Data Driven Water Quality Analysis Based on Hybrid Classification Technique

Ms. Swapnali D. Mahadik[1], Dr.Anup Girdhar[2]

[1](PhD Research scholar, TMV,Pune
Assistant Professor, Navinchandra Mehta Institute of Technology and Development, Mumbai,India)
[2](CEO-Founder, Sedulity solutions & Technologies, Ph.D.Guide, TMV,Pune,India)

Abstract: *Water quality prediction plays a very important role in today's world. Water quality is always related to wellness of its neighbors. Water quality is affected by many factors such as human activities, Industrial wastes, chemicals, biological components, etc. The discharges added by these pollution sources affect the pH value of water. Thus it is difficult to describe them in quantitative using any accurate mathematical model.This paper shows the hybrid approach known as genetic algorithm- support vector regression, which searchesbest suitable SVR parameters using genetic algorithm and then adopts best parameters to construct SVR model. For effective SVR model, parameters should be set carefully. The implementation will demonstrates the model is basically based on Root Mean Square Error. This will be very helpful to predict water quality.Also this paper classifies data analytical approach and Water quality prediction model for water quality assessment.*
Keywords: *Water Quality Prediction Model, Hybrid Approach, Genetic Algorithm, Support Vector Regression*

I. Introduction

Water is one of the most essential natural resource for the existence and survival of the entire life on this planet. But now a days discharges added by the pollution sources affect the pH value of water.This gives rise to many water-related problems such as water-borne diseases in organisms and deaths of aquatic animals and so on.[1]

Over the last decades water quality management has become increasingly complex. Increasing demand of society regarding quality of water, use and protection of water bodies leads to new views and strategies towards policy for water quality management.Water quality evaluation has been conducted using conventional approaches in all these years. These approaches involve manual collection and assessment of raw data. Water environments have been confronted with very serious pollution and increasing stress because of increasing demands for water by expanding industry, urbanization, agriculture, human population. Overcome with such kind of issues requires water quality sampling, monitoring and prediction of quality.

Water quality evaluation plays a very important role to monitor and control water pollution. The main objective of water quality evaluation is to check the quality of water to identify if a given sample of water is suitable enough for a given purpose. May be out of two samples of a given water from same resource but the different location may differ the taste because of their characteristics. May be out of two similar samples of water user might give preference to water which has good test. So test can become a quality evaluation parameter. The main objective of water quality evaluation is to decide quality standards or parameters for a water body based on desired use of water body.[1]

Here for this study sample data can be obtained from the USGS i.e. United States Geological Survey online resources for some particular time interval. The main objective of this model is to measure water parameters and analyzes the performance by applying hybrid model of Root Mean Squared Error (RMSE) and Support Vector Regression technique. [3]Also the classification method can also be applied, which is a two phase process, training and testing. Training is process where sample water parameters can be trained as per the requirement in supervised or unsupervised mode. And testing is the process to of checking how classifier has trained and labeled the unseen samples.[7]

Here the structure of the paper is as follows. In Section II gives the details about literature review which has been referred for this paper. Section III gives detailing about methodology such as Genetic Algorithm and support vector regression which has been used for proposed model. Section IV explains about process model or a flow chart for GA-SVR Model implementation. And finally section V concludes the studyof hybrid classification process model and direct and suggests the future investigation of water quality analysis.

II. Literature Review

The growing use of techniques in water quality management is seen in recent literatures. As few author are suggesting minimal number of parameters but with the acceptable accuracy to evaluate water quality. G. Kang et.al (2017) has reviewed the process for quality improvement of water based on Artificial Neural Network (ANN), Radial basis Function Network (RBFN) Model, Deep Belief Network Model.[1]shuangyinliu et. Al(2018) proposed hybrid approach of support vector machine with particle swarm for developing the accuracy of predictions.[7]Yafra Khan et.al (2014) has also suggested the prediction model using Artificial Neural Network(ANN).[3]Author (2014) has suggested a hybrid model that combines pattern based model with statistical group decision model is used to take normal range of parameter value which enhanced prediction for hourly sensor movementShuangyin Liu et.al (2013) has proposed Support vector regression model to solve the aquaculture water quality prediction problem based on Root mean Square Error(RMSE) and mean absolute percentage error(MAPE).[2]

III. Methodology

Genetic Algorithm:

Genetic Algorithm is a technique which can search large and complicated spaces using from natural genetics and evolutionary principle. [2]This algorithm always works with population of every genetic factor which is known as chromosome. This algorithm consists in maintaining a population of chromosomes, which represent solution to a problem. It is suitable to large scale problems. Initially GA starts with random selection of population. And then proceed with the operations i.e. selection, crossover and mutation. The selection operation selects the best individuals from the population which will help to reproduce the next. The crossover operation helps to exchange the gene of two selected individuals to produce new individuals. The mutation operation inverts certain bits of the selected individual to increase diversity.

