International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

A Review of Literature on Water Resource Management Using Data Mining Techniques

Pratap Singh Solanki [1], R. S. Thakur [2]

[1]Scientist-B,Central Water & Power Research Station, Khadakwasla R.S.,Pune, India

[2]Associate Professor, Computer Application Department, MANIT, Bhopal, India

Abstract: ***Water, a limited natural resource and most precious elements for all live beings on planet. It is not only essential for human being and plants but also very essential for Agriculture, Environment, Household, Power generation, Industries, Navigation, and Recreation etc. India is an agriculture country and about seventy percent of Indian population depend on the agriculture. Our economy is also based on agriculture. This article aims to review the studies related to use of data mining techniques in the field of water resource sector for Water Management. Presently, the Water Resource Management has become most challenging, interesting and fascinating domain around the word since last many years. Scientist tries to predict the Rainfall, Flood Warning, Water Inflow, Water Availability and Requirements etc. based on huge available metadata using various methods. In this article, we try to search the use of data mining techniques for predicting the inflow, drought possibility, weather report, rainfall, evaporation, temperature, wind speed etc. This paper provides the survey of some literature and work done by the researches using various algorithms and modeling method viz. Associations rules, Classification, Clustering, Decision Tree, and Artificial Neural Network etc.. Data Mining is a collection of techniques for efficient automated discovery of previously unknown, valid, novel, useful and understandable pattern in large databases thus facilitating decision makers to make proactive, knowledge-drive decision..***

Keywords: Data Mining, Water Resource Management, Weather forecasting, Rainfall, Inflow.

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. ***Data mining is the search for relationships and global pattern that exist in large databases but are hidden among the vast amount of data. These relationship represented valuable knowledge about the database and, if the database is a faithful mirror, of the real world registered by the database (Holshemier & Siebes, 1994)*** *[1]***.** 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. [19] India is endowed with a rich and 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. (MoWR Annual Report 2014-15). 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. To handle the different types and enormous volume of data, Data mining techniques are required. Data Mining (DM) is the technology that blends traditional data analysis method with sophisticated algorithms for processing of large volume of data. The Information and Communication Technology have become the integrated part of all sectors and contributing immensely to develop the efficient application for management of various resources due to this development, huge amount of data generated in various field. Data mining technology have the capacity to process the huge amount of data to extract the new and useful information which will help massively for integrated water resource and disaster management.

2. Date Mining

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. This field grew out of the limitation of current data analysis techniques in handling the challenge posed by these new types of data sets. Data Mining does not replace other area of data analysis, but rather takes them as the foundation for much of its work.Data mining is the search for relationships and global pattern that exist in large databases but are hidden among the vast amount of data. These relationship represented valuable knowledge about the database and, if the database is a faithful mirror, of the real world registered by the database (Holshemier & Siebes, 1994) [1]. Data mining is an approach for information extraction from huge amount of data stored in the database (Miller and Han, 2001) [2]. Data mining helps for pattern finding, association clustering and discover the undiscovered information that lie hidden in huge database. Data mining used for predicting the future action related to Retail Industry, Finance Analysis, Banking, weather forecasting, climate change, water management, flood controlling Telecommunication, Biological Data Analysis, Intrusion Detection, Scientific Analysis etc. The potential of using data mining technologies to enhance the decision making process, promptly , accurately and well in advance.

3. Data Mining History

Statisticians have been using computers for years to carried-out the hypothesis experiments on sample data or collected data. During 1970 to 1990 various applications has been developed and after this fuzzy logic analysis has evolved. In this period, the processing capabilities and data storing capacity with multidimensional data model, of the computer system also increased. This statistical analysis are proved a good methodology for proving the relationship among the data attributes but as the volume of database and attributes increased, these method become very time consuming and challenging process.In 19770, Artificial Inelegancy (AI) also shown the capabilities to analyse the data and during 1980s, Artificial Intelligence Algorithm designed to enable machine to learn. This machine learning process became realistic tools to deal with the huge data set. So, Data Mining (DM) is the technology that blends traditional data analysis method with sophisticated algorithms for processing of large volume of data. Therefore the flavour of these Statistics and AI technologies together came into form of DM Technology. In fact, DM is the synthesis of many other technology also viz. Data Management System, Visualization etc. Presently, DM technology are having capacity to handle and process the huge amount of data with the capabilities of classify the data sets, prove the association of attributes with other attributes, partition the data set into similar cluster etc. The process of collection of data, data abstraction, cleaning, use of Data mining tools to find patterns and verification, visualization of the model etc. are known as of Knowledge Discovery in Database (KDD).

