Analysis decision-making system for aquaculture water quality based on deep learning

Jing Su, Jiahong Chen*, Jiawei Wen, Wende Xie, Mingjie Lin

Faculty of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang, Guangdong, 524000, China

*Corresponding author’s e-mail: jhchan3@163.com

Abstract. The quality of water is a key factor that determines whether aquaculture will be successful or not. Currently, aquaculture is mainly based on worker's experience, which is not capable of precisely controlling the quality of the aquaculture water. The main function of the existing system for aquaculture water quality is still to manage water quality, but hardly to analyse or visualize the data of water quality. Furthermore, fewer works for prediction of future water quality are made. To tackle these issues, we design an analysis and decision-making system based on deep learning for aquaculture water quality. We combine the latest artificial intelligence and big data technology to achieve the functions of the system such as the visualization of data, the prediction of future water quality factors and early-warning of water quality, which helps the users to realize scientific aquaculture.

1. Introduction

Currently, since aquaculture is based on previous or current data, the obtained data only reflects the historical status of aquaculture [1,2]. This leads to a lagging and passive situation for water quality management. Farmers often do not realize that water quality conditions have seriously deteriorated until serious consequences occur. Therefore, it is necessary to employ the novel management approaches which can carry out early warning on the deterioration of water quality and formulate corresponding emergency measures to minimize losses and achieve scientific farming. The future trend in aquaculture must be information-based farming. A series of aquaculture measures based on available data will be implemented to achieve a step-by-step transition from manual empirical aquaculture to data-based aquaculture. Drawing on previous approaches to water quality decision system [3,4], we design a more applied aquatic data analysis and decision support system by combining the current popular front-end and back-end technologies [5,6,7], Spark and recurrent neural network (LSTM) [8].

We use the true data from the farm in accordance with the actual requirements. Integrating the latest artificial intelligence technology and the specialization in aquaculture we develop a water quality data analysis and decision support system for aquaculture. Our system achieves the aquaculture industry's water quality data analysis, prediction and data visualization functions. In addition, we provide multiple models for users to choose under different situations. The system improves the disadvantages of the existing approaches in aquaculture which mainly relies on experience in aquaculture and outperforms the other water quality prediction system.

The rest of the paper is as follows. In Part II, we introduce the architecture and functionality of the system. In part III, we elaborate the design of the system. In Part IV, data processing part is described. In Part V, we summarize the paper.
2. Architecture and functionality of the system

2.1. System architecture
The system uses the architecture of Browser and Server. It is divided into three layers: control layer, business layer and persistence layer on the basis of Model-View-Controller design pattern [7]. The specific architecture is shown as Figure 1.

![System architecture diagram](image)

**Figure 1. System architecture diagram.**

2.2. System's functions
This system contains several main components. They will be introduced in sequence below.

2.2.1. User management. The user can use the functions of this system after registering and logging in. The user management module also includes the basic functions such as changing passwords and viewing permissions.

2.2.2. Management of water quality data. Users can enter dissolved oxygen, water temperature, PH, ammonia nitrogen, time as well as the number of pond, and they can also select a different pond number to view the historical data about dissolved oxygen, water temperature, pH, ammonia nitrogen of the pond. In addition, users can modify or delete each piece of historical data. As the data entered at a time may be very large, the system provides the function of batch entry of water-quality data through Excel, making it more humanistic, powerful, and reducing repetitively physical labor.

2.2.3. Water quality model training. The system provides users with several high-performance models such as Xgboost [9], BP [10] and LSTM [8], which enables users to choose one according to the different requirements. After selecting the type of the training model users also need to select the water quality factors which are used as the input of the training model for water quality prediction. Then they can predict future trends in water quality factors through the well-trained model.

2.2.4. Projections of water quality data. The system will display the information about all the trained models after users select the water quality factor they plan to predict. Users can select the model with better training results by the recent training record to predict the water quality factor they need. By clicking on the prediction button the system can show the prediction results in the form of a line chart which enables users to more easily discover the development trend and law of data.

