A COMPARATIVE ANALYSIS FOR SMART WATER RESOURCE USING DATA MINING TOOLS

R.Sasikala [*1]
[*1] Department of Computer Science, National College (Autonomous), Trichirappalli, TamilNadu, India

Abstract

Cities are expecting dramatic population growth and so it will need new and intelligent infrastructure to meet the needs of their citizens and businesses. The water provided by the 5 cities (Chennai, Trichy, Madurai, Coimbatore, Thanjavur) is not sufficient for the use of citizen. In this project we have only considered the water resources available in the different area in trichy. All resources are mapped in the Google map using KML (Keyhole Markup Language) platform .The research also deals with providing a graphical view for the availability of ground water resources of the 5 cities. The groundwater flow model for the study city was formulated by using input data, such as the location of water resources and appropriate boundary conditions. This project needs to collect data from various sources and analysis those data with some datamining tools for predict or decision making process. After collections of various data, main task it to maintain data apply transformation and preprocessing of large data sets for that data mining tools is required. Now a day's various tools for data mining are available either as open-source or commercial software. It includes wide range of software products, from comfortable problem-independent data mining suites, to business centered data warehouses with integrated data mining capabilities and to early research prototypes for newly developed methods. These projects are discussed about various available data mining tools and compare their utilities. We use WEKA, Orange, R Studio, Tinn R, R tools for comparative study about the water resource analysis.

Keywords: Data Mining; Bacteriostatic Water; Weka Tool; Algorithm.

1. Introduction

The exponential growth of electronic data , data storage capacity and powerful computers, leads to develop the machine learning methods for knowledge representations in addition to traditional analysis methods. The huge and large data are someway converted to knowledge or information. Climate affects the human society in all possible ways. Knowledge of available water, water

requirement, weather condition, climate etc. is essential for business, society, agriculture, navigation, transportation, aviation etc. The volume of water resources data in the worlds is increasing day by day and various studies are carried-out on these data for a decision making process. To handle this enormous volume of water data, data mining techniques are required, which predict the results for future action related to weather forecasting, climate change, water management, flood controlling etc. vast diversity of natural resources, water being the most precious of them. Water security, water management and its development is of immense importance in all walks of human life and also for all living beings. Integrated water management is essential for environmental sustenance, sustainable economic development of the country and for bettering human life. In past, India witnessed major human tragedy and property loss due to a heavy rains floods, drought, water crises etc. And it was caused due to lack of knowledge, awareness and absence of efficient and integrated water resource management system. Therefore it is very essential to manage the water for optimum use. Knowledge of available water data is essential for business, society, agriculture, navigation, transportation, environment etc. Presently, the prediction of accurate water availability and water requirement has become most challenging problems around the word since last many years. Predicting the actual availability of water in advance based on predicted incoming volume of water to reservoir will help immensely to take the decision for operation of Reservoir. This will also help for reservoir gate operation to avoid the flood in case if the predicted volume of water is goes beyond the danger level of reservoir or lock the water if predicted volume of water is less than required water level. The volume of water resources data in the worlds is increasing day by day. These data are used for decision making but the traditional techniques are not capable to handle and process the enormous volume of data. Different aspects of water resources are studied and the results combine for holistic outlook

2. Literature Review

2.1. An Internet of Things-Based Model for Smart Water Management

In this paper [1] "An Internet of Things-based model for smart water management" Water is a vital resource for life, and its management is a key issue nowadays. Information and communications technology systems for water control are currently facing interoperability problems due to the lack of support of standardization in monitory and control equipment. This problem affects various processes in water management, such as water consumption, distribution, system identification and equipment maintenance. OPC UA (Object Linking and Embedding for Process Control Unified Architecture) is a platform independent service-oriented architecture for the control of processes in the logistic and manufacturing sectors. We provide an architecture for sub-system interaction and a detailed description of the physical scenario in which we will test our implementation, allowing specific vendor equipment to be manageable and interoperable in the specific context of water management processes.

