DATA MINING TECHNIQUES FOR WATER QUALITY MONITORING SYSTEM

Ms. Swapnali D. Mahadik
PhD Research scholar, TMV,
Asst. Prof. MCA Department,
YMT College of Management, Navi Mumbai.
swapn.mahadik30@gmail.com

Dr. Anup Girdhar
Ph.D. Guide, TMV
anupgirdhar@gmail.com

ABSTRACT

Water is one of the most basic element supporting life and environment for every leaving as well as non-leaving thing. For predicting a water quality now a day's lots of techniques are implemented like Data Mining, Remote Sensing, etc. Data Mining is becoming the most popular technique for handling huge amount of water and its related data. At present The Central Pollution Control Board (CPCB) provides data about water and its quality which is very difficult to understand so it is necessary to build a data model to monitor and analyze the water quality based on the defined parameters. This paper represents how to handle the large amount of data with the help of Data Mining basics and clustering technique with the K-means to analyze the water quality based on the predefined water parameters.

KEYWORDS: *Central Pollution Control Board (CPCB), Clustering, Data Mining, K means Clustering, Remote Sensing.*

I. INTRODUCTION

The quality of water may be described in terms of the concentration and dissolved state of some organic and inorganic materials present in the water. The quality of water is judged by authenticate standards. World Health Organization has recommended guidelines for water quality standards, which can be taken as base values for formulating other values for predicting a quality.

The purpose of proposed water quality monitoring system is to gather sufficient and relevant data to assess spatial variations in water quality. Water quality reflects the composition of water as affected by natural processes and human activities, and therefore a need to establish water quality in the natural hydrological cycle [10]. This paper represents how Data Mining techniques are used to find and analyze water quality at different areas depending upon their attributes. The diversity and scope of information to be processed for proper trend analysis towards water quality management makes the use of Data Mining concepts all the more appropriate.

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 [1]. The traditional monitoring methods have many shortcomings, such as long cycle, discontinuities in time and space, higher cost and so on [8]. The integration of remotely sensed data, GIS,

Copyright © 2017 [Cyber Times]. All Rights Reserved.

GPS techniques provides an important input to the Data Mining tools for monitoring and assessing various water resources. Also Remote sensing techniques are widely used in water quality assessment [13].

The water quality index indicated that most of the sampling locations come under good category indicating the suitability of water for human use [9].

Where these techniques have helped to find and estimate pollutants in water with lower costs and greater potential [7].

II. DATA MINING

Data Mining is one of the important research areas motivate to find the meaningful information from large number of datasets [14].

At every level, every organization collects required data and processed to extract the expected solutions. When existing data is already loaded, some extra information can be extracted from the large amount of data and this process can be called as Data Mining.

In other words, in Water Monitoring System Data Mining can be considered as an approach to determine the valid and understandable data patterns in huge databases which give better and simple interpretations.

Proposed water Quality Monitoring system follows Data Mining process as shown in Figure 1, which includes several stages such as data cleaning, data selection, data reduction, data extraction, data interpretation and finally report analysis.

Even in the In the Knowledge Discovery Database process, data cleaning and preprocessing is main step before finalizing the Data Mining algorithms.

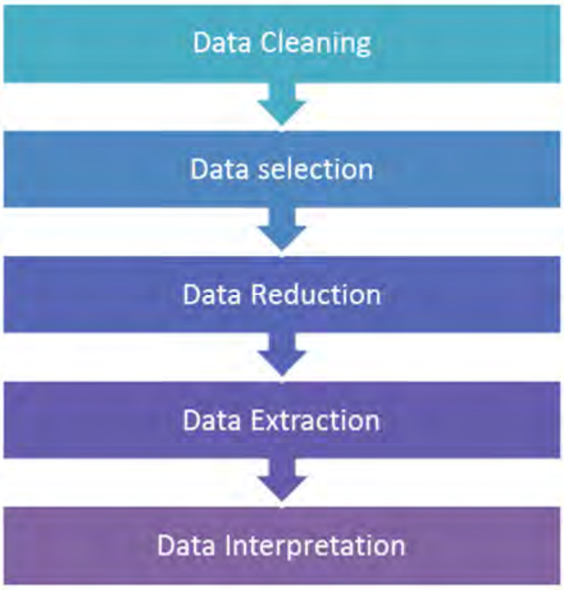

Figure 1: Steps of Data Mining Process for WQMS

Proposed Water monitoring system provides each and every stage flexible to perform its own process and provide input to next process. Data cleaning includes cleaning of data which removes missing, noise, duplications and erroneous issue [15]. Data Selection provides some of the data in order to be used in analysis to maintain the integrity. Data Reduction gives facility to reselection of data i.e. from the selected data it reduces unused data from the water attribute point of view. In the next stage after data investigation, data patterns are extracted to interpret the final results i.e. prediction of water quality.

