Implementation of water quality management platform for aquaculture based on big data

Zhibin Peng
Faculty of Mathematics and Computer Science
Guangdong Ocean University
Zhanjiang, China
905906851@qq.com

Yuefeng Chen*
Faculty of Mathematics and Computer Science
Guangdong Ocean University
Zhanjiang, China
* yuefengch71@126.com
*Corresponding author: Yuefeng Chen

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

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

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

***Abstract*—In order to ensure the quality and quantity of aquaculture, aquaculture farmers need to grasp the water quality in time. However, most farmers have to collect water quality data manually at present, and cannot store and reuse that information rapidly. This paper aims to use SpringBoot framework and JPA framework to build a big data platform of acquisition automation and visualization, which realizes the data analysis and display of heterogeneous water quality and breeding information. The platform can make the water quality prediction and real-time warning. Meanwhile, it realizes the management of robots, users and breeding experts. The application of this platform will bring better social benefits to aquaculture farmers.**

Keywords: aquaculture;big data; water quality warning;data visualization

I. INTRODUCTION

With modern technology development, automation of aquaculture monitoring technology has been improved. However, most farmers use robot or sensor equipment only to get some characters of water for one time, and that information usually is not stored and reused to forecast breeding status. In order to increase productiveness and scale of aquaculture, it's necessary to develop water quality management platform to improve the automation and intelligence of aquaculture industry.

At present, a lot of work is focused on the research of aquaculture robot platform. The paper [1] studied the application of ROV underwater automatic collection of water quality samples and all-round multi-parameter water quality detection. The paper [2] made a contribution to the combination of big data analysis technology, cloud computing technology and online water quality monitoring technology. Four important indicators for water quality monitoring in aquaculture are proposed, and their importance to water quality are revealed [3].This paper [4] proposed a new integrated method to predict the dissolved oxygen content in water, which improves the prediction accuracy. Combined spectral processing technology with machine learning, the researchers used digital image to analyze water quality [5]. The paper [6] summarized and compared the water quality models of three kinds of offshore waters, and mentioned the coupling direction of the new technology. There are many limitations in the development of traditional water quality management platform, which cannot make full use of the collected data to process and change into useful data. Instead of paying attention to the prospect of intelligent detection and the significance of water quality data, most research of aquaculture are invested on machine hardware, such as sensor precision, remote sensing range and endurance ability, and that makes shortcomings in further analyzing and storing the large amount of accurate data collected.

In order to solve these problems and use data to provide practical help to the aquaculture industry, we develop the Water Quality Management Platform for Aquaculture (WQMPA).

II. PLATFORM ARCHITECTURE

The platform background use SpringBoot [7] framework combined JPA [8] framework to develop system framework, as well as front-end use Highcharts [9] framework. Android [10] use MPAndroidChart framework to display data changes. In this way, farmers can use APP and PC to view real-time data of aquaculture robots easily, and experts on aquaculture can also

The research is supported by National Innovation and Entrepreneurship Training Project for College Student (China,2019)

978-1-7281-9837-8/20/$31.00 ©2020 IEEE
DOI 10.1109/CIBDA50819.2020.00024

view problems encountered by farmers and abnormal data and provide some substantive help. At the same time, the data of aquaculture robots based on Hadoop framework can receive all kinds of isomerization data in real time. The TensorFlow [11] framework and keras [12] framework provide water quality prediction and real-time warning functions to facilitate the scientific breeding of aquatic products by farmers.

The core of WQMPA is big data mining and analysis, and its data sources are mainly sensor data, user active upload data and Internet data. Then the massive data is stored by HDFS, and MapReduce [13] can do parallel offline computation. The core technology of the platform is Hadoop, combines machine learning, deep learning and other technologies [14,15] to calculate and mine the data, and finally provides data visualization, breeding prediction, environmental prediction and other services. The architecture includes five layers：Data source, Data server, Data calculation & mining, data service and Data view, which is shown as Figure 1.