2.2. Water and Energy Integration: A Comprehensive Literature Review of Non-Isothermal Water Network Synthesis

In this paper[2] "Water and energy integration: A Comprehensive Literature review of non-isothermal water network synthesis" Syntheses of non-isothermal water networks consisting of

water-usages, wastewater treatment, and heat exchanger networks has been recognised as an active research field in Process Systems Engineering. However, only brief overviews of this important field have so far been provided within the literature. This work presents a systematic and comprehensive review of papers published over the last two decades and highlights possible future directions within this field. This review can be useful for researchers and engineers interested in water and energy integration within process water networks using systematic methods based on pinch analysis, mathematical programming, and their combination. We believe that this research field will continue to be active in the near future due to the importance of simultaneous optimising processes, water and energy integration for achieving profitability and sustainability within process industries.

2.3. Water Network Optimization with Wastewater Regeneration Models

In this paper [3] **"Water Network Optimization with Wastewater Regeneration Models"** The conventional water network synthesis approach greatly simplifies wastewater treatment units by using fixed recoveries, creating a gap for their applicability to industrial processes. This work describes a unifying approach combining various technologies capable of removing all the major types of contaminants through the use of more realistic models. The following improvements are made over the typical superstructure-based water network models. First, unit-specific short-cut models are developed in place of the fixed contaminant removal model to describe contaminant mass transfer in wastewater treatment units. Short-cut wastewater treatment cost functions are also incorporated into the model. In addition, uncertainty in mass load of contaminant is considered to account for the range of operating conditions. Furthermore, the superstructure is modified to accommodate realistic potential structures. We present a modified Lagrangean-based decomposition algorithm in order to solve the resulting nonconvex Mixed-integer Nonlinear Programming (MINLP) problem efficiently. Several examples are presented to illustrate the effectiveness and limitations of the algorithm for obtaining the global optimal solutions.

2.4. Review of Optimization Models for Integrated Process Water Networks and their Application to Biofuel Processes

In this paper [4] **"Review of optimization models for integrated process water networks and their application to biofuel processes"** This paper provides an overview of recent development in the area of optimal synthesis of process water networks in which a major goal is to reduce the freshwater consumption by the reuse and recycle of process and treatment streams. The recent models can globally optimize these networks through mixed-integer nonlinear programming techniques. We discuss the application and impact of these techniques to biofuel plants, which are known to consume large amounts of water

2.5. Integrated Water Management Design Criteria Report

In this paper [5] **"Integrated Water Management Design Criteria Report"** The three-fold purpose of this project was to develop a product design criteria methodology to assess water saving products and systems; identify water products and systems that perform well against the design criteria; and comment on any potential commercial opportunities. Different products scored differently against different criteria, with the best overall 'score' coming from a combined

'system', comprising a low flow shower head, a water efficient washing machine, a 9,000 litre rain water tank and greywater reuse. Two potential commercial opportunities were identified; the installation of a relatively small 200 litre rainwater tank attached to the side of the house to supply toilet water only; and a 'modular' tank system where small storage blocks of 200 to 300 litres each could be connected up in irregular shapes to fit under decks, etc.

3. Methodology

3.1. Dataset

A collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer. A data set is organized into some type of data structure. In a database, for example, a data set might contain a collection of business data like calculating PH value, comparing with WEKA tool, Orange, R Studio, Tinn R,R tools. The database itself can be considered a data set, as can bodies of data within it related to a particular type of information, such as sales data for a particular corporate department.

3.2. WEKA

Weka is a landmark system in the history of the data mining and machine learning research communities, because it is the only toolkit that has gained such widespread adoption and survived for an extended period of time. The Weka or woodhen (Gallirallus australis) is an endemic bird of New Zealand. It provides many different algorithms for data mining and machine learning. Weka is open source and freely available. It is also platform-independent
The GUI Chooser consists of four buttons:
- Explorer: An environment for exploring data with WEKA.
- Experimenter: An environment for performing experiments and conducting statistical tests between learning schemes.
- Knowledge Flow: This environment supports essentially the same functions as the Explorer but with a drag-and-drop interface. One advantage is that it supports incremental learning.
- Simple CLI: Provides a simple command-line interface that allows direct execution of WEKA commands for operating systems that do not provide their own command line interface.