III. DATA GENERATION

The study suggests appropriate number of water samples can be collected and the selected water samples with their quality parameters are measured by carrying out the above mentioned stages. Quality parameters can be pH, Biological Oxygen demand (BOD), Sulphates, nitrates, chloride, dissolved solids, sodium, potassium, phosphate, etc. These data can generate and parameters can be analyzed on the basis of predefined locations.

Copyright © 2017 [Cyber Times]. All Rights Reserved.

Water Quality Index is also computed to reduce the large amount of water quality data to a single numerical value that expresses the overall water quality at a certain location and time based on several water quality parameters. It is also defined as a rating reflecting the composite influence of different water quality parameters on the overall quality of water. [3].

The main aim of water quality index is to turn complex water quality data into information that is understandable by the public. This Water Quality Index based on some very important parameters can provide a simple indicator of water quality.

The Water Quality Index measures the scope, frequency, and amplitude of water quality and then combines the three measures into one score. This calculation produces a result between range of 0 and 100. The higher the score the better the quality of water.

The scores are then ranked into one of the five categories described below:

- **Excellent**: (WQI Value 95-100) - Water quality is protected with a virtual absence of impairment; conditions are very close to pristine levels. These index values can only be obtained if all measurements meet recommended guidelines.
- **Very Good**: (WQI Value 89-94) - Water quality is protected with a slight presence of impairment; conditions are close to pristine levels.
- **Good**: (WQI Value 80-88) - Water quality is protected with only a minor degree of impairment; conditions rarely depart from desirable levels.
- **Fair**: (WQI Value 65-79) - Water quality is usually protected but occasionally impaired; conditions sometimes depart from desirable levels.
- **Marginal**: (WQI Value 45-64) - Water quality is frequently impaired; conditions often depart from desirable levels.
- **Poor**: (WQI Value 0-44) - Water quality is almost always impaired; conditions usually depart from desirable levels.

WQI scores are computed for each public water supply system that has been sampled in a sampling season [4].

IV. ANALYTICAL DATA REVIEW

Based on character and quality of various information sources, water records are processed to identify the records and met the quality objectives and then imported to statistical analysis packages. With this data can be analyzed using statistical methodology for spatial, temporal parametric trends [2].

As study suggests Cluster analysis evaluates spatial variations of water quality. Sampling can be categorized into various groups such as seasonal change, location wise change, etc. this groups can produce different water quality levels and pollution degrees as per the geographical locations. The main objective of cluster analysis is to reduce the number of monitoring stations to reduce the frequency of sampling and to minimize the number of parameters for which the samples should be tested [5].

Applications of Cluster Analysis:

Cluster Analysis organizes sampling entities into discrete groups, such as maximum similarities within a group and minimum among group of criteria. This is an exploratory analysis tool for solving classification problems [11], [15]. Its object is to sort cases, data, or objects (events, people, things, etc.) into groups or clusters. The resulting clusters of objects should exhibit high internal (within-clusters) homogeneity and high external (between-clusters) heterogeneity. [12]

Copyright © 2017 [Cyber Times]. All Rights Reserved.

V. K-MEAN CLUSTERING

In proposed water monitoring system, K-mean clustering is a statistical technique used for identifying various identical groups of observations using certain characteristics of water. Then the groups are identified so that the attributes of water are closer to the mean of their group than to the mean of any other group.

This K-mean clustering follows the below given steps:

a. Initialize the procedure by separating the similar water parameters into specific clusters, either randomly or using some information about the water attributes.
b. Estimate the Euclidean distance between each member and means of k clusters, and regroup the water attributes to nearest clusters.
c. Recomputed the mean of each cluster and repeat the second step

Repeat the second and third steps until there are no changes in cluster membership.

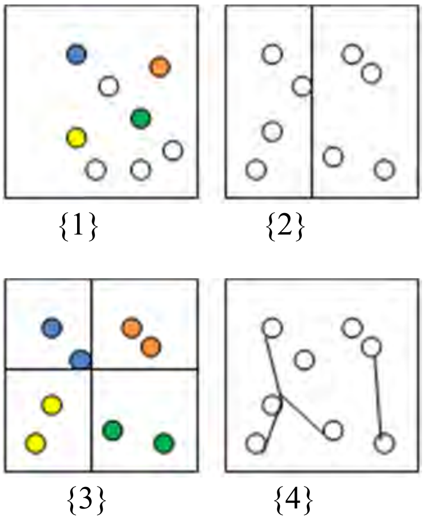

Figure 2: Standard K mean algorithm

1. Here as shown in Figure 2 , {1} represents randomly generated within the data domain
2. Stage {2} represents K clusters are created by associating every observation data with the nearest mean
3. In stage {3}, after re-computing centroid of each K cluster becomes the new mean. In stage {4}, step 2 and 3 are repeated until no longer changes in relationship. [6]

VI. CONCLUSION

This paper represents how Data Mining techniques are efficient to handle large amount of water related data in a systematic way to analyze the water quality. This analysis has given the stepwise process of Data Mining for proposed water quality monitoring system. This paper also summarizes the ongoing researchers involving Data Mining techniques in this area. Basically overall research analysis represents Data Mining Techniques with water quality Indexing and Clustering to study the various water resources to improve the water quality which may help in other environmental related things.