Support Vector Regression:

Support vector regression works on similar principles as Support vector Machine. Support vector regression is a technique which follows principle of structural risk minimization, which has been successfully used for data classification and regression. [4].Generally simple regression technique tries to minimize the error rate. While in SVR the error fits within a certain threshold.

SVR can be extended to remote sensing applications in the management of land and water resources. SVR is mainly use to construct empirical models retrieving water quality variables using remote sensing images. The result shows that SVR can implement any non-linear mapping and produce better prediction than the traditional statistical method. [6]

IV. Process Model

This analysis system is composed of Data Acquisition, Data Standardization, pre-processing, Result and application i.e. evaluation of water quality parameters content.
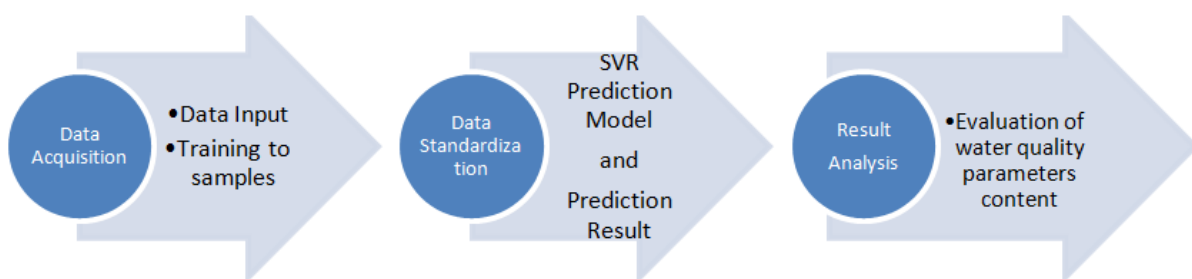

Figure no:1 Flow Chart for Water quality analysis system based on SVR Model.

And the implementation process of GA-SVR model is shown in figure 2. As mentioned in flow chart this system also consists of Data Acquired, Feature Extraction, Fault diagnosis, Training and Testing results and finally prediction.
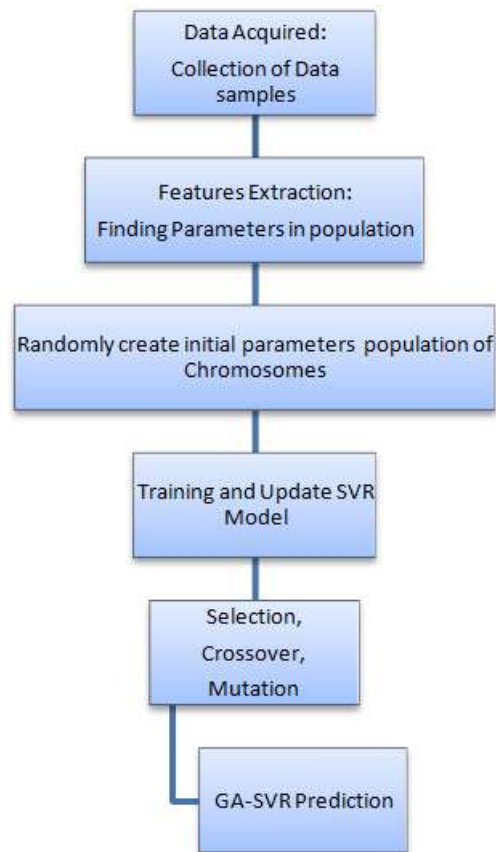
Figure no 2: GA-VSR Model

At a very initial stage parameters and their values should collect very carefully. After that feature extraction has to be proceeding with selected water parameters based on the requirements. Based on this approach SVR with genetic algorithm randomly creates initial parameters population of chromosomes. After this in training session fitness of the data set can be easily find out by the model. Therefore the fitness function is defined as the Root Mean Square Error cross-validation on the training data set.

