www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 3 Issue 6 June, 2014 Page No. 6299-6306

Water Quality Prediction Using Data Mining techniques: A Survey

Shoba G, Dr. Shobha G.
Student, Dept of Computer Science &Engg.,
RVCE, Bangalore, India
E-mail: shobagoals@gmail.com

Prof & HoD, Dept of Computer Science &Engg.,
RVCE, Bangalore, India
E-mail: shobhag@rvce.edu.in

Abstract:

Data Mining is the set of activities used to find new, hidden, or unexpected patterns in data. Many organizations are now using these data mining techniques. Research in data mining continues growing in business and in learning organization over coming decades. Data mining methods may be classified by the function they perform or by their class of applications. Using this approach, four major categories of processing algorithms and rule approaches emerge: 1) Classification, 2) Association, 3) Sequence and 4) Cluster. This paper explores various data mining techniques like Artificial Neural Network, Back propagation, MLP, GRNN, Decision Tree etc. used in prediction of water quality. This survey focuses on how data mining techniques are used for water quality Prediction is analyzed.

Keywords: *ANN, SVM, GRNN, IDTL, Decision Tree, ANFIS, CA, FA, GIS*

I. INTRODUCTION

Water Quality is a major concern around the world. Water quality is affected by a wide range of natural and human influences. The most important of the natural influences are geological, hydrological and climatic, since these affect the quantity and quality of water available. Their influence is generally greatest when available water quantities are low and maximum use must be made of the limited resources; for example, high salinity is a frequent problem in arid and coastal areas. If the financial and technical resources are available, seawater or saline groundwater can be desalinated but in many circumstances this is not feasible. Thus, although water may be available in adequate quantities, its unsuitable quality limits the uses that can be made of it. Although the natural ecosystem is in harmony with natural water quality, any significant changes to water quality will usually be disruptive to the ecosystem.

The effects of human activities on water quality are both widespread and varied in the degree to which they disrupt the ecosystem and/or restrict water use. Pollution of water by human faeces, for example, is attributable to only one source, but the reasons for this type of pollution, its impact on water quality and the necessary remedial or preventive measures are varied. Faecal pollution may occur because there are n community facilities for waste disposal, because collection and treatment facilities are in adequate or improperly operated, or because on-site sanitation facilities (such as latrines) drain directly into aquifiers. The effects of faecal pollution vary. In developing countries intestinal disease is the main problem, while organic load and eutrophication may be of greater concern in developed countries (in the rivers into which the sewage or effluent is discharged and in the sea into which the rivers flow or sewage sludge is dumped). A single influence may, therefore, give rise to a number of water quality problesm, just as a problem may have a number of contributing influences. Eutrophication results not only from point sources, such as wastewater discharges with high nutrient loads, but also from diffuse sources such as run-off from livestock feedlots or agricultural land fertilized with organic and inorganic fertilizers. Pollution from diffuse sources, such as agriculture run-off, or from numerous small inputs over a wide area, such as faecal pollution from unsewered settlements, is particularly difficult to control.

The quality of water may be described in terms of the concentration and state (dissolved or particulate) of some or all the organic and inorganic material present in the water,

together with certain physical characteristics of the water. It is determined by *in situ* measurements and by examination of water samples on site or in the laboratory The main elements of water quality monitoring are, therefore, on-site measurements, the collection and analysis of water samples, the study and evaluation of the analytical results, and the reporting of the particular location and time at which that sample was taken. Once purpose of a monitoring programme is, therefore, to gather sufficient data (by means of regular or intensive sampling and analysis) to assess spatial and/or temporal variations in water quality. In this paper various data mining techniques are used to find water quality at different areas like river, reservoir,etc.

In this paper, various data mining algorithms are analyzed in prediction of water quality in different area.

II Related Works

Hao Liao and Wen sun [1] has proposed feed forward back propagation and improved decision tree (IDTL) method for water quality prediction. Forecasting model based on improved decision tree combines artificial neural network and decision tree algorithm the IDTL predicts n+1 month based on the known N month. Parameters considered are dissolved oxygen DO five-day biochemical oxygen demand BOD_5 and ammonia-nitrogen levels. IDTL model predicts 85% of accuracy compared units decision tree model which predicts accuracy 70%. This approach has advantage in the clustering of data when processing the data with stress on inter-dependence between parameter.

