Aquaculture Water Quality Prediction using Smooth SVM

Wijayanti Nurul Khotimah

Abstract—***Aquaculture, aqua farming, is the farming of aquatic organism such as fish, crustaceans, mollusk and aquatic plan. There are many factors that influence the production of aquaculture such as food stocks, protection from other predators, and water quality custody. In modern intensive river aquaculture management, water quality prediction plays an important role. The water quality indicator series are nonlinear and non-stationer. Hence, the accuracy of the commonly used conventional methods, including regression analyses and neural networks, were limited. A prediction model based on Smooth Support Vector Machine (SSVM) is proposed in this research to predict the aquaculture water quality. SSVM is an algorithm that is used for solving no linear function estimation problems. The data used in this research are data of river in Surabaya collected for two years. The data have twenty variables that indicate water quality such as temperature, turbidity, color, SS, pH, alkalinity, free CO2, DO, Nitrite, Ammonia, Copper, phosphate, sulfide, iron, Hexavalent Chromium, Manganese, Zinc, Lead, COD, and Detergents. From 520 instance data, we used 5-fold for the experiment. The Root Mean Square Error (RMSE) of the experiment is 0.0275. This value shows that SSVM proven to be an effective approach to predict aquaculture water quality.***

Keywords—Aquaculture, Water Quality, Smooth Support Vector Machine.

I. Introduction

The water quality is determined by its physical, chemical and biological parameters [1]. Physical parameters are related to sight, touch, taste, and smell, such as: temperature, turbidity, colour, suspended solid, and pH. Because water is an excellent solvent, then a lot of materials are soluble in water. Thus, the materials are chemical parameters such as: alkalinity, CO_2, DO, Nitrite (NO_2), manganese, and iron (Fe). Biological parameters are related to the existence of micro-organisms in water, such as: coliform, escheria coli [2].

Water quality monitoring is one of the highest priorities in environmental protection policy [3]. Ascertaining it is crucial before use for various intended purposes such as potable water, agriculture, recreational, industrial water uses, and aqua farming.

Aqua farming is the farming of aquatic organism such as fish, crustaceans, molluscs and aquatic plan. One of the factors that influence the production of aquaculture is water quality custody. Therefore water quality prediction plays an important role in modern intensive river aquaculture management. Once the water quality deteriorates and the aquatic organism are in poor environment, their quality decline, they will get disease and dead easily. It will cause great economic losses to the farmers [4].

Many researches to predict water quality has been done. Feifei Li et al. established short-term forecasting model to predict dissolved oxygen using back-propagation (BP) and autoregressive (AR) [5]. Palani et al. developed a neural network model to forecast the amount of dissolved oxygen in seawater [6]. However, neural networks suffer from a few weaknesses, such as over-fitting.

In this research, we propose SSVM to predict water quality (especially dissolve oxygen). The rest of the paper is organized as follows. In the next section, we begin by explaining and giving some simple preliminary analysis of the methods, the water quality parameters, Smooth Support Vector Machine, and evaluation parameter (RMSE). In section III, we explain about data collection, pre-processing, and experiment result. The conclusions of this paper are shown in section IV.

II. Method

A. Water Quality Parameter

Water quality parameters are divided in to three parts physical parameter, chemical parameter, and biological parameter.

1) Physical parameter

a. Temperature

Temperature is important for aquatic organism because it influences growth, reproduction, respiration, heart rate, enzymes activity and other physiological processes. Temperature also affects the level of dissolved oxygen in water [7].

b. Turbidity

The higher the turbidity, the shallower lights penetrate the water.

c. Colour

Colours that appear on the water are usually caused by the contact of water with decayed organic such as leaves and wood. The decayed organic will be decomposed. It produces compounds that leach into the water and cause the water to be colour. However, well water should be physically colourless' [8].

d. Suspended Solid (SS)

Suspended Solid (SS) is solids contained in the water. It typically contains organic and inorganic compounds. Excessive SS can harm fish and other aquatic organism because it can cover the gills and reduce the solar radiation. Further the reduction of solar radiation will affect the natural food chain [9].

e. The degree of acidity (pH)

The degree of acidity or pH is a scale that indicates level of acid or alkaline of a solution. The pH value of normal water is neutral between 6 and 8. pH less than 7 is acidic, while pH greater than 7 is alkaline. Very acidic or alkaline water will interfere the survival of aquatic organism because it causes the disruption of metabolism

[1]Wijayanti Nurul Khotimah is with Departement of Informatics Engineering, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia. E-mail: wijayanti@if.its.ac.id.

and respiration of them [8].