1. Data Acquisition:

As shown in process the very first component is Data acquisition. In which, collect data samples from different geo locations at different time intervals. This can detect the parameters using the wireless sensor, such as pH, Dissolved Oxygen (DO), solar radiation, temperature, etc. As we employ the SVR for water quality parametric values of different locations considered the following points: Linear separating Hyperplanes and Multipleseparating Hyperplanes. This gives the classification of data.Even there are chances to collect faulty data sets in between the collection.

2. Data Standardization:

The original data sets can be collected after a specific time interval in a continuous time series. After collection of a data to train the data set data should be in standardize format. In order to eliminate the dimension differences the following formula can be used for data normalization and standardization.

$$X' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Here x represents the original data point , and x_{max}-$x_{min}$are the maximum and minimum values in data set respectively. [2]

Basically the parameters can select on trial and error basis.The design of a water quality model based on SVR does not restrict to standard procedure to determine the parameters to identify the water quality.

3. Result Analysis:

As mentioned above parameters can be selected on trial and error basis. Generally dissolved oxygen and temperature of water plays a very important role in quality of water. Then the subsequent monitoring parameters can also be compared and analyzed. For evaluating and comparing the performance of the SVR

model we can employ the Mean absolute percentage error (MAPE) and Root Mean Square Error (RMSE) to calculate or to evaluate accuracy. If the error values are smaller performance is better of the model.

V. Conclusion

Water quality prediction is very important in today's world.The actual implementation of hybrid approach of Genetic algorithm and Support vector Regression optimization provides reliable data on water quality prediction of large scale. The model which is studied here based on the hybrid approach for predicting water quality can be used to analyze prediction of water quality with different parameter. Implementation of some other measures may also improve the performance of SVR and GA.It can be observed that through this study, hybrid approach of Genetic algorithm and support vector regression process gives better water quality indications based on the given parameters. In this model operations of the genetic algorithm i.e. crossover, mutation may get different problems. So, by using advanced techniques to update this model with the new feature and services can be proposed in future development. This can be help to provide early warnings of the change of water quality to society.

References

[1].   G. Kang, J. Z. Gao and G. Xie, "Data-Driven Water Quality Analysis and Prediction: A Survey," 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService), San Francisco, CA, 2017, pp. 224-232. doi: 10.1109/BigDataService.2017.40

[2].   A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction ShuangyinLiuabHaijiangTaiaQishengDingaDaoliangLiaLongqinXubYaoguangWeiaMathematical and Computer Modelling Volume 58, Issues 3–4, August 2013, Pages 458-465

[3].   Predicting and analyzing water quality using Machine Learning: A comprehensive model, Y. Khan and C. S. See, 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY, 2016, pp. 1-6. doi: 10.1109/LISAT.2016.7494106

[4].   A hybrid approach of support vector machine with particle swarm optimization for water quality predictionW. Xuan, L. Jiake and X. Deti, "A hybrid approach of support vector machine with particle swarm optimization for water quality prediction," 2010 5th International Conference on Computer Science & Education, Hefei, 2010, pp. 1158-1163.
        doi: 10.1109/ICCSE.2010.5593697,Date Added to IEEE Xplore: 30 September 2010,Electronic ISBN: 978-1-4244-6005-2

[5].   Quality measurement classification for water treatment using neural network with reinforcement programming for weighting optimizationM. F. Dinniy, A. R. Barakhbah and E. M. Kusumaningtyas, "Quality measurement classification for water treatment using neural network with reinforcement programming for weighting optimization," 2016 International Conference on Knowledge Creation and Intelligent Computing (KCIC), Manado, 2016, pp. 126-133. doi: 10.1109/KCIC.2016.7883636 ,

[6].   Applying support vector regression to water quality modeling by remotesensing data,Xili Wang, Chansheng He, 10 Aug 2011

[7].   Fault diagnosis of water quality monitoring devices based on msvm and rbdt, shuangyinliu 1, longqin xu1, qiucheng li2,3, xuehua zhao4, daoliang li3, Received November 6, 2017, accepted January 17, 2018, date of publication February 9, 2018, date of current version May 9, 2018.

[8].   A hybrid prediction model for monitoring of River water quality in the USN system, Hoontae Kim , Minsoo Kim, October 2014, Interanational journal of Distributed sensor Networks , Volume 2015, Article ID 849287

[9].   Water quality assessment analysis by using combination of Bayesian and genetic algorithm approach in an urban lake, China ,LikunYangaXinhuaZhaoabSenPenga, Ecological Modeling, Volume 339, November 2016