Miao qun, Gao ying, Liu Zhiquiang and Tan Xiadhui [2] proposed to evaluate surface water quality a comprehensive water quality identification index, based on single factor water quality identification index tool. This method evaluates water quality qualitatively and quantitatively. The result of Valuation shows that the river water quality standards reach the regional water environment function III, which is in accordance with spatial and temporal distribution of pollutants from the evaluation process the comprehensive WQI calculated simple operability strong and the conclusion of scientific rationality.

Yu-Hu, Hai-feng and Cong [3] proposed multiple statistical methods including cluster analysis (CA) and factor analysis (FA) to evaluate the temporal –spatial variations of water quality of Luzhi river system in plain river-net areas, Suzhou, China. The results show that water quality of Luzhi river system exists obvious temporal-spatial distribution characteristic, The sampling periods were categorized into two groups corresponding to seasonal changes in water quality, Three groups were classified based on sampling sites which represented different water quality levels and pollutions degrees. Three factors were extracted using FA method and they reflected nitrogen and phosphorous levels of the water, organic and physical conditions respectively. The varifactors obtained from FA indicate that the parameters responsible for water quality variations of different monitoring site are mainly related to corresponding geographical location pollution source situations and hydrodynamic characteristics. The study provides critical information for water quality management in the Luzhi River –net areas.

Yang ,Li and Zhang [4] developed two adaptive neural networks based predicting models are implemented for river water quality. One predicting model is used for anticipating the lower course water quality by measuring the upriver water quality and another model is estimating the future state with current water quality. Parameter (biological oxygen demand),dissolved oxygen, chemical oxygen demand, temperature, velocity of stream, L(water level), ΔDO(DO variation),ΔTOC(TOC variation), ΔV(V variation),and ΔL(L variation). The output variables are selected as DO and COD. Here two Neural Network is two MISO (multi –In singular – out) feed forward NN, with single hidden layer. So it can avoid intricacy net structure, enhance training speed, and add predicts items expediently. In this paper, a technique called time window removement is employed to carry out d-step-ahead prediction. Namely, the predicted value Y(K+1) as datum to predict Y(K+2),and go on. The learning algorithms with orthogonal basis transfer function for state and dynamic neural network is given. Both the neuron numbers and orthogonal basis transfer function can be established automatically in training process. Local extremum problem does not exist in this method. Accordingly, the simulation results prove that the two adaptive NNs approaches have high precision, good adaptability and extensive applicability.

Tin-suo ,Tling – Lis and Crun-yan [5] proposed Artificial neural networks successfully used to model primary production and predict one-step weekly algae blooms in reservoir. Parameters-turbidity(TUR), PH, chloride(CHL), hardness(HAR), alkalinity(AIK), ammonia(NH4), nitrite nitrogen (NO3), total organic nitrogen (TON), chemical oxygen demand (COD), temperature (T) and chlorophyll-a (CHL-a). Chlorophyll-a model to predict the water quality for two days later is too established to assist the determination of the operation of emergency system for water treatment process. Chlorophyll-a of source water is Tianjin two days later is established and tested for simulation on the basis of systematic method of neural network. The conclusions are Correlation coefficient and index clustering is used to analyze the previous data of water quality in Tianjin. Meanwhile, 26 schemes of input variables are put forward and each scheme is tested by ten times of experiments. Then chlorophyll-a, turbidity, water temperature, ammonia, PH and alkanity are determined as the proper input variables of the model. Through different methods, 27 schemes of input variables and their combinations are established for experiment optimization. By simulation testing, the correlates coefficient b/w prediction and real values of chlorophyll-a reaches 0.88.Then the accuracy can go up to 85% when the amt of chlorophyll-a divided into four grades.

