An Improved Data Fusion Method for Water Quality Sensors

Zhaoli Wang\(^1\), Jian Xiao\(^2\) and Yiwei Yang\(^2\)

\(^1\)Chengdu Academy of Environmental Sciences, Chengdu, China, 610072
\(^2\)College of computer and Information, Hohai University, Nanjing, China, 211100
Email: lzhxu@hhu.edu.cn

Abstract. This paper proposes an improved data fusion method for water quality sensors. First, a deep learning network model is formed by stacking automatic encoder and sparse automatic encoder, so as to realize feature mining and sparse representation of sample data. Second, after large-scale sample training, the network model may fit complex nonlinear functions, and has certain generalization ability for low-quality sample data. As a result, the accuracy of prediction and classification can be improved. The experimental results demonstrate that the proposed method can obtain higher classification accuracy.

1. Introduction

Water resources control the change of ecological environment and national economic development. It is currently facing with problems such as deterioration and pollution, so water quality monitoring of water resources is needed to provide reference data and basic measures for water quality assessment and prediction of water resources. According to the characteristics of the sensor data in the water quality monitoring system, the data fusion technology is applied to the multi-source data fusion processing, and the comprehensive evaluation results are obtained, which can provide important basis for the maintenance and management of the current water quality. Therefore, in order to better monitor and protect water resources, it is of great significance to explore the data fusion algorithm of water quality sensors.

In recent years, some research results have been achieved in the field of water quality sensor data fusion. Du et al. proposed a cluster grey fusion prediction model to predict the trend of water turbidity [1]. The model adopts the clustering fusion method and the grey prediction algorithm to process the data collected by the sensor, and the result of clustering fusion is used as input data of grey predictive control. The output data of grey prediction control is compared with the fusion data in order to determine the predicted turbidity value. Najah proposed an augmented wavelet de-noising technique with Neuro-Fuzzy Inference System (WDT-ANFIS) based on the data fusion module for WQP, and the technology can handle with water quality data containing noise and predict water quality parameters, which help policy makers report water quality conditions [2]. He proposed a fusion algorithm based on the combination of self-regression model and fuzzy c-means clustering, and the water quality indicators monitored by various sensors were fused to detect the abnormal water quality [3]. Taormina uses a deep learning model to detect physical attacks on the water supply system of the pipe network, and the sensor data is used to train the automatic encoder to map the network status, so that the entire model is sensitive to the sensor data recorded during the attack [4].

Deep learning can not only construct a network model for processing information in parallel, but also has outstanding characteristics such as self-organizing, strong adaptability, and good robustness, which has advantages for processing water quality sensor data with incomplete information and noise problems. It can achieve a good combination of water quality data. Therefore, in this paper we use deep learning to fit complex nonlinear functions in deep network with the advantage of large sample
data, and the data fusion algorithm for water quality sensor based on the mixer encoder is proposed. Simulation results show that the proposed algorithm achieves better results for water quality sensor data classification and evaluation.

2. Proposed Method

By using deep learning to use deep network to fit complex nonlinear functions under the advantage of large sample data, a water quality sensor data fusion algorithm based on stacked hybrid encoder is proposed.

Autoencoder [5, 6] is a typical three-layer symmetric neural network composed of an input layer, a hidden layer, and an output layer. The function between the input layer and the hidden layer is to encode the input data sample, and the main function between the hidden layer and the output layer is to decode the output data of hidden layer. In addition, sparse Autoencoder [7] reconstructs the input sample data more efficiently by finding a set of "overcomplete" base vectors, and finds patterns and structures hidden within the input data. The sparse constraint is added on the basis of the Autoencoder. By adding constraints to the response of each unit of the hidden layer, only a few nodes of the hidden layer unit are active, and most of the nodes are in a suppressed state.

2.1. Network Structure

The network model of the proposed algorithm has a total of five layers. The first layer is the input layer, and the middle three layers are stacked hybrid encoders composed of Autoencoder and Sparse Autoencoder to realize the feature mining and extraction of the input data. The last layer is the Softmax classifier [8, 9], which realizes the classification of output after extracting features of the stack type mixing encoder.