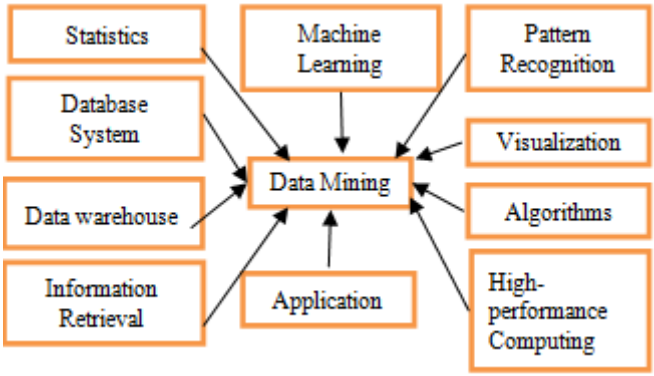

Figure 1.1: Data mining adopts techniques from many domains [18]

4. Water Resource Management Data

The volume of water resources data Viz. Rainfall, Temperature, Humidity, Irrigation, Reservoir data etc. in the worlds have been increasing very fast. These data have been used for only decision making but these stored data have not fully utilised to explore the new and hidden information, which may predict the future possible action. The traditional techniques are not capable to handle and process the enormous volume of data. Data mining technology may help to process the huge amount of data to extract the new and useful information which will help massively for integrated water resource and disaster management. These massive databases offers opportunity to developed new methodology and concept for integrated water management as well as benefits in education also for further enhancing the research based on new findings.

5. Literature Survey

There are many important research studies and experimental work carried-out using various data mining techniques which prove the importance and need of the these technology for water resource management.

Orazio G. et al. [4] Used new data mining technique named Evolutionary Polynomial Regression (EPR) and uses polynomial structure whose exponents are selected by an evolutionary search, thus providing symbolic expressions. He carried-out a case study from UK water distribution system to predict the pipe burst failure rate of water supply system. He identified some parameter viz. pipe age, material and diameter, soil corrosively, meteorological condition, traffic loading, internal pressure etc., but these data are very much difficult to obtain. The same dataset was analysed by Savic et al. (2003) by means of a classification approach [3].

Elia G. P. [5] used Decision tree because decision support tool very often used due to simplicity and easy to understand. They applied classification and regression trees (CART) algorithm to predict the weather. They used the data collected from Hong Kong which recorded between 2002 and 2005, using free data mining specialized software WEKA (Waikato Environment for Knowledge Analysis) to build the decision tree. They prepare the database and used year, month, average pressure, clouds quantity, relative humidity, average precipitation and temperature data. They

transformed these data to use in WEKA which encompasses visualization tools and algorithms for data analysis and predictive modelling with graphical user interface. They predicted with certain accuracy, the average temperature for future months.

M. Kannan et al. [6] used regression method and applied Karl Pearson Correlation Coefficient for finding how many centimetres rainfall fall in the particular region to predict the rainfall fall in the future year using multiple linear regression approach. They acquired the Five years Data of winter season (September, October, November) from Statistical Department of Tamilnadu, Chennai. They discussed about rainfall forecasting methods employed in weather forecasting. Fundamentally, Empirical and Dynamic methods. The empirical method is based on historical rainfall data and its relationship with atmospheric and oceanic variables. E.g. the widely used prediction is regression, artificial neural network, and fuzzy logic. In dynamic approach, physical models are used to generate the prediction and implemented using numerical rainfall forecasting. They applied empirical method for short term forecast of rainfall. The results show that the predicted values are lie below the computed values but as an approximate value.

Pinky S.D. [7] used data mining techniques for Forecasting monthly rainfall of Assam. She used Multiple Linear Regression and processed the six years data collected from Regional Meteorological Center, Guwahati, Assam, (India) .Her model considers temperature, wind speed and mean sea level as predictors. Her model results show the acceptable accuracy. She found 63% accuracy in variation of rainfall in proposed model. Wind direction predictor not included because of some data collection constraints otherwise the results might be more accurate.