Tinsuoler, and Tinglin huang [6] proposed Decision making tree to forecast the chlorophyll level of the next day. Real data testing shows the forecast results are good. The authors determine by analyzing three factors that affect the chlorophyll level of next day as follows. The current content of chlorophyll, the solar radiation and the change of DO. The results of simulation show that the prediction accuracy can reach 80%.In this study, they collected 115 daily measured data about indirect monitoring of raw water quality of algae and solar irradiance data from online monitoring and direct-line artificial monitoring of chlorophyll content of raw water. This paper enables online monitoring data and artificial monitoring data with same dimension, combined with the algal growth dynamics, they transform several online monitoring data of dissolved oxygen and solar irradiance data is one day into one data per day, that is mean calculating the average standard deviation and average 100 sets of data are used to train and set up decision making tree model which is to predict the level of chlorophyll is next day.

Sirilak and siripun [7] built a multilayer perception(MLP) neural network using the levenberg –macquardt (LM) algorithm to classify water quality of canals in Bangkok. Parameters are pH value, dissolved oxygen(DO),and biochemical oxygen demand(BOP).The classifications of canal water quality are based on surface water standards. The lower the number class, the better the quality of water quality. The results indicate that the NN perform with a high accuracy classification percentage of 99.34%.The cost and time of water resource management could be minimized. LM-Optimization algorithm outperforms simple gradient descent and other conjugate gradient methods in a wide variety of problems. It approximates steepest –descent method.

Liu and He Xue [8] proposed fuzzy c-means clustering method to classify and assess rural surface water quality based on monitoring data from 33 typical stations in 23 rural rivers and 4 reservoirs in Lianyungang city. In monitoring 33 stations, the results are classified into 3 clusters in terms of water quality. The first cluster is nearly or better than the national surface water grade II standards, while the second and third clusters worse than the national grade v standards and the index values in the third cluster exceed those in second cluster. And also consideration of principle of maximum membership degree, 27 stations classified into first cluster, 5 stations in 2^{nd} cluster and 1 station into 3^{rd}, means most of the monitoring stations meet the national grade II or III surface water standard and the overall quality of rural surface water in Lianyungang city remains good, but exist some sections contaminated with ammonia nitrogen and petroleum, so safety and risk assessment system is necessary. In this paper, the water quality clustering is carried out by using math's programming method with the FCM function of math's embedded. Fuzzy C-means clustering is a kind of automatic clustering method without any intervention or supervision. As it does not require much more auxiliary information and therefore is early to be accepted.(Among all the water quality indexes, seven indexes including permanganate index,BOD5,ammonia nitrogen, petroleum, volatile phenol, total Hg and total Pb are chosen for the following water quality clustering and assessment.)

Jorge Camejo, Osvaldo Pacheco and Miguel Guevara [9] proposed artificial intelligence techniques in combination with Delphi method is applied for classification of drinking water quality. The Delphi technique which is used National Sanitation Foundation (NFS) was improved using kumar method and PART, ANN and KNN algorithm are used for water quality classification in real time. By comparing the three algorithms by iterations, KNN algorithm has given slightly better results. By using kumar method, they automatically compute water quality index to classify the datasets elements into 5 classes (excellent, good, medium, bad, very bad). Selecting best performed real time features on results of classified datasets. Then, exploring MLT (eg. Decision tree, Artificial Neural Networks and K-nearest neighbor) for producing DWQ (Drinking water quality) classifiers. Here, they perform the classification of two classes (good and Pmedium) out of 5 possible categories, due to the absence of vectors in the datasets. Parameters are pH, Dissolved oxygen, Nitrate and temperature.

Elaheh, Farzad , Rehan, Mina, Homayoun and Manuel. J. Rodriguez [10] proposed fuzzy rule based system with final fuzzy evidential reasoning to incorporate ambiguity and uncertainty involved in drinking water quality assessment with linguistic environment. The proposed method has advantage which has ability to deal with uncertain water quality parameters (WQP's) and to model nonlinear relationship between WQP's and the relative water quality. A mamdani type inference is used to fire the weighted rules that are designed to assess the relative quality of drinking water in the WDN (Water Distribution Network). The input weights are evaluated based on AHP method or subjective judgment of the expert. Two methods including "minimum" and "maximum" are common to be applied to obtain a new fuzzy set as output. 'Minimum' operator has been used to evaluate new output fuzzy memberships. In other words, each rule's output will be combined to form single fuzzy set using a fuzzy aggregation operator. Maximum operator is used as aggregating operator. Fuzzy Dempster-Shafer theory of evidence is used to perform the final aggregation of individual fuzzy rule-based systems for each group of microbial and physiochemical parameters. Dempster-shafer is used to overcome drawbacks of Bayesian assignments. The methodology used reports (H-VH) quality for drinking water at most of the sampling locations in Quebec City main WDN. The fuzzy rule based results from microbial and physiochemical group is combined using fuzzy evidential reasoning. The decision making about the recommended

water quality grade is performed on comparison between pignistic probabilities.