2) Chemical Parameter

a. Alkalinity

Alkalinity is the quantity of anions in the water which can neutralize hydrogen cation.

b. Carbon dioxide (CO_2)

Carbon dioxide (CO_2) is a normal component in all natural water and gas that is soluble in water. In nature, CO_2 consist of CO_2-free and CO_2-bound which depends on the pH.

c. Dissolved Oxygen (DO)

Dissolved oxygen (DO) is the most common parameter that is used to measure the water quality. Low DO adversely affects the survival of aquatic organism. Oxygen need of fish is different based on their species and age. DO level with range between 3-6 mg/liter is critical almost for all kind of fish [9].

d. Nitrite (NO_2)

In natural water, concentration of nitrite (NO_2) is very small. The presence of nitrite describes that the biologic process of destroying organic material is happening.

e. Manganese

Manganese is a metal ion that can cause problems in water supply systems. Ground water containing manganese are always lack of dissolved oxygen and excess of carbon dioxide.

3) Biological Parameter

The existence of organism in water can be used as biological parameter. The example of the bacteria are:

a. Coliform
b. Escheria coli

B. Smooth Support Vector Machine (SSVM)

Nowadays, Support Vector Machine (SVM) becomes a method that is constantly evolving and increasingly popular in machine learning. This method has several advantages over other methods such as neural network. Those advantages are [10].

a. Assuring global optimum
b. Parameters that must be estimated relatively few
c. The model is stable
d. Relatively simple to use
e. Successfully applied in most real cases.

Many SVM developments proposed to improve its performance and efficiency. Lee and Mangasarian [11] have proposed a new formula of SVM with linear and non linear kernel for classification using smoothing method. The method is called Smooth Support Vector Machine (SSVM). The basic concept of SSVM is changing primal SVM formula into non smooth optimization problem without constraint. Hence, the objective function of the optimization problem is not differentiable. Therefore smoothing function was used in purpose to achieve differentiable objective function. This method was solved by Newton Armijo algorithm.

SSVM has important mathematical properties such as strong convexity and differentiable infinity. Those properties proofed that when the smoothing parameter approaching infinity, SSVM converges to the unique solution of the optimization problem [11]. The optimization problem of SSVM is shown in the following equation:

$$\min_{(w,\gamma,y)\in R^{n+1+m}} \frac{v}{2}y'y+\frac{1}{2}(w'w+\gamma^2)$$
$$\text{s.t. } \mathbf{D}(\mathbf{A}w-\mathbf{e}\gamma)+y\geq \mathbf{e} \qquad (1)$$
$$y\geq \boldsymbol{e}$$

Constraint in Equation (1) can be written as follows:

$$y=(\mathbf{e}-\mathbf{D}(\mathbf{A}w-\mathbf{e}\gamma))_+ \qquad (2)$$

Thus, y in Equation (1) with ***(e-D(Aw-eγ))***$_+$ and convert SSVM problem in Equation (1) into the following equation:

$$\min_{(w,\gamma)}\frac{v}{2}\left\|(\mathbf{e}-\mathbf{D}(\mathbf{A}w-\mathbf{e}\gamma))_+\right\|_2^2+\frac{1}{2}(w'w+\gamma^2)$$
(3)

Here, function plus $(x)_+$ is defined as $(x)_+=$ max $\{0,x_i\}$, i = 1,2,3…n

Objective function in Equation 2.5 can not be differentiate two times. Lee and Mangasarian [11] (2001) applied smoothing technique by using $p(x,\alpha)$ instead of using x_+.

$$p(x,\alpha)=x+\frac{1}{\alpha}\log\left(1+\varepsilon^{-\alpha x}\right),\ \alpha>0 \qquad (4)$$

Where α is the smoothing parameter. Therefore a new model of SSVM:

$$\min_{(w,\gamma)\in R^{n+1}}\Phi_\alpha(w,\gamma):=\min_{(w,\gamma)\in R^{n+1}}$$
$$\frac{v}{2}\left\|p(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}w-\boldsymbol{e}\gamma),\alpha)\right\|_2^2+\frac{1}{2}(w'w+\gamma^2) \qquad (5)$$

The SSVM model above can be solved with Armijo Newton algorithms as follows:

a. Newton Armijo Algorithm:

Begin with initialization $(w^0,\gamma^0)\epsilon R^{n+1}$.