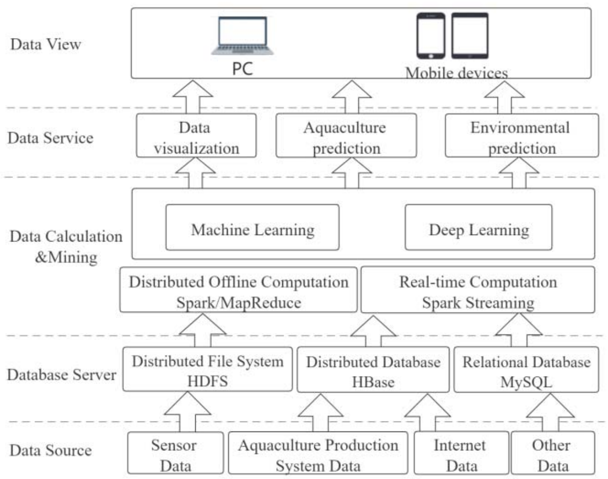

Figure 1. Architecture of WQMPA

III. SYSTEM ANLYSIS

There are three types of users in the system: farmers, aquatic experts and administrators. The main modules of WQMPA include user permission validation, platform management, data display and extraction, and water quality prediction and warning. Functional structure diagram is shown as Figure 2.

A. Permission Validation

All users are required to register and login. Without logging, users can't use the system. In order to prevent the platform from being cracked by violence, the system processes the user password with md5 algorithm and protects the login module with authentication. This protects platform security effectively.

B. Platform Management

The super administrator has all the privileges and can use all the functions of the system, such as: user management, view background data， user usage and so on. The farmers can manage the data of the aquaculture robot, so as to control the robot better and work effectively. Experts can also evaluate the data collected by the breeding robot through the authority granted by the farmers, which can help the breeding process.

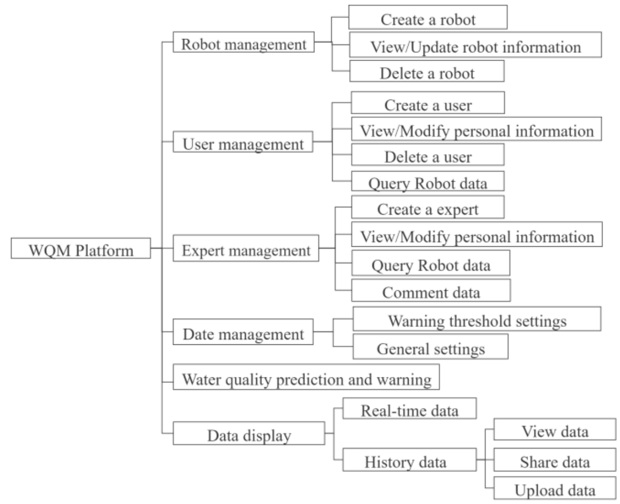

Figure 2. Functional structure diagram of platform

C. Data Presentation and Extraction

The data collected by robots can be displayed most directly through data visualization. The front-end application framework can be used to enable users to observe and use the data in the most intuitive and convenient way. Farmers can view the robot information directly, expert comments and other information data from the WQMPA platform. Moreover， the collected data can be stored into the system background database, so that the farmers can download their interested water quality data. And the data will be displayed and extracted in the form of line charts, bar charts, pie charts and Excel sheets.

D. Water Quality Prediction and Warning

The quality of water quality is related to PH value, dissolved oxygen, ammonia nitrogen, nitrite, acidity and alkalinity, etc. but on the whole, the problem of water quality is not just a single index, and it is difficult for farmers to judge which factors cause the sick fish. The system focuses on the use of big data technology to analyze the collected data and predict the water quality. When the water quality is likely to exceed or fall below the water quality safety level in the future, the platform will give users early warning, so as to reduce the economic losses of farmers.

IV. PLATFORM DESIGN

In the platform, we use SpringBoot framework designed seven entity class. They are Warning class, User, Data class, Robot class, Token class, Processor class, Amplitude class. Warning and Data are the two core entity classes in the platform, and the big data processing and water quality Warning are realized through the methods of these two classes.

The relationships between the seven entity classes are shown in figure 3.