Maqbool Ali, Ali Mustafa qamar [11] presented average linkage (within groups) method of Hierarchical clustering using Euclidean distance for finding water quality index. Multilayer perceptron is used for classification. For clustering, k-means as well as Hierarchical techniques using different intervals were applied. For classifications, Nearest Neighbors, using Euclidean and city block distances. Neural networks, using MLP and Radial basis function, and support vector machines techniques are used average linkage (within groups) method of Hierarchical clustering using Euclidean distance is an accurate unsupervised learning technique for finding water quality index. And for classifications, multi layer perceptron has been found to be more accurate supervised learning technique. Higher values of fecal coliforms were found in the months of march, june, july and October.

Fi-John chang and Ya-ting chang [12] built adaptive neuro fuzzy inference system a 1-3 hours ahead a prediction model for reservoir management. Two ANFIS models for water level forecast is developed, one with human decision of reservoir outflow as input variable, another without. Fuzzy logic is recommended for modeling reservoir operation to solve the inherent imprecision and vagueness characteristics in reservoirs. ANFIS possesses good capability of learning, expensing, constructing and classifying. Since ANFIS allows extracting fuzzy rules from numerical data or expert knowledge and adaptively constructs a rule base. It can further tune complicated conversion of human intelligence to fuzzy systems. The Bell-shaped membership function is used. The bell-shaped membership functions have one more parameters than Gaussian membership functions, so a non fuzzy set can be approached when the free parameter is tuned. The model with human decision as input variable has consistently superior performance with regard to all used indexes than the model without this important input. ANFIS model provide accurate and reliable water level prediction for next three time steps.

Natarajan,Samson and Ganapathiram [13] proposed fuzzy logic and GIS for ground water quality for potability. The decision on water quality assessment for potability gives water is desirable, acceptable and not acceptable as per guidelines from BIS and WHO regulatory bodies. But, in border line cases of water quality parameters, the different types of uncertainties are involved at various part of experimental and measurement process right from sampling, sample storage, processing and analysis. Monitored sets of data and limits should not be crisp set, but fuzzy sets. One way of avoiding the difficulty is uncertainty handling in water quality assessment is by introducing a margin of safety or degree of precaution before applying a single value to drinking water quality standards. The analysis functions in a GIS allow manipulation of multiple themes of spatial data to perform overlays, buffering and arithmetic operations on the data with its spatial analysis capabilities. 66 rules were designed for physiochemical water quality parameters as group I, 73 rules were designed for group II. Results from group I and group II are combined with group III, to assess final classification of water. A total of 27 rules were fired using mamdani amplification of min and max operator for final assessment of ground water quality in fuzzy logic models. The results from all the three groups were aggregated to assess the final classification of water. To assess drinking water quality of ground water samples, 181 rules are fired. A fuzzy rule based system is generated in which users classify the water with respect of different parameters, all connected using AND operator. Delphi's technique is applied to converge the feedback of various users to a single value. Fuzzy logic simulation approach has shown to be practical, simple, and useful tool to assess ground water quality assessment for potability. 79 sampling wells for 3 years monitored and major recorded data revealed that concentrations of cations and anions, were above maximum, desirable for human consumption. EC is important parameter within input parameter used in this modeling.79 samples bore wells were considered. According to the result, the ground water meets WHO drinking water 67.5% and 92.5% during pre monsoon. As sampling stations 24 and 73 were non potable due to vicinity of waste water discharging areas and solid waste dumping sets. During pre monsoon satisfies 67.5% and 92.5%.