3.3. R-Programming Tool

This is written in C and FORTRAN, and allows the data miners to write scripts just like a programming language/platform. Hence, it is used to make statistical and analytical software for data mining. It supports graphical analysis, both linear and nonlinear modeling, classification, clustering and time-based data analysis. It's a free software programming language and software environment for statistical computing and graphics. The R language is widely used among data miners for developing statistical software and data analysis. Ease of use and extensibility has raised R's popularity substantially in recent years. Besides data mining it provides statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others.

3.4. R Studio

This is very popular since it is a readymade, open source, no-coding required software, which gives advanced analytics. Written in Java, it incorporates multifaceted data mining functions such as data preprocessing, visualization, predictive analysis, and can be easily integrated with WEKA and R-tool to directly give models from scripts written in the former two..

3.5. ORANGE

Orange Data mining: Orange is an open source data visualization and analysis tool. Orange is developed at the Bioinformatics Laboratory at the Faculty of Computer and Information Science, University of Ljubljana, Slovenia, along with open source community. Data mining is done through visual programming or Python scripting. The tool has components for machine learning, add-ons for bioinformatics and text mining and it is packed with features for data analytics. Orange is a Python library. Python scripts can run in a terminal window, integrated environments like PyCharm and PythonWin, or shells like iPython. Orange Data mining In Orange, data analysis.

3.6. Tinn R

Tinn R is a software package that enables users to integrate with third party machine learning package written in any programming language, execute classification analyses in parallel across multiple computing node, and produce Html reports of classification results.

4. Result and Discussion

A Water Resource dataset has been used with no of items each for analysis. A set of rules are obtained by Orange tools applying Navie Bayes, KNN and Random Forest Algorithm. By analyzing the data, and giving different support and confidence values, we can obtain different number of rules. During analysis it found that Navie Bayes is much faster for large number of transactions as compare to KNN and Random Forest. It takes less time to generate frequent itemsets. We work on Water Resource data which contains 0.596 transactions. All the results are collected from Pentium Dual core processor with 1. 73GHz speed and 1 - GB RAM

Tools	Chennai Dataset	Coimbatore Dataset	Madurai Dataset	Thanjavur Dataset	Trichy Dataset
Weka tool (Execution time)	0.0	0.0	0.0	0.0	0.0
Orange tool (Execution time)	0.2	0.1	0.1	0.2	0.2
R Tool (Execution time)	0.8	0.4	0.5	0.3	0.4
RStudio (Execution time)	0.6	0.12	0.11	0.9	0.7
Tinn R (Execution time)	0.3	0.5	0.9	0.6	0.8

Table showing comparison of various dataset in different tools
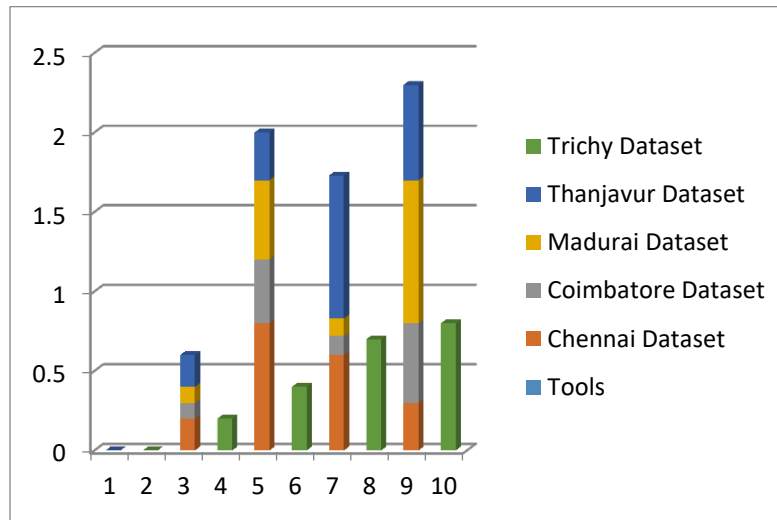
Chart showing comparison of various dataset in different tools