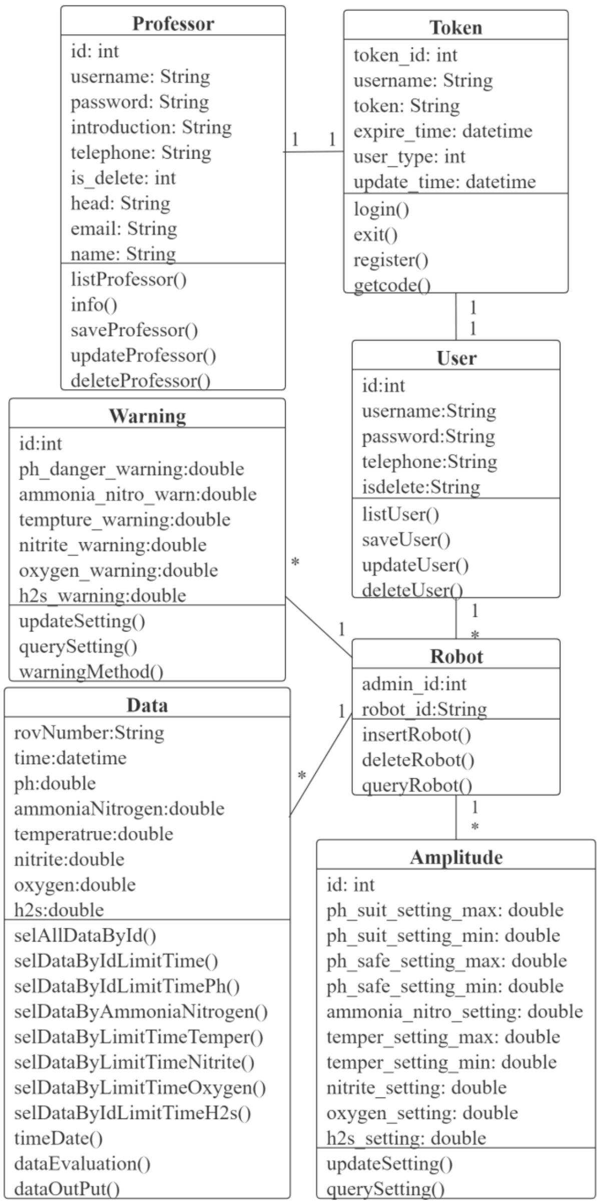

Figure 3. Entity class of WQMPA

V. KEY TECHNOLOGIES

A. Big Data technologies

This platform is a high reliability, high-performance, column-oriented and scalable distributed storage system, mainly using distributed storage system HBase technology to build a large-scale structured storage cluster on PC Server. Among them, HBase is located in the structured storage layer; Hadoop HDFS provides the low-level storage support with high reliable performance for HBase. Hadoop MapReduce provides high-performance computing power for HBase, and Zookeeper provides stable service and failover mechanism for HBase. Data acquisition model is shown in Figure 4. Kafka, the distributed publish-subscribes message system, provides the data cache and guarantees the high throughput of the system. When gathering information reach the peak, Flume will make adjustments between the data producer and the data receiver to ensure that it can provide smooth data. For size of data is too big to load them all into memory for analysis, HBase native data format can't satisfy our basic requirements of data processing. Hadoop MapReduce will distribute the data to each node of the cluster for operation, and the data obtained from each node will be constructed into the structure of balanced binary tree. By traversing the balanced binary tree, we can get the required data. The high-performance computing power of MapReduce facilitates the statistical analysis of water quality, and enables us to process data according to different business requirements.

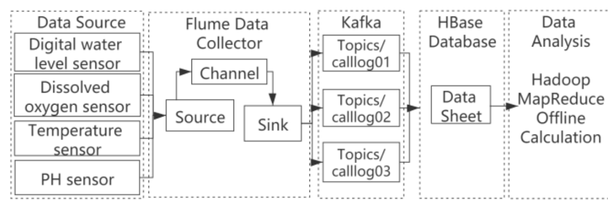

Figure 4. Diagram of data processing

B. Neural Network Model

Using the Tensorflow framework and Keras, the system builds LSTM-RNN neural network [16,17] and CNN neural network respectively to realize the function of water quality data prediction and water quality warning.