Rahimi and Mokarram [14] proposed fuzzy logic for assessing the ground water quality. Four parameters Ec, Mg, Na and Ca are selected for water quality analysis. First, the interpolation using of 214 sampling points are developed for each of the parameters with kriging method for production map for each one of parameters model. Second, the calculation of the fuzzy memberships for Ec, Mg, Na and Ca is evaluated using linear function. Finally, for ground water quality, is required to calculate the convex combination of raster values containing the different fuzzy parameters. The weights are calculated through Analytical Hierarchical process (AHP) which relies on pair wise comparison matrices which relate to different components and are assigned values according to their relative importance. The fuzzy theory showed 14.22% of study area as having high ground water quality, 19.86% having moderate quality, and 48.9% have marginal quality and 16.99% have poor quality. In order to evaluate and present a better method eight different points were randomly chosen and Mg, Na, Ca and Ec were measured. The points are plotted on the comparison of the fuzzy and Wilcox methods. The development of GIS has contributed to facilitate the mapping the groundwater quality results, both Boolean and fuzzy, but the topological rules implied in GIS software are based on crisp theory.

Sundarambal palani, sie-yui-liong, pavel tkalich [15] proposed ANN models such as Back propagation (BP) (Wardnet) and GRNN to predict salinity, water temperature , dissolved oxygen and chl-a concentrations in Singapore coastal water both temporally and spatially using continuous weekly measurements of water quality variables a different stations. A prediction capability which was tested is found to be faster than from process-based model with minimal input requirements. According to performance in water quality parameters forecasting, the wardnet is better architecture for the temperature and salinity models, but GRNN is better for temperature and salinity models, but GRNN is better for DO and chl-a models. A time lag up to $2\Delta t$ appeared to suffice to yield good simulation results. Simulation accuracy, measured in the Nash- Sutcliffe coefficient of effiecincy (R^2), ranged from 0.8 to 0.9 for the training and overfitting test data. BP is especially good at fitting high dimensional, continuously-valued functional approximations to data.

Vahid Nourani, Tohid Rezapour Khanghah and Milad sayyadi [16] proposed feed forward neural network (FFNN) with BP to estimate the quality of treated water using qualitative data of zarrineh rood water treatment plant. The objective of BP is to find the optimal weights, which would generate an output vector, as close as possible to the target values of the output vector, with selected accuracy. Values of temperature , pH, opacity, total hardness and quantity of calcium before the treatment process were considered as input of the models to predict the values of Ec and TDS (after treatment process) which can determine the qualitative condition of treated water. FFNN models to predict TDS and Ec in term of correlation coefficient (more than 0.90) and coefficient of determination (more than 0.70) in the calibration and verification steps. The results of the proposed model were compared with other classical and practical models and FFNN proved high merit in predicting the parameters of the treated water.

Chen xiaopong and huang he [17] applied Immune feed forward neural network (IFNN) for fault detection of water quality monitoring equipment. The network architecture, activation functions, learning parameters, and training methods encoded in a binary string so that network can be optimized using the immune algorithm. The network inputs are pH, conductivity, temperature, and dissolved oxygen at time t and previous time step t-1. If the difference between the current values of one parameter is excessive, the system is in error. Thus the input variable is an 8-dimensional vector. In addition to the normal working status, the fault status is divided into five types, pH sensor failure, conductivity sensor failure, temperature sensor failure, water cycle controller failure, and line fault. The network output is then a 5-dimensional vector corresponding to the normal state and four faults. An output variable bit of 1 indicates the system is in corresponding state. IFNN performance is compared with that of EFNN based on the number of training generations, epochs, the convergence speed, and actual MSE that meets the minimum MSE. The optimized network has been successfully applied to fault detection in a water quality monitoring system. Test show that the Immune network is faster and more accurate than the feed forward network.

Kwok-wing chau and nitin muttil [18] employed data mining using box plots and multivariate statistical analysis using factor analysis for spatio-temporal analysis of coastal water quality data from Tolo harbor, Hong Kong. The results from statistical analysis of box plots revealed pronounce spatial and temporal patterns and the heterogeneity of the parameters have been studied. The studies of spatial heterogeneity shows, out of three monitoring stations in the Harbour su zone, TMZ are the most susceptible to eutrophication with the highest nutrient and algal biomass concentrations. The factor analysis indicates nutrient and hydro – biological processes to be most dominant and the external environment factors seem to be relatively less dominant. So by using factor analysis nutrients from external sources like pollutants loadings from point sources have been significantly reduced. So steps to control pollution loadings from external sources, it is necessary to undertake steps to control pollutant loadings from internal sources also.