2.2.5. Data visualization. Users can view the water quality data for each pond. The system can draw a line graph based on previous data for each pond to reflect the law of previous water quality data. After
completing the training for water quality data model, the system will draw two folded lines with different colors for the previous data and predicted data. By comparing the gap between two folded lines on the graph, users can get a more intuitive comprehension of how well the model predicts the outcome. In addition, the system will also produce their error values with two folded lines. When predicting water quality data, the system will give the predicted results of the water quality factors in the form of a line graph according to the factors and models users select. The previous data and predicted results are shown in different colors in the line graph.

2.2.6. Water quality warning. Users can set the warning value of dissolved oxygen, water temperature, pH, ammonia nitrogen according to the actual situation in their region. Once the predicted results exceed the warning value, the system will immediately mark the predicted points in red in the visualization interface, and send the alarm information to the user at the same time which facilitates the users to adopt the corresponding strategy in advance to avoid the deterioration of water quality.

3. System design

3.1. Database design
In the system we design five database tables. The water quality table is used to record water quality data such as dissolved oxygen, water temperature, pH and so on. The user table is used to record user information such user name and user secret. The role table is used to record role information including admin, VIP and general user where different roles have different permissions. The model table is used to record historical training models and model evaluation, but the specific model data is stored on the local host. The warning table records the warning values set by the user. If the predicted value exceeds the record, the user can be alerted by mobile messaging or email. The database is designed as Figure 2.

![Figure 2. ER diagram.](image)

3.2. System class diagram
In this part, we mainly introduce the entity classes designed in accordance with the requirements of the system. These entity classes are WaterQuality class, User class, Role class, Model class, WaterQualityHelper class and Warning class. Each class has its attributions and methods which are responsible for implementing the system’s functions such as manipulating the database by addition, deletion, changes and selection or providing support for the later implementation of complex logic. The class diagram is as shown in Figure 3.
3.3. System module

This system exploits modular development to achieve the separation of a frontend and a backend. The main modules contain not only the frontend, Java backend, Python sub-backend, but also the Spark local cluster mode. The frontend part is quickly built by Vue.js and a set of iVew UI components based on Vue.js[11]. The visualization part employs the open source visualization library ECharts. The system uses the Vue framework to implement the user interface.

For the Java backend, the service uses Maven[12] for project management such as managing the dependencies and libraries the water quality decision project required. The system with Maven has the advantages of reusability which facilitates the deployment of the system in different ponds and enhances the maintainability of the system.

For the python backend, we use Django as a sub-background for training and predicting water quality factors. Then, we employ the keras, tensorflow framework to build neural network models. Simultaneously, we construct the machine learning model by the scikit-learn framework.

Finally, the system partially uses Spark [13] (local Spark cluster) and Ubuntu 16.04 LTS to run programs for analysing and predicting water quality data. Spark provides an implementation library that allows users to load data into cluster memory and query it multiple times.

4. Data analysis

There are many factors that affect the water quality in aquaculture [1]. Several crucial factors are dissolved oxygen, water temperature, pH and ammonia nitrogen. We mainly study them in this paper. Specifically, the system predicts the water quality data from the fourth day according to the data from the previous three days.

4.1. Standardization of data

In the phase of the data processing, we first perform the standardization of data which aims at eliminating the difference between the features in the dimension and reducing the dynamic range of the index value. By this way the convergence speed and accuracy of the model is greatly improved.

The data are mapped to the range of 0 to 1 according to Equation.1.

\[ x' = \frac{(x-\mu)}{\sigma} \]  

(1)
Where $\mu$ is the mean of all sample data and $\sigma$ is the standard deviation of all sample data.

4.2. Models
Our analysis system is mainly based on deep learning. Taking into consideration of the expansibility of the system, we use the multiple models for the prediction of aquaculture water quality. The prediction approaches used in the system are divided into two categories. One is machine learning model such as Xgboost, and the other is deep learning model such as BP or LSTM.

When the dataset is not enough large, the Xgboost model is a better choice for the prediction. Nevertheless, the neural network model such as LSTM will make better performance on the large dataset. Therefore, users can choose the most appropriate model for prediction according to the actual situation.

We mainly introduce the deep learning model. In the train phase for the LSTM model [14, 15], we exploit mean-absolute error (MAE) as the loss function of the system. Here MAE refers to the average of the error between the prediction value and the true value of the sample. Additionally, we use Adam algorithm as the optimizer which is suitable to solve problems with the large-scale data or parameters. Finally, the model can gain a higher accuracy in the water quality prediction by making a tiny adjustment to the hyper parameters.