REFERENCES

[1] ShobaG Student, Dept of Computer Science &Eng. , Dr. Shobha G. Prof & HOD, Dept. of Computer Science &Eng., June 2014, “water Quality prediction Using Data Mining Techniques: A survey”, International Journal of Engineering and Computer Science ISSN: 2319-7242 Volume 3 Issue 6, ***Page No. 6299-6306***

[2] Dr. Dave Hargett, HolliHargett, and Steve Springs, Prepared by Pinnacle Consulting Group Division of North Wind, Inc. Submitted to Saluda-Reedy Watershed Consortium 27 July 2005, Submitted to

[3] Saluda-Reedy Watershed Consortium, “water Quality Data-Mining, Data Analysis, and Trends Assessment”, 2005, Research Project ,Page no.6, available at http://www.friendsofthereedyriver.org/files/files/Final%20Report%20data_mining.pdf [Accessed 5th March 2017]

[4] Kamakshaiah.Kolli, R. Seshadri, PhD, August 2013, “Ground water quality assessment using Data Mining techniques “, International Journal of Computer Applications (0975 – 8887) Volume 76– No.15, Department of environment and Climate Change, available at http://www.ecc.gov.nl.ca/waterres/quality/drinkingwater/dwqi.html [Accessed 17th March 2017]

[5] NeetuArora, Amarpreet Singh Arora, Sidhhartha Sharma, Dr.AkepatiS.Reddy, “use of Cluster

Copyright © 2017 [Cyber Times]. All Rights Reserved.

Analysis-A Data Mining tool for improved water quality monitoring of River Satluj" International Journal of Advanced Networking Applications (IJANA), ISSN No. : 0975-0290- no. 63

[6] <https://en.wikipedia.org/wiki/K-means_clustering> [Accessed 15th March 2017]

[7] Muntadher A. SHAREEF*, Abdelmalek TOUMI*, Ali KHENCHAF*, 2014 "Estimation of Water Quality Parameters Using the Regression Model with Fuzzy K-Means Clustering", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 6, 2014

[8] Zheng ZHOU, LiangmingLIU,Yuanling ZHAO, 15 and 20 June 2010, Nessebar, Bulgaria ,"Design of the water quality monitoring system for Inland lake based on remote sensing data" ,3rd International conference on cartography and GIS , 2010, Page no. 8, available at <https://cartography-gis.com/pdf/53_Zhou _China_paper.pdf> [Accessed 13th March 2017]

[9] Rajkumar V. Raikar1, Sneha, M. K, 2012, "Water quality analysis of bhadravathiTaluk Using GIS-a case study", INTERNATIONAL JOURNAL OF ENVIRONMENTAL SCIENCES Volume 2, No 4, ISSN 0976 – 4402.

[10] Twaha A. Basamba, KassimSekabira, C.Marykayombo, Paul ssegawa,2013, "Application of Factor and cluster analysis in the assessment of sources of contaminants in Borehole water in Tanzania ", Pol.J.Environ, stud.Vol.22 no.2 , 337-346

[11] NurgulOzbay, SuheylaYerel, Huseyin Ankara, June 9-10 2009, "Investigation of Cluster Analysis in surface water in Yesilirmak river", 1st International Symposium on Sustainable Development, 2009, Page no 237, available at <http://eprints.ibu.edu.ba /468/1/ISSD2009-SCIENCE-3_p237-p240.pdf>, [Accessed 8th March 2017]

[12] Emad.A.Mohammad Salah, Ahmed M. Turki, EetharM.Al-Othman, December 2012, "Assessment of Water Quality of Euphrates River Using Cluster Analysis", Journal of Environmental Protection, 2012, 3, 1629-1633

[13] Xing-Ping Wen1 and Xiao-Feng Yang, "Monitoring of water quality using Remote Sensing Data Mining", Knowledge Oriented Applications in Data Mining available at <www.intechopen.com>.[Accessed 15th March 2017]

[14] S.Brintha Rajakumari*, Dr.C.Nalin, September 2016, "Data Mining analysis of iron contaminant in river water quality data", research Article www.ijptonline.com, CODEN: IJPTFI, ISSN: 0975-766X, Vol. 8 | Issue No.3 | 18112-18116

[15] Hui Zou 1,2, Zhihong Zou 1, and Xiaojing Wang, "An Enhanced K-Means Algorithm for Water Quality Analysis of The Haihe River in China" International Journal of Environmental Research and Public Health 2015, 12, 14400-14413; DOI:10.3390/ijerph121114400, ISSN 1660-4601

Copyright © 2017 [Cyber Times]. All Rights Reserved.