F.Karimipour, M. R. Delavar, and M.Kinaie [19] explore the relationships and future trends of data that are essential for water quality management using Geospatial information system (GIS). GIS and geospatial data mining is used for information extraction which are considered for the importance of spatial component in water quality management. The results regarding spatial distribution of water quality parameters are achieved using GIS. The results have clearly identified the relationship between number and location in industrial pollutions and water quality indicators to be used in environmental protection and land use planning.

Sirilak areerachakul and Siripun sanguansintunkul [20] proposed CART and MLP neural network using Levenberg-Marquardt algorithm to classify water quality. K-folds cross validation to find model classify water quality applied to CART and MLP. According to results, CART performs with high accuracy classification percentage of 99.96% while MLP show percent accuracy of 98.82%. CART has advantage of expressing regularities explicitly and thus being convenience to inspect for water quality validity. Levenberg – Marquardt has faster convergence and more accurate towards error minimum. As a result the cost and time of water resource management could be minimized.

Iuana, Dusan, Ljiljana, comic, aleksandar, marina and nenad [21] proposed combination of cluster analyses (CA) and classification for analyses of total coliforms in Gruza and Grosnica reservoirs. Classification is done using decision tree, it is possible to detect the relation of total coliforms and some other physiochemical or biological parameter in the

reservoirs. Results indicate that it is possible with help of CA, to separate localities and depths showing the difference in number of TC. CA with k-means algorithm gives satisfying results about the analyses of physicochemical and biological parameters in water monitoring. The system determines the accuracy of analyses by validity percentage. The analyses show that the number of total coliforms is connected to anthropogenic activity, the amount of organic matter, as well as to presence of bacterial community which is not dominant or characteristics for the specific reservoir.

Toochukwa and Francisca [22] proposed multilayer feed forward network in prediction of cryptosporidium and giardia peak concentrations in Kano river. Eight water quality parameters are used to predict crypotospordium peak concentration and seven parameters were used to model giardia concentration in Kana river, Nigeria. The ANN models correctly predicted oocysts and cysts concentration with accuracy of 90% and 92% respectively. The neural Network model gave excellent result. ANN has ability to learn through training and resembles biological neural network in structure.

Saiful, haffir, shah and nooritawati [23] built sensors of pH and Do are connected to stamp circuit that acted as universal asynchronous receiver/ transmitter (UART) whilst temperature sensors is connected to microcontroller. Then fuzzy algorithm is developed based on water quality standards for decision making. thus both hardware and software are integrated and initial findings showed that the proposed system is apt for monitoring of water quality.

Li and wang [24] proposed bayes for water quality assessment. Bayes is very effective and simple method and has low computational complexity. Its feasibility is verified through water quality monitoring data of three gorges and its evaluation on water quality is impersonal and reasonable. So , it can be applied in the practical assessment. The results show water quality condition in three gorges reservoir area is good, but ammonia nitrogen content in it is higher.

Ji-hong, Biao, cai-lian, chang and Jun [25] applied fuzzy comprehensive evaluation method for monitoring data of water quality evaluation in liuziazhuang of Qing zhang river. The results show changes about water quality is not obvious. Effective measures for water quality monitoring to protect the health of people, seek the sustainable development of society, economy, and environment, and make use of water resources and water environment forever.

Yang , zengchuen, Qing hang [26] constructed ANN with shuffled frog leaping algorithm for optimization a combined water quality assessment model. SFLA was applied to train the initialized data from water quality criteria for optimizing the connection weights and thresholds of neural network. It can provide information for decision makers. This model provides a new way for water quality assessment.

Xue , chen [27] proposed back propagation network model for water quality evaluation of surface water based. Back propagation network model strong classifying ability; it can complete the comprehensive level evaluation in water quality assessment. Compared with traditional evaluation methods, the assessment results of BP network model are more objective and reasonable. The BP network model can also select a random number of evaluation parameters according to different requirements to establish different water quality evaluation models, which shows strong flexibility and adaptability as well as bright future of this method.