4.3. Model prediction
The system employs the root mean square error (RMSE) to evaluate whether the model predicts good or bad, as shown in Equation 2.

$$\sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_{test}^{(i)} - \hat{y}_{test}^{(i)})^2}$$ (2)

Where $y_{test}^{(i)}$ is the predicted value of the sample, $\hat{y}_{test}^{(i)}$ is the actual value of the sample, and the value of m is the sample size. The lower the RMSE, the smaller the gap between the actual and predicted values will be.

Figure 4 shows the predicted PH value of a LSTM model where the predicted value is in the blue line while the actual value is in the red line. The two lines are quite close which stands for the relative good effect for the model prediction.

5. Summary
We design a novel analysis and decision-making system for water quality in aquaculture in line with the actual needs of the aquaculture environment. The system mainly includes three modules: data management module, data visualization module and data analysis module.

Compared with the existing systems for aquaculture, our system combine deep learning and Spark framework to provide the users with more scientific management way. Furthermore, the system
provides more intuitive interfaces of visualization. The system greatly improves the efficiency of breeding and has a high application value.

Acknowledgments
The work was financially supported by the Innovation and Entrepreneurship Training Program for University Students in Guangdong Province (CXXL2019088).

References
[1] Ma, Z. (2013) The Construction of DSS in Water Quality Management of Intensive Litopenaeus vannamei shrimp tanks[D]. Shandong: Ocean University of China.
[2] Zhang, J. (2018) Research status and trend of key technologies of aquaculture informatization. Friends of farmers getting rich, (18):239
[3] Liu, S. (2014). Prediction and Early-warning of Water Quality in Aquaculture Based on Computational Intelligence [D]. China Agricultural University.
[4] Xu D. (2016) Design of cross platform aquaculture management system based on machine learning [D] (Master's thesis, Taiyuan University of Technology).
[5] Lin, C., Liu, D., Pang, W., & Wang, Z. (2015). Sherlock: A semi-automatic framework for quiz generation using a hybrid semantic similarity measure. Cognitive computation, 7(6), 667-679.
[6] Liu, D., & Lin, C. (2014). Sherlock: a Semi-Automatic Quiz Generation System using Linked Data. In International Semantic Web Conference (Posters & Demos), Italy.
[7] Zhang, L., Wang, Y. (2018) Research on MVC Model Based on SpringBoot MicroServices Architecture[J]. Journal of Anhui electronic information Career Technical College, 17(4):1-9.
[8] Li, R., Lin, C., Collinson, M., Li, X., & Chen, G. (2019). A Dual-Attention Hierarchical Recurrent Neural Network for Dialogue Act Classification. In 2019 SIGNLL Conference on Computational Natural Language Learning (CoNLL), Hong Kong.
[9] Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015) Xgboost: extreme gradient boosting. R package version 0.4-2, 1-4.
[10] Ye, Y., & Zhuang, Y. (2012) Water Quality Prediction Technology Based on BP Neural Network[J]. Journal of Beihua University(Natural Science) ,2012,13(04):493-496.
[11] Wang, Z. (2018) The Design and Implementation of Vue.js Based Development Platform [D].Guangdong University of Technology.
[12] Lv, Y. (2018) SpringBoot framework in web application development[J]. Science and Technology Innovation Herald. 8, 168-173.
[13] Wang, X. (2016) The Research On Classification And Regression Tree's Parallelization Based On Spark Platform[D] (Master's thesis, Chongqing: Chongqing University).
[14] Su, J., Lin, C., Zhou, M., Dai, Q., & Lv, H. (2018). Generating Description for Sequential Images with Local-Object Attention Conditioned on Global Semantic Context. In Proceedings of the Workshop on Intelligent Interactive Systems and Language Generation (2IS&NLG).
[15] Chen, Y., Cheng, Q., Fang, X., Yu, H., & Li, D. (2018) Principal component analysis and long short-term memory neural network for predicting dissolved oxygen in water for aquaculture[J]. Transactions of the Chinese Society of Agricultural Engineering, 34(17).