1) Prediction of water quality data:

First,the water quality data is converted to vector and put into LSTM-RNN model.Then the water quality data of the next period is predicted in real time by using the time series model characteristics of RNN, and the data is fed into the trained warning model to realize early warning. The character of LSTM is to add a valve nodes in each layer, except for the RNN structure. There are three types of valves: forget gate, input gate and output gate. These valves, which can be opened or closed, will be used to determine whether the previous memory state of the model network has reached the threshold output in this layer and thus be added to the calculation of the current layer. As shown in the figure 5, the valve node uses the *sigmoid* function to calculate the memory state of the network as the input. If the output reaches the threshold, the valve output is multiplied by the calculation result of the current layer as the input of the next layer. By contrast, the output result is forgotten. The weight of each layer including the valve node, will be updated during back propagation training in each model.

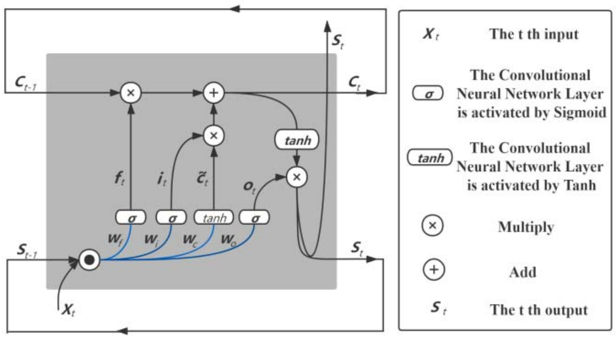

Figure 5. Diagram of LSTM calculation and judgment process

2) Water quality warning

Based on four important indicators of water quality, we built CNN neural network to realize water quality warning function. We can obtain water quality data: dissolved oxygen, ammonium nitrogen, nitrite, PH value, through aquaculture robots. Then compare to labeled warning data (provided by the experts according detail breeding variety), and as labeled data, they also be converted into a vector and input into the model to realize the traditional supervision training and learning. After the training, the real-time water quality data can be made decision to warning data or normal data.

C. Data Visualization

Data visualization [18] is the most direct demonstration of the platform for farmers, which has high real-time requirements. On Android, use the Jetpack component officially recommended by Google to implement the layered architecture of MVVM. The business logic of view and data is developed in different modules, which greatly improves the scalability and stability of the project. In the Jetpack component, the Room implementation interacts with the built-in SQLite data on the phone, to save technologies that need to be persisted during the use of the system to the phone. And Retrofit is responsible for network interaction. The aquaculture data from robots are first sent to the server, and then the platform interacts with the background server through Retrofit to obtain real-time data. At the same time, some files with large volume will be saved to the local file by Room. The next access can be directly obtained from the local, reducing the pressure of network requests. The request code of Room and Retrofit are encapsulated in Repository, and can be invoked in the system through its object class. ViewModel is responsible for the operation of data, which originally are written in the Activity and Fragment. It means that the View Activity and Fragment only need to complete the presentation of data, making the task assignment of the function clearer. After the ViewModel has processed the data, it sends the data to the Activity and Fragment through LiveData, which is the bridge between them. ViewBinding and DataBinding play assistant roles, and are also responsible for the data presentation of the Activity and Fragment, making the code more concise. The Lifestyle framework is built into each component framework to ensure that the life cycle of the Jetpack components is consistent, reducing memory leaks due to improper references.

VI. CONCLUSION

The purpose of this platform is to predict the change of water quality parameters, monitor abnormal phenomena, prevent the occurrence of diseases in time, and reduce the cost of breeding. We hope to extract different characteristics contained in mass breeding data, carry out effective classification and detection, and provide effective information mined to farmers or relevant departments, so as to realize intelligence and visualization of aquaculture and provide early warning information for relevant departments in time. The platform has high prospective and high value of research and development.

ACKNOWLEDGMENT

The research is supported by National Innovation and Entrepreneurship Training Project for College Student (China, CXXL2019022).