Sinin, nannan , zhennan and fengbing [28] proposed BP and RBF for evaluation of water quality in Fuyang river in handan city. The results of evaluation of water quality grade are III. RBF is superior to BP neural network training process. BP neural network readily falls into local minimum. RBF with Guassian function is used. The traditional method need to determine the connection weights artificially. So affects the reliability of evaluation models.

Zhenxiang , Qiang and Dong liu [29] built a fuzzy comprehensive evaluation model based on entropy weight method (FCE-EW) to assess the actual state of underground water quality. The entropy weight method can describe the influence weight of each index on the water quality evaluation objectively driven by sample data. FCE-EW is easy to operate, so it is widely used to evaluate water quality. The calculation results compared to that RAGABP method and PPC, but FCE-EW were more objective.

Changjun and Qinghua [30] proposed a grey clustering model to evaluate water quality. The evaluation results based on the assess of water quality of 20 sections in Suzhou river was compared with that of traditional method, the performance of proposed model is practically feasible in the application of water quality assessment and its application is simple. The grey clustering which can overcome the disadvantages of single factors can reflect the water quality at present.

III Conclusion

In this paper an analysis is presented for water quality prediction using various data mining techniques at different locations. According to analyses Back Propagation is used frequently.

REFERENCES

[1] Hao Liao and Wen Sun, "Forecasting and Evaluating Water quality of Chao Lake based on an Improved Decision Tree Method", International Society for Environmental Information Sciences, Procedia Environmental Sciences 2, 970-979, 2010.

[2] Miao Qun, Gao Ying, Liu Zhiqiang and Tan Xiaohui , "Application of Comprehensive Water Quality Identification Index in Water Quality Assessment of River", Global Congress on Intelligent System , 2009.

[3] Yu-Hu Zhang, Hai-feng Jia, Cong Yang, "Assessment of Spatial-Temporal variations in Surface water quality of Luzhi river system in Plain River-Net Areas, Suzhou, China", IEEE, 2011.

[4] Yang Yamin, Li Ying and Zhang Yun, "The Orthogonal Basis Neural Network Based Prediction Modeling for River Water Quality", Proceedings of American Control Conference, Arlington, VA June 25-27, 2001.

[5] Jin-Suo LU, Ting-Lin HUANG and Chun-Yan WANG, "Data Mining on Source Water quality (Tiajin,China) for forecasting Algea Bloom Based on Artificial Neural Network (ANN)", World Congress on Computer Science and Information Science, 2009.

[6] Jinsuo Lu and Tinglin Huang, "Data Mining on forecast Raw Water Quality from online Monitoring station Based on Decision-Making Tree", Fifth International Joint Conference on INC,IMS and IDC, 2009.

[7] Sirilak Areerachakul and Siripun Sanguansintukul, "Water Quality Classification using Neural Networks: Case Study of Canals in Bangkok, Thailand, Institute of Electrical and Electronics Engineers, 2009.

[8] Liu Fu-Cheng and He Xue-Zhao, "Application of Fuzzy c-means Clustering for assessing rural surface Water Quality in Lianyungang City", Fifth conference of Measuring Technology and Mechatronics Automation, 2013.

[9] Jorge Camejo, Osvaldo Pacheco , Miguel Guevara, "Classifier for Drinking Water Quality in Real Time", Foundation for Science and Technology, IEEE, 2013.

[10] Elaheh Aghaarabi, Farzad Aminravan, Rehan Sadiq, Mina Hoorfar, Homayoun Najjarana and Manuel J. Rodriguez, "A Fuzzy Rule-Based Approach for Water Quality Assessment in the Distribution Network:, IEEE, 2013.

[11] Maqbool Ali, Mustafa Qamar, " Data Analysis, Quality Indexing and Prediction of Water Quality for the Management of Rawal Watershed in Pakistan", IEEE, 2013.

[12] Fi-John Chang, Ya-Ting Chang, "Adaptive neuro-fuzzy inference system for prediction of water level in reservoir", Advances in Water Reources, 29, 1-10, 2005.