Softmax classifier mainly solves multi-classification problems and is based on Logistic regression. It not only gives the classification result of the data sample, but also gives the corresponding probability of the result. Suppose $I$ labeled training samples are \( \{(x^1, y^1), \ldots, (x^I, y^I)\} \), the training sample was changed to \( \{(C_3^1, y^1), \ldots, (C_3^I, y^I)\} \) after the feature extraction of the stacked mixed encoder, the dimension of the sample input $C_3$ is 6. Category label $Y$ can take 5 different values, so the label for the training sample is $y^i \in \{1, 2, 3, 4, 5\}$. Assume that $p(y = j | C_3^i)$ represents the probability, the data sample is determined to be class $j$ in the case of input $C_3$. When the probability of belonging to a category is the largest, it is determined as this category. Therefore, for the proposed classifier of network model, the output will be a 5-dimensional vector, and the output will be:

\[
h_y(C_3^i) = \begin{bmatrix}
p(y^i = 1 | C_3^i; \phi) 
p(y^i = 2 | C_3^i; \phi) 
\vdots 
p(y^i = 5 | C_3^i; \phi)
\end{bmatrix}
\]

(1)

$C_3^i$ represents the output of the ith training sample after the third level hidden layer. $\phi$ is the model parameter of the Softmax classifier, which is composed of the classifier parameters corresponding to 5 categories, and the relationship is shown as follows:

\[
\phi = [\phi_1^T, \phi_2^T, \phi_3^T, \phi_4^T, \phi_5^T]^T
\]

(2)

The output result after normalization is:
\[ h_k(C^j) = \frac{1}{Z} \begin{bmatrix} \exp(\phi_j C^i) \\ \exp(\phi_j C^i) \\ \vdots \\ \exp(\phi_j C^i) \end{bmatrix} \]  

(3)

The \( Z \) expression is:

\[ Z = \sum_{j=1}^{5} \exp(\phi_j C^j) \]  

(4)

The loss function of the Softmax classifier is:

\[ J(\phi) = -\frac{1}{l} \sum_{i=1}^{l} \sum_{j=1}^{S} s\{y^i = j\} \log \frac{\exp(\phi_j C^i)}{\sum_{k=1}^{S} \exp(\phi_k C^i)} + \frac{\tau}{2} \sum_{j=1}^{S} \sum_{j=1}^{S} \phi_{jj}^2 \]  

(5)

\[ s\{y^i = j\} \] is an indicative function. If the value in braces is false, the output will be 0; if the value is true, the output will be 1. The second term is the L2 regular term, which is the weight attenuation term. It is used to punish the parameters whose weights are too large and make the parameters converge to the global optimal value. \( \tau \) is the weight coefficient.

2.2. Network Training

Training is divided into two phases: unsupervised pre-training and supervised training [10, 11]. The unsupervised pre-training stage represents the learning process, which usually uses a non-labeled data set to perform layer-by-layer pre-training of the hidden layer. The parameters of other layers remain unchanged, and the output characteristics of the previous layer are expressed as the input of the current layer. After completing the training of the layer network, the layer network output feature is also used as the input of the next layer. Supervised learning stage can be divided into supervised classification learning and supervised network fine tuning. Supervised classification learning uses labeled data to train the Softmax classifier. The output data of the sample data processed by the stack hybrid encoder is used as the classifier input. The BP algorithm is used to train and learn the Softmax network. The supervised network fine-tuning uses the network parameters obtained by unsupervised pre-training and supervised classification learning as initialization parameters of the entire network model, and then fine-tunes the entire network through supervised target learning. Finally a deep learning network model with predictive target can be trained.

2.3. Specific Steps

The implementation process of the proposed algorithm can be divided into two stages. Firstly, the network is trained by training samples to determine the parameters of the whole network model. The second stage uses the trained network model to carry on the comprehensive forecast classification of water quality data. The algorithm implementation step of stack type hybrid encoder network model is as follows:

1. The overall structure of the network model is determined, the network parameters are initialized in a random way, and the training sample data set is determined;
2. The training sample data is used as input without tags. The network parameters of the first hidden layer are trained using the Autoencoder method, and the output of the first hidden layer is calculated using the trained network parameters;
3. The output data processed in step 2 is used as the input of the second hidden layer, and the network parameters of the second hidden layer are trained using the Sparse Autoencoder method. The output results of the first two hidden layers are calculated using the trained network parameters;
4. The output of step 3 is used as the input of the third hidden layer, and the network parameter training of the hidden layer is performed using the same method as in step 3;

5. Take the output of step 4 as input to the Softmax multi-classifier, and then use the tags of the sample data to train the network parameters of the Softmax classifier;

(3) Supervised network fine-tuning

6. The network parameters obtained in steps 2, 3, 4 and 5 are used as the initial values of the parameters of the entire deep learning network (3 hidden layers and 1 Softmax output layer), and then the entire network parameters are fine-tuned with labeled sample data. Finally we obtain the optimal network parameter value;

7. The test sample data is predicted by using the trained network mode

3. Experimental Results

To verify the effectiveness of the fusion algorithm of water quality sensor data based on the Stack Hybrid Encoder proposed in this paper, this algorithm is compared with classification algorithm based on BP neural network, classification algorithm based on RBF neural network, classification algorithm based on deep belief network model and classification algorithm based on Stack Sparse Encoder.