Neha Khandelwal et al. [8] presented a MLR equation to predict rainfall using four different climatic factors for Jaipur city, Rajasthan, India, for selecting these factors the author used Pearson correlation coefficient and then use the result to determine the drought possibility.
In his work, she used Jaipur rainfall dataset which is collect from the India Meteorological Department and used average temperature, gas emission, pressure, winds direction, wind speed. After doing whole study of the data they extract useful data for the research area because there are so many factors as we include are greater impact on the climate. So they extract some climatic factors.

Folorunsho Olaiya [9], investigate the use of data mining techniques in forecasting rainfall, maximum temperature, evaporation and wind speed. This work he carried-out using Artificial Neural Network and Decision Tree algorithm. He used meteorological data collected between 2000 and 2009 from the city of Ibadan, Nigeria. In his work he used C5 decision tree classification algorithm to generate decision tree and rules for classifying weather parameter viz. maximum and minimum temperature, rainfall, evaporation, wind speed in terms of month and year. His works show how this parameter has influenced the weather observed in these months over the study period.

Paras et al. [10] developed the model for weather forecasting. They said it is simple because it uses simple mathematical equation. The recorded weather data which is time series data, of a particular station. They predicted maximum and minimum temperature, relative humidity. Relative humidity also predicted using time series of maximum and minimum temperature and rainfall. Model is obtained from MLR equation and the coefficient of these regressions has been used to estimate the future weather conditions. His model shows that the estimated results are satisfactory.

P. Hemalatha. [11] used decision tree classification scheme to predicate the weather report for guidance of ship using Global Positioning System (GPS). GPS helps in identifying the area in which the ship is currently navigating. The weather report on that area is compared with the existing database and decision is made in accordance with the output obtained from the Data Mining Technique. This decision about the weather condition of the navigating path is then instructed to the ship. In her experiments, they placed GPS transmitter in the ship and receiver in the instructing station. Next, they analysed weather report database is the training data and decision tree constructed from these training data set. The GPS transmitter sends the information of the latitude and longitude of its current position to the nearby satellite and the satellite sends this information to the nearest satellite of instructing station. The receiver in the instructing station receives the GPS data and the weather information for that particular location is collected. After that the Decision Tree is traversed using this weather information and the required information is obtained and predicted decision sent to ship and ship navigates as per predicted results.

Badhiye S.S. et al. [12] presented a literature survey on temperature and humidity data analysis for future values prediction using various data mining technique. In his survey, he found that k-Nearest Neighbor classification is an easy to understand and easy to implement classification technique. Despite its simplicity; it can perform well in many situations. k-Nearest Neighbor is particularly well suited for multi-modal classes as well as applications in which an object can have many class labels.

Kalyankar et. al. [13] presented a paper on analysing the meteorological data using data mining technique. They applied knowledge Discovery Process to extract knowledge from Gaza city weather dataset and applied the Clustering data mining techniques.

S. Mohan et al. [14] discussed about the data mining application in water resource. Water resource data has much variability because they represents natural phenomenon therefore this variability makes the modelling as a challenging process. The continues monitoring of water resources started recently and data may generates hourly, daily, monthly, yearly etc. Specifically, as per application of data mining, the database may be classified into two category i.e. surface water hydrology and groundwater hydrology. In Surface Water Hydrology, Rainfall and its associated runoff, evaporation from water bodies and reservoirs, which regulates flow downstream, are the key components of the

surface water hydrology. The prediction accuracy of the first two components ensures the reliability on the derived reservoir operational policy. The predictive data mining techniques viz. Artificial neural network, Navie Bayesian etc. may apply. In groundwater hydrology, increasing water demands due to urbanisation emphasises the necessity of groundwater resource for future sustainability. This situation motivated the researchers in modelling the groundwater to a high level of accuracy by incorporating the dynamics. Groundwater models simulate the fluctuation in water level due to recharge or discharge from the aquifer over a time period and spatial distribution. Continuous monitoring of observation wells encouraged researchers to develop data driven models as an alternate for numerical modelling. To extract the knowledge of this huge groundwater database, data-mining algorithms are applied by researchers. It is observed that even though compared to surface water data, the groundwater database has a longer data history, application of data mining is still an unexplored area and it has huge research potential in groundwater modelling. Artificial neural network and Association Rule Mining (descriptive data modelling) etc. may apply.