[13] Natarajan Venkat Kumar, Samson Mathew and Ganapathiram Swaminathan, "A Hybrid Approach towards the Assessment of GroundWater Quality for Potability: A Fuzzy Logic and GIS Based Case Study of Tiruchirappalli City, India" , Journal of Geographic Information System, 2, 152-162, 2010.

[14] Rahimi.D, Mokarram. M, "Assessing the groundwater quality by applying fuzzy logic in GIS environment – A case study in Southwest Iran", International Journal of Environment Sciences, Volume 2, No 3, 2012.

[15] Sundarambal Palani, Shie-Yui Liong, Pavel Tkalich, "An ANN application for Water Quality Forecasting", Marine Pollution Bulletein 56, 1586-1597, 2008.

[16] Vahid Nourani, Tohid Rezapour Khanghah and Milad Sayyadi, "Application of the Artificial Neural Network to monitor the quality of treated water", International Journal of Management and Information Technology", Volume 3, No 1, January, 2013.

[17] Chen Xiaoping, Huang He, "Immune Feedforward Neural Network for Fault Detection", Tsinghua Science and Technology, ISSN 1007-0214 07/16 pp272-277 Volume 16, Number 3 ,June 2011.

[18] Kwok-wing chau and Nitin Muttil, "Data mining and multivariate statistical analysis for ecological system in coastal waters", Journal of Hydroinformatics, 09.4, 2007.

[19] F. Karimipour, M.R.Delavar and M.Kinaie, "Water Quality Management Using GIS data Mining", Journal of Environmental Informatics, 1726-2135/1684-8799, 2005.

[20] Sirilak Areerachakul, Siripun Sanguansintukal, "Classification and Regression Trees and MLP Neural Network to Classify Water Quality of Canals in Bangkok, Thailand, International Journal of Intelligent Computing Research, Volume 1, Issue 1/2 , March/June 2010.

[21] Ivana D.Radojevic, Dusan M.Stefanovic, Ljiljana R. Comic, Aleksandar m.Ostojic, Marina D.Topuzovic and Nenad D. Stefanovic, "Total coliforms and data mining as a tool in water quality monitoring" , African Journal of Microbiology Research vol. (6), pp.2346-2356, 16 March, 2012.

[22] Toochukwu c.Ogwueleka, Francisca N.Ogwueleka, "Data mining Application in Predicting Cryptosporidium SPP.OOcysts and Giardia SPP. Cysts concentration in Rivers", Journal of Engineering Science and Technology, Vol. 5, No.3, 342-349, 2010.

[23] Saiful Azwad Saipol Azhar, Haffiz Johar, Shah Rizam Mohd Shah Baki, & Nooritawati Md Tahir, "Optimization of Water Quality Monitoring Based on Fuzzy Algorithms", IEEE Conference on systems, Process & control, 13-15 December, 2013.

[24] Li Chuanqi, Wang Wei, "Assessment of the Water Quality near the dam area of Three Gorges Reservoir Based on Bayes", The 1st International Conference on Information Science and Engineering, 2009.

[25] Ji-hong Zhou, Biao Zhang, Cai-Lian Hao, Chang-jun Zhu, Jun-guang Zhao, "The Fuzzy Evalutaion

of Surface Water Quality in LiuJiaZhuang of QingZhang River", International Conference on Environmental Science and Information Application Technology, 2009.

[26] Yan Zhao, ZhegChuan Dong and QingHang Li, "ANN based on SFLA for Surface Water Quality Evaluation", International Conference on Transportation, Mechanical, and Eletrical Engineering, December 16-18, Changchun, China, 2011.

[27] XUE Xicheng, CHEN Yan , "Water Quality Evaluation of Surface Water Based on Back Propogation Neural Network", 2011.

[28] Simin Li, Nannan Zhao, Zhennan Shi, Fengbing Tang, "Application of Artificial Neural Network on Water Quality Evaluation of Fuyang River in Handan City", 2010.

[29] Zhenxiang Xing, Qiang Fu, Dong Liu, "Water Quality Evaluation by the Fuzzy Comprehensive Evaluation based on EW method", Eighth International Conference on Fuzzy Systems and Knowledge Discover, 2011.

[30] Changjun Zhu, Qinghua Liu, "Evalution of water Quality using Grey Clustering", Second International workshop on Knowledge Discovery and Data Mining, 2009.