The classification algorithm based on BP neural network is denoted as BP, classification algorithm based on RBF neural network is denoted as RBF, classification algorithm based on the Stack Hybrid Encoder is denoted as SHE, classification algorithm based on deep belief network model is denoted as DBN, classification algorithm based on Stack Sparse Encoder is denoted as SSE.

3.1. Loss Function Value

Firstly, the experiment is designed to compare the variation of loss function value between the algorithm proposed in this paper and the stack sparse encoder algorithm when performing supervised network fine tuning. The target accuracy of setting the network fine tuning is $1 \times 10^{-4}$. When the output error is less than the set target precision, the iteration training is stopped, and the curve diagram of the loss function value is shown Fig 1.

The figure shows that when the SSE method is tuned for network fine-tuning, the convergence rate is faster at the beginning of the algorithm model tuning and the fine-tuning reaches the target precision after 858 iterations, the loss function value is 0.0123. The SHE method proposed in this paper performs network fine-tuning. When the number of iterations reaches 391, its loss function value starts to be smaller than the loss function value of the SSE method. The set target accuracy is reached after 865 iterations, and the loss function value is 0.0094. Although the convergence speed of the SSE method is faster at the beginning, the value of the loss function ultimately achieved is larger than the value of the loss function of the algorithm model proposed in this paper. Since the loss function value represents the error between the sample target output and the actual output, it can be concluded that the SHE method has better sample learning ability and can better fit the sample internal feature relationship when performing sample data training, that is, the complex nonlinear mapping capability between input and output is stronger.

![Figure 1. Loss Function Value Curve.](image)
3.2. Prediction Classification Accuracy

The predicted classification accuracy of the three models is shown in Table 1. It can be seen that the proposed SHE method can predict the classification accuracy of the sample data of the water quality sensor by 87% before the network fine tuning, it reaches 99.70% after fine tuning. It shows that fine tuning has a positive and important influence on the prediction accuracy of network model. The proposed algorithm is 0.80% higher than the predictive classification accuracy of the SSE method, and is 1.6% higher than the predictive classification accuracy of the DBN method. Moreover, it can be seen from the table that the deep learning method can achieve higher classification accuracy in large sample conditions than the shallow neural networks such as BP and RBF. Therefore, the proposed SHE method has a better classification effect under large sample water quality standards.

| Method | BP | RBF | SSE | DBN | SHE |
|--------|----|-----|-----|-----|-----|
| Accuracy before fine-tuning | none | none | 84.30% | 34.00% | 87.70% |
| Accuracy | 88.70% | 94.90% | 98.20% | 97.60% | 98.60% |

4. Conclusion

This paper proposes the data fusion algorithm of water quality sensor based on Stacked Hybrid Encoder. The algorithm solves the problem of shallow neural network learning ability under large sample conditions. It can well fit the complex non-linear function relationship within the data, and avoid Overfitting problem. The experimental results show that the algorithm can predict the classification accuracy in a large number of samples.

5. Acknowledgement

This work is supported by Water Conservancy Science and Technology Innovation Project in Guangdong Province (No. 2020-04).

6. References

[1] Du Y, Wei K, Shi Y, et al. Infrared detection and clustering grey fusion prediction model of water quality turbidity. Infrared & Laser Engineering, 2016, 45(10):1028002.

[2] Najah A A, ElShafie A, Karim O A, et al. Water quality predict-ion model utilizing integrated wavelet-ANFIS model with cross-validation. Neural Computing and Applications, 2012, 21(5): 833-841.

[3] He H M, Hou D B, Zhao H F, et al. Multi-parameters fusion algorithm for detecting anomalous water quality. Journal of Zhejiang University, 2013, 47(4):735-740.

[4] Taormina R, Galelli S. Real-Tim-e Detection of Cyber-Physical Attacks on Water Distribution Systems Using Deep Learning// World Environmental and Water Resources Congress. 2017:469-479.

[5] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. Nature, 1986, 323(6088): 533.

[6] Jiang X, Zhang Y, Zhang W, et al. A novel sparse auto-encoder for deep unsupervised learning// Sixth International Conference on Advanced Computational Intelligence. IEEE, 2013:256-261.

[7] Sun W, Shao S, Zhao R, et al. A sparse auto-encoder-based deep-p neural network approach for induction motor faults classification. Measurement, 2016, 89:171-178.

[8] Guo R, Shen M, Huang D, et a-l. Recursive estimation of transition probabilities for jump Markov linear systems under minimum Kullback–Leibler divergence criterion. Iet Control Theory & Applications, 2015, 9(17):2491-2499.

[9] Chen M, Weinberger K, Sha F, et al. Marginalized denoising auto-encoders for nonlinear representations//International Conference on Machine Learning. 2014: 1476-1484.

[10] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. Neural Computation, 2006, 18(7): 1527-1554.
[11] Yang Y W. Research on data fusion algorithm of water quality sensor. *Nanjing: Hohai University*, 2019.