5. Conclusion

In this study the major conclusions derived from the water quality prediction modelling using deep learning approach are outlined as the results carried out from the study concludes that using unsupervised learning, data with variation can be predicted at acceptable water accuracy rate, Results show that WEKA and ORANGE Tool has high variation compared to the other tools. PH has not much variation in data and hence, it is stable as compared to other and has a little variation during summer as the temperature affects the water quality during summer. This system can be implemented on system to continuously monitor the quality of the water. It can be helpful to monitor the quality of water in any uncertain condition. Here we find out the best as Weka tool and orange tool its execution time will be more faster as compare to Tinn R Tool, RStudio and R tools for comparative study about the water resource analysis.

References

[1] T. Robles, R. Alcarria, D. Mart´ın, and A. Morales, "An Internet of Things-based model for smart water management," in Proc. of the 8th International Conference on Advanced Information Networking and Applications Workshops (WAINA'14), Victoria, Canada. IEEE, May 2014, pp. 821–826.
[2] Ahmetović, Elvis, Nidret Ibrić, Zdravko Kravanja, and Ignacio E. Grossmann. 2015. "Water and energy 37 integration: A comprehensive literature review of non-isothermal water network synthesis." Computers & Chemical Engineering 82:144-171. doi: 10.1016/j.compchemeng.2015.06.011.
[3] Yang, Linlin, Raquel Salcedo-Diaz, and Ignacio E. Grossmann. 2014. "Water Network Optimization with Wastewater Regeneration Models." Industrial & Engineering Chemistry Research 53 (45):17680-17695. doi: 10.1021/ie500978h.

[4]   Grossmann, Ignacio E., Mariano Martín, and Linlin Yang. 2014. "Review of optimization models for integrated process water networks and their application to biofuel processes." Current Opinion in Chemical Engineering 5 (0):101-109. doi:http://dx.doi.org/10.1016/j.coche.2014.07.003.

[5]   Yoo, ChangKyoo, TaeYoung Lee, Jiyong Kim, Il Moon, JaeHak Jung, Chonghun Han, Jong-Min Oh, and In-Beum Lee. 2007. "Integrated water resource management through water reuse network design for clean production technology: State of the art." Korean Journal of Chemical Engineering 24 (4):567-576. doi: 10.1007/s11814-007-0004-z.

[6]   "Planning of water quality monitoring systems Technical document UNEP/GEMS Water", 2012.

[7]   "Status of water quality in India -2010, Central pollution control board, Ministry of Environment and forests", 2010.

[8]   P. Tirkey. etal.,, "Water quality indices- important tools for water quality assessment,"IJAC, Vol. 1, pp. 15-28, 2013.

[9]   P. Chaudhry. etal.,, "Water Quality Assessment of Sukhna Lake of Chandigarh City of India; Hydro Nepal," Vol. 12, pp. 26-31, 2013.

[10] S. Malek. etal.,, "Dissolved Oxygen Prediction Using Support Vector Machine," Vol. 8, pp.153-160, 2014.

[11] A. Tarmizi. etal.,, "Dissolved Oxygen Prediction Using Support Vector Machine in Terengganu River," pp.2182- 2188, 2014.

[12] R. Mohammadpour. etal.,, "Prediction of water quality index in constructed wetlands using support vector machine," Vol. 22, pp.6208–6219, 2015.

[13] "Water Resources Engineering" by R.K. Linsley and J.L.H. Paulhus published on ACM Symp. Information, Computer and Comm. Security (ASIACCS), pp. 31-03, 2017.

[14] "Hydroelectric Handbook" by W.P. Creager and J.D. Justin published on 17th ACM Conf. Computer and Comm. Security (CCS), pp. 735-737, 2010.

[15] "Irrigation Water Resources and Water Power Engineering" by P.N. Modi Seventh Int'l Conf. Information Security Practice and Experience (ISPEC), pp. 83-97, 2011.