REFERENCES.

[1] H. Gu, L. J. Zhang, X. Liu, X. J. Xue, and L. Zhang, “Application of multifunctional submersible ROV in water environment monitoring,” Beijin Water, vol. 5, pp. 14–17, May 2018.

[2] B. Qin, “Discussion on the application of big data in the on-line monitoring of water quality,” Low Carbon World, vol. 6, pp. 30-31, June 2019.

[3] Z. W. Xu, “Determination and analysis of water quality index and its application in aquaculture,” Jilin Animal Husbandry and Veterinary Medicine, vol. 35, pp. 18-20, May 2014.

[4] K. Ozgur, A. Meysam, and G. AliReza, “Dissolved oxygen prediction using a new ensemble method,” Environmental Science and Pollution Research. Germany, vol. 27, pp. 9589-9603, Mar 2020.

[5] Y. H. Li, X. Wang, Z. X. Zhao, S. Han, and Z. Liu, “Lagoon water quality monitoring based on digital image analysis and machine learning estimators,” Water Research, vol. 172, pp. 115471, Apr 2020.

[6] Y. F. Lu, Z. B. Ding, W. Lei, P. Chen, and X. Chen, “Review of coast quality models: Research process and outlook,” Marine Sciences, vol. 44, pp. 161-170, Feb 2020.

[7] H. Suryotrisongko, D. P. Jayanto, A. Tjahyanto, “Design and Development of Backend Application for Public Complaint Systems Using Microservice Spring Boot”, Procedia Computer Science, pp. 736-743, 2017.

[8] Y. Y. TaN, “Research and Implementation of Data Query Based on Hibernate JPA and JQuery Framework,” Computer and Modernization. Vol. 1, pp. 196-198, 2012.

[9] D. Z. Bian, J. T. Li, “Research of Real-Time Data Presentation of Cloud Monitoring Based on Highcharts,” Advanced Materials Research, vol. 3716, pp. 847-850, 2015.

[10] W. Enck, M. Ongtang, and P. McDaniel, “Understanding Android Security,” IEEE Security & Privacy, vol. 7, pp. 50-57, Mar 2009.

[11] B. Pang, E. Nijkamp, Y. N. Wu, “Deep Learning With TensorFlow: A Review,” Journal of Educational and Behavioral Statistics, vol. 45, pp. 227-248, 2020.

[12] Vidnerova, Neruda, “Evolving keras architectures for sensor data analysis,” Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, pp. 109-112, Nov 2017.

[13] J. J. Li, Y. J. Liu, J. Pan, P. Zhang, W. Chen, and L. Z. Wang, “Map-Balance-Reduce: An improved parallel programming model for load balancing of MapReduce,“Future Generaton Computer Computers-The International Joural of ESCIENCE,” vol. 105, pp. 993–1001, 2020.

[14] D. Liu, L.H. Cheng, "Sherlock: a Semi-Automatic Quiz Generation System using Linked Data." In International Semantic Web Conference (Posters & Demos), pp. 9-12,2014.

[15] C. Lin , D. Liu , W. Pang , Z. Wang, Sherlock: "A semi-automatic framework for quiz generation using a hybrid semantic similarity measure. Cognitive computation", 7(6), pp:667-679,2015.

[16] J.Y. Yi, , Z. Q. Wen, , J. H. Tao,H. Ni,and B. Liu, "CTC Regularized Model Adaptation for Improving LSTM RNN Based Multi-Accent Mandarin Speech Recognition". Journal of Signal Processing Systems, vol:90, pp:985–997 ,2018.

[17] X. Li , and G. Chen, "A Dual-Attention Hierarchical Recurrent Neural Network for Dialogue Act Classification". In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL) ,pp. 383-392, 2019.

[18] X. Chen, X. H. Zhu, and Y. Yue, "Dynamic Data Visualization for a Water Quality Monitoring Network," Proceedings of 2017 5th International Conference on Mechatronics,Materials,Chemistry and Computer Engineering(ICMMCCE 2017), pp. 1596-1599, Jul 2017.