Nikhil Sethi et al. [15], presented paper on study of the data mining techniques for predicting the rainfall and its use for farmers and agriculture. He discussed about both empirical and dynamical approaches. He performed empirical statistical experiments to evaluate the accuracy of rainfall prediction using Multiple Linear Regression (MLR) Analysis and used the regional rainfall data taken from Udaipur city, Rajasthan, India. Precipitation, cloud cover, average temperature and vapour pressures used as predictors. There experiments involved data collection, data pre-processing and data selection, reduction of explanatory predictor, building model using regression and validity check. Collection of data is most important part of this complete process. The weather dataset collected from Indian meteorological department which is in excel form on monthly as well as yearly basis. They removed some noisy and unwanted data and filled the missing value as part of data pre-processing. The selected the relevant data from the database for determining the correlation of data. After that they reduced the predictors which have high inter correlation with others because many highly inter correlated explanatory variables may substantially increase the sampling variation of the regression coefficients, and degrade the model predictive ability. After reduction of explanatory predictors, they build model with the use of training data using linear regression technique. The model builds over training data set with test with test data to check the accuracy.

Divya Chohan, et al. [16], conducted a literature survey on Data Mining Technique for Weather Prediction and found that the weather prediction using data mining yields good results and can be consider as an alternative to the traditional meteorological approaches. His study describes the capabilities of various algorithms. Major techniques like decision tree, lazy learning, artificial neural network, and clustering and regression analysis are best suited data mining techniques for this application.

Gettha [17] conducted a literature survey on Data mining techniques in Agriculture. She discussed the various data mining techniques and there benefits for Agriculture Sector using Data Mining and Information Technology. She described how Information Technology and Data mining techniques can help farmers to take the decision regarding better yield. She thoroughly conducted the literature survey and found that the popular data mining techniques viz. K - nearest neighbor, Decision tree, Artificial Neural Networks, Bayesion network, Support Vector Machine , Fuzzy set, K – means etc. have been used in agriculture sector. She also discussed on various data mining application in which proved very helpful for solving the different agricultural problems.

Omkar [20] conducted experiment for predicting the inflow at Hatnur Reservoir in Maharashtra located at Tapi River using regression Analysis. They developed their own modeling software and process the ten years rainfall and discharge data to get the regression coefficient. The calculated regression coefficient and observed discharge used to predict the inflow which is expected to come at Hatnur Reservoir.

6. Objective for Literature Review

From the literature review it is envisaged that the research will encompass the following objectives.

i) Use of the advance technology and existing enormous water recourse available data to predict the water availability and needs, in advanced.
ii) Development, promotion and use of water resource data and decision support tools.
iii) Help set the standard for efficient and need based water management system.
iv) Help to avoid the flood and drought situation (Disaster Management).
v) Help to store and provide the water for urbanization, Navigation, Fishing, Irrigation, Industries etc. as per future needs.
vi) Advances in Computing Technologies will be made online real time or quasi-real time water management modelling possible.
vii) Help to control the global warming.

7. Conclusion

In this literature review we found that, the thinking offered by the advent of computer technology highly complementary to some of the goals of water management. Services delivered by the technology are interactive, fast and multidimensional. In this way Data mining offer us a much-needed opportunity to deliver scientific findings and information to stakeholders and decision makers for providing collective decision-making tools. An integrated water management scale simulation model may be built and it may serve as a core for water management design to provide a conceptual basis for understanding the performance of the water management system. It is not the amount and quality of information that is crucial for the success of water management, but how well the information is disseminated, shared and used by the decision makers. In this respect the Data Mining Technology offers a wealth of

opportunities for the decision-making process. Data mining is an advance and powerful technology which helps in extracting hidden predictive information from huge databases and make capable to decision makers to make timely, knowledge-driven decisions. In this literature review we tried to review the various popular data mining techniques for weather prediction and climate change using various approaches and algorithm based on the important research work carried-out using different techniques. The Karl Pearson Correlation Coefficient, Multiple linear regression approach, C5 algorithms are used to generate the decision tree and rules for classifying the weather parameters. Decision Tree is easy to understand, interpret and works faster. Artificial Neural Network can be also used to detect the relationship between the input parameter and generate results based on observed patterns. Decision tree, lazy learning, artificial neural network, and clustering and regression analysis are suited data mining techniques.