Compute (w^{i+1},γ^{i+1}) as follows:

b. Newton Direction: computing direction $d^i\ \epsilon R^{n+1}$ by solving $n+1$ linear equation with n+1 variable as follows:

$$\nabla^2\Phi_\alpha\left(w^i,\gamma^i\right)d^i=-\nabla\Phi_\alpha\left(w^i,\gamma^i\right)' \qquad (6)$$

c. Armijo Stepsize: Choosing stepsize $\lambda_i\ \epsilon R$:

$$\left(w^{i+1},\gamma^{i+1}\right)=\left(w^i,\gamma^i\right)+\lambda_i d^i \qquad (7)$$

With $\lambda_i\ \epsilon R=max\{1,1/2,1/4,\ldots\}$ therefore:

$$\Phi_\alpha\left(w^i,\gamma^i\right)-\Phi_\alpha\left(\left(w^i,\gamma^i\right)+\lambda_i d^i\right)\geq-\delta\lambda_i\nabla\Phi_\alpha\left(w^i,\gamma^i\right)d^i \qquad (8)$$

Where $\delta\epsilon(0,1/2)$.

Loop until $\nabla\Phi_\alpha\left(w^i,\gamma^i\right)=0$.

By using the Newton-Armijo algorithm can be seen that the main difference between the smoothing approaches with SVM is that SSVM solve problem using linear equations, while SVM using quadratic programming problem.

C. Root Mean Square Error (RMSE)

RMSE is a frequently used measure of the difference values between predicted value that was produced by model and actual value observed from environment. The RMSE is defined by the square root of the mean square error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}} \qquad (9)$$

Where X_{obs} is observed values, X_{model} is predicted values at time i, and n is the number of the data.

III. Result and Discussion

A. Data Collection

The understanding of a reciprocating compressor can The number of data that we collected is 520. Thus consist of 20 parameters. The detail of those parameters is shown in Table 1. From those parameters we used DO (Dissolved Oxygen) as parameter that will be predicted. Since the process to measure the value of DO by experiment is complex (It needs 2 days experiment) and dissolved oxygen is important factor affecting the growth of aquaculture. For experiment, we used k fold validation. The result of the experiment will be explained in the next chapter.

B. Data Pre-processing

1) Normalization

Based on Table 1, the range value among the parameters is very high. For example, the range value of free CO_2 is between 0 and 105.24 while the range value of iron is between 0.02 and 1.78. Figure 1 and Figure 2 show the value of free CO_2 and iron respectively. In order to eliminate dimension differences, the data were normalized and standardized to the range [0, 1]. The following formula was used for this process.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (10)$$

Here, x denotes the original data point, x_{min} and x_{max} are the minimum and maximum values in data set, respectively.

2) Remove the Null Value

From 520 collected data, there were some days the value of the parameters are null. For example the concentration of manganese that is shown in Figure 3. In this figure, most of the values are zero. In this experiment we discarded the data that contains a lot of null value. Those parameters are Manganese, Seng, Timbel, COD, and detergent. As the consequences, we did not use those parameters as predictors.

C. Experiment Result

The result of dissolved oxygen prediction using SSVM is shown in Figure 4. The figure shows the comparison between prediction value of this proposed method and actual value (original data). We used 5 fold for the experiment. The RMSE value of the prediction is 0.0275. This value shows that SSVM proven to be an effective approach to predict dissolved oxygen.

IV. Conclusion

Water quality prediction is very important for intensive aquaculture. It can help provide early warning of the change of water quality and reduce the loss of aquaculture. The method introduced here employs SSVM approach for predicting water quality, especially dissolved oxygen concentration. Hopefully the SSVM forecasting method to predict water quality can help avoid economic losses causes by water quality problems.

References

[1]. Sargaonkar, A., & Deshpande, V. (2003). Development of an overall index of pollution for surface water based on a general classification scheme in Indian context. Environ. Monit. Assess , 43-67.

[2]. Putri, O. (2013). Implementation of Quality Management System Using Association Rule Mining and Storet for Raw Water Quality Parameter Predictions. Surabaya.

[3]. Simeonov, Einax, Stanimirova, & Kraft. (2002). Environmetric modeling and interpretation of river water monitoring data. Anal. Bional. Chem , 374 898-905.

[4]. Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., & Wei, Y. (2013). A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Science Direct , 458-465.

[5]. Feifei, L., Daoliang, L., & Daokun, M. (2010). , Dissolved oxygenprediction in apostichopus japonicas aquaculture ponds by BP neural network andAR model. Sensor Letters 8 , 95.

