zielona Góra. Applied Energy 188(2017) 356-366.

[9] Chen Caisen, Cheng Zhiwei, Wang Yubo, et al. Research on cache timing template attack based on KNN algorithm. Journal of Computers (Taiwan), vol. 31, no. 1, p. 184-191, February, 2020.

[10] Konstantin Posch, Maximilian Arbeiter, Juergen Pilz. A novel Bayesian approach for variable selection in linear regression models (2020) 144.

[11] Zhang Yu, Yang Fei, you Mingxi, Li Guolong, long Yiyu. Research on evaluation model of Track Irregularity Based on gbdt. Railway architecture 60(08)(2020) 111-114.

[12] Binggang Wang. Early Warning Method of Marine Products Network Marketing Risk Based on BP Neural Network Algorithm 103(sp1)(2020) 177-181.

[13] Shaohua Wang, Jianli Jiang, Xiaobing Lu. Study on the Classification of Pulse Signal Based on the BP Neural Network 08(05)(2020) 104-112.

[14] Aijun Deng, Huan Zhang, Wenyan Wang, et al. Developing Computational Model to Predict Protein-Protein Interaction Sites Based on the XGBoost Algorithm 21(7)(2020).

[15] Saleh Abbas, Bahtiar, Wahyono et al. Anomaly Detection in Climate Data Using Stacked and Densely Connected Long Short-Term Memory Model. Journal of Computers (Taiwan), vol. 31, no. 4, p. 42-53, August, 2020.

[16] Kai Yang, Kaiping Yu, Hui Wang. A hybrid method of multi-objective particle swarm optimization and k -means clustering and its application to modal parameter estimation in the time–frequency domain 26(9-10)(2020) 769-778.

[17] Xu, Yun-Xin, Niu, Li-Chao, Yang, Huan, et al. Optimization of Lithium Battery Pole Piece Thickness Control System Based on GA-BP Neural Network 14(7)(2019) 978-986(9).