References

[1] Holsheiler, M., and Siebes, A. (1994) Data mining: *The search for knowledge in databases,* Technical Report CS-R9406, Centrum voor Wiskunde en Informatica (from Web).
[2] Miller, H.J and Han, J. (2001) Geographic Data Mining and Knowledge Discovery, Taylor and Francis (From Web).
[3] Savic D.A., Bessler F and Walters G.A., (2003) Comparison of Data Mining Methods to Statistical Approaches For Pipe Burst Risk Analysis, IWA-IAHR joint conference: Pumps, Electromechanical Devices and Systems PEDS.
[4] Orazio G,Dragan A.S.,Daniele L. (2004) Data Mining for Management and Rehabilitation of Water Systems: The Evolutionary Polynomial Regression Approach. Wasserbauliche Mitteilungen Heft 27. Institutfür Wasserbau und THM der TU Dresden ; 285-296.
[5] Elia Georgiana Petre (2009) A Decision Tree for Weather Prediction. BULETINULUniversităŃii Petrol – Gaze din Ploiesti Vol. LXI No. 1/2009; 77-82.
[6] M. Kannan, S. Prabhakaran and P. Ramachandran (2010) Rainfall forecasting using data mining technique. International Journal of Engineering and Technology Vol.2 (6); 397-401.
[7] Pinky Saikiya Datta (2010) Prediction of rainfall using data mining technique over Assam. Indian Journal of Computer Science and Engineering Vol. 5 (2), 2014; 85-90.
[8] Neha Khandelwal and Ruchi Davey (2012) Climatic Assessment of Rajasthan's Region for Drought with Concern of Data Mining Techniques. International Journal of Engineering Research and Applications (IJERA) Vol.2 (5); 1695-1697.
[9] Folorunsho Olaiya (2012) Application of Data Mining Techniques in Weather Prediction and Climate Change Studies. I. J. Information Engineering and Electronic Business; 51-59.
[10] Paras, Sanjay Mathur (2012) A Simple Weather Forecasting Model Using Mathematical Regression. Indian Research Journal of Extension Education Special Issue Vol.(I), 2012; 161-168.
[11] P.Hemalatha (2013) Implementation of Data Mining Technique for Weather Report Guidance for Ships Using Global Positioning System. International Journal of Computational Engineering Research Vol. 3 (3); 198-202.
[12] S.S. Badhiye, Chatur P.N. and Wakode B.V. (2013) Temperature and Humidity Data Analysis for Future Values Prediction using Clustering Technique: An Approach. International Journal of Computational Engineering Research Vol. 3 (3) ; 198-202.
[13] M. A. Kalyankar, S.J. Alspurkar (2013) Data Mining Technique to Analyze the Metrological Data. International Journal of Advance Research in Computer Science and Software Engineering Vol 3 (2); 114-118.
[14] S. Mohan , N. Ramsundram (2013) Data-mining models for water resource applications. ISH Journal of Hydraulic Engineering. Vol.19 (3); 211–218.
[15] N. Sethi, Kanwal Garg (2014) Exploiting Data Mining Techniques for Rainfall Prediction. International Journal of Computer Science and Information Technology Vol. 5 (3), 2014; 3982-3984.
[16] D. Chouhan, Jawahar Thakur (2014) Data Mining Technique for Weather Prediction : A Review. International Journal of Recent and Innovation Trends in Computing and Communication Vol. 2 (8); 2184-2189.
[17] Geetha M.S.C. (2015) A Survey in Data Mining Techniques in Agricultures. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 3 (2); 887-892.
[18] Jiawei Han,Micheline Kamber, Jian Pei (2014). Data Mining Concept and Techniques (3rd edition). Elsevier;pp 23,327-328;330-332;350-351;355-356.
[19] Annual Report of Ministry of Water Resources, Government of India, 2015. Web Link address http://wrmin.nic.in/writereaddata/Annual_Report_2014-15.pdf; p 1.
[20] Omkar Potekar,Siddharth Roman,Swapnil Dhumal,Akshay Dhadave, Pratap Singh Solanki (2016) Implementation of DSS using Regression Analysis for Dam Management, International Journal of Research in Engineering and Technology, Vol 05 (05), 2016;199-206.

Author Profile

Pratap Singh Solanki received his Master of Computer Application (MCA) degrees from Devi Ahilya Univarsity, Indore (Madhyapardesh) in 1999. He is working in Central Water & Power Research Station (CW&PRS, Govt. of India, MoWR, RD & GR), Khdakawasla R.S., Pune, India since October 2002 and presently working as Scientist-B. He has been working in the field of Software Development, Database Management, Networking, Web Technology, and Water Resource Sector. This review of literature work has been carried-out for the fulfillment of Ph.D. course work under the guidance of Dr. R.S. Thakur, Associate Professor, MANIT, Bhoal (M.P.).