[6]. Sundarambal, P., Shie-Yui, L., & Pavel, T. (2009). Development of a neural network for dissolved oxygen in seawater. Indian Journal of Marine Science, 38 (2) , 151-159.

[7]. Wahida, N. (2013, January 8). Retrieved September 9, 2013, from http://nurulwahidadotme.wordpress.com/2013/01/08/58/

[8]. Pratama, H. Y. (2010, Oktober 15). Retrieved September 9, 2013, from http://pratamahendrik.blogspot.com/2010/10/analisa-air.html

[9]. Wardana, Y. (2012). Parameter Pencemaran Air.

[10]. Abe, S. (2010). Support Vector Machines for Pattern Classification 2nd edition. New York: Springer- Verlag.

[11]. Lee, Y. J., & Mangasarian. (2001). A smooth support vector machine. Journal of Computational Optimization and applications , 20: 5 – 22.

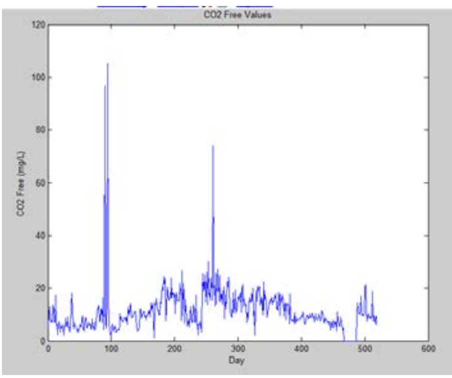

Figure 1. The Range Value of Free CO_2

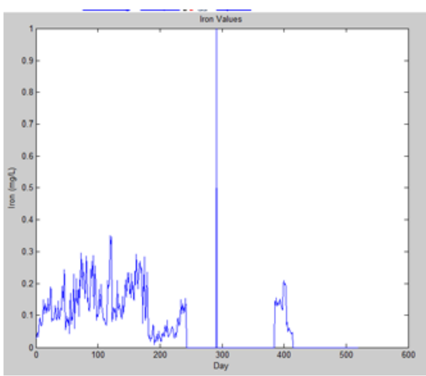

Figure 2. The Range Value of Iron

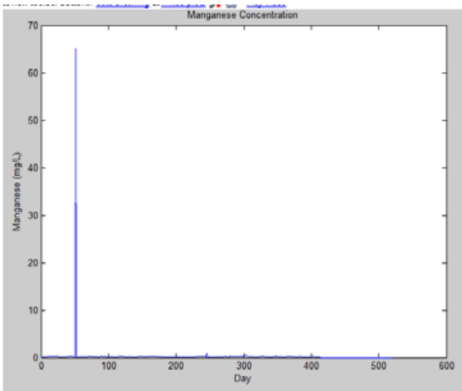

Figure 3. The Concentration of Manganese

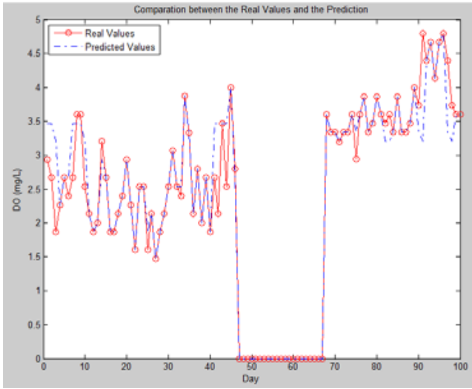

Figure 4. The Comparison between Predicted Value and Actual Value of DO

TABLE 1.
DATA FEATURES

No	Name	Unit	Range Value
1	Temperature	^{o}C	21.8-30.4
2	Turbidity	NTU Scale	3.8-1370
3	Color	Pt Co Scale	5.808-1983.8
4	SS	mg/L	0-398
5	pH		7.11-8.18
6	Alkalinity	mg/L	0-278.6
7	Free CO2	mg/L	0-105.24
8	DO	mg/L	4.27-57
9	Nitrite	mg/L	0.05-1.207
10	Ammonia	mg/L	0.05-4.34
11	Copper	mg/L	0-3.564
12	Phosphate	mg/L	0.06-8.839
13	Sulfide	mg/L	0-1.65
14	Iron	mg/L	0.02-1.78
15	M Heksavalen	mg/L	0-0.34
16	Manganese	mg/L	0-65
17	Zinc	mg/L	0-0
18	Plumbum (Pb)	mg/L	0-0
19	COD	mg/L	0-0
20	Detergent	mg/L	0-0