ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.2, Pages : 04 - 07 (2014)
Special Issue of ICACE 2014 - Held on March 10, 2014 in Hotel Sandesh The Prince, Mysore, India

Parametric Analysis of Water Resource Data (E-Governance Projects) Using Data Mining Techniques

Himanshu Trivedi, Dr. Amit Dutta
AISECT University Bhopal, India, Himanshu61@gmail.com

Abstract— E-Governance refers to Electronic Government/Easy Government.

E-Governance is a cognitive process in providing citizen-centric information to the public or an organization. In today's revolutionary world of technologies, industries (public & private) are growing everyday with innovative ideas and IT (Information Technology) support. To manage shoulder-to-shoulder support with such growing companies, the government needs an extra edge to introduce IT technologies in their policies. Thus, E-Governance concept is introduced to give them opportunities with extensive options to introduce IT based services to various sections in government. Beside industries, citizens will get primarily benefits with the introduction of E-Governance. Their daily chaos with government agencies will get more eased now with various innovative IT projects. Rise in scope of E-Government towards citizen-centric services has been one of the most prominent developments in government sectors worldwide. Globally, E-Governance has setup a benchmark in various simplification of their work deliverance to citizens and industries.

In this paper **we propose E-governance implementation for Water resource department** with the use of data mining techniques. For this we intend to frame a logical architecture design for WRD. On the basis of this architecture we will develop the logical design of database. We aim at demonstrating the need of hydrological information need of different hydrological data users and techniques of using data mining in e-governance applications, with special reference to water resource department Madhya Pradesh.

Keywords:- e- Governance, Data mining, water resource department.

Introduction

Data mining is a technique of discovering analytical data from large amount of data stored in data warehouse, relational database and other data repositories. Due to large amount of data in digital forms and need for turning such data into analytical information and knowledge for broad application and decision support system. Data mining refers "*Knowledge discovery in database*"

Data mining is a dominant new technology with great potential to help users focus on the most important information in their data base. Data mining tools predict future trends and behaviors, allowing planning to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems.
Data mining consist of sequence of following steps:

Data integration
Data Cleaning
Data transformation
Data Mining
Data Analysis

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries.

Data mining is a powerful tool that can help us find patterns and relationships within our data. But data mining does not work by itself. It does not eliminate the need to know our BI, to understand our data, or to understand analytical methods. Data mining discovers hidden information in our data, but it cannot tell us the value of the information to our organization.

E-governance & Data mining

Data mining is a broad category of applications and technologies for gathering, storing, analyzing and providing access to data to help the decision makers in making decisions. Typically, any Mining (DM) application includes large data base, decision support systems, query and reporting, OLAP (On Line Analytical Process), statistical analysis & mining DM, therefore is well suited for e-Governance applications in the G2G (Government to Government) and G2C (Government to Citizen) environment.

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.2, Pages : 04 - 07 (2014)
Special Issue of ICACE 2014 - Held on March 10, 2014 in Hotel Sandesh The Prince, Mysore, India

Data Mining has been successfully deployed in numerous application areas (Brachman et. al 1996). Within the government the focus of data mining is predominantly on the management of interaction between the government and citizen or business. In these application data mining can extract knowledge from transactional data, can lower the burden of dealing with large volume of transactions and can improve decision making abilities.

Need of Data Mining in E-Governance

Many departments associated with water resource department for hydrological data system like Fisheries, agriculture, PHE, pollution control board, NABARD, Farmers associations, local water bodies, irrigation, funding bodies and the government in other words; the stakeholders would be interested in to know "What are the timely & quality of services offered in water resource system.
The stakeholders raise these questions with one or more interest: Improving the services in water resource system, Satisfaction for Hydrological data users group (HDUG's), providing data information timely & accurately. Government accountability and policy making; funding agencies for funding decisions. To find an answer to the questions on quality of services in water resource system, these stakeholders depend on various sources of data.
One area of interest for regulating authorities is to monitor the uses and availability of water as per needs. The above new technologies enable interactive data analysis and adhoc reporting (INMON 1996). The data mining can primarily "*Knowledge discovery in database*". Every water data users have different need of water data also the period of the information can vary from each other. In this context the difference between time & type of request for data and availability of data in departmental database is important for planning & proper implementation of government schemes as well as disaster management system. With the help of e-governance application, the process can be made automated or time bound. All details needed for the accumulation & predication can be fetched by a data mining techniques. A data mining can help to provide a fair availability of data.

The Data Mining Cycle

The four stages of the righteous cycle of data mining are:

1. **Categorize HDUG's troubles issues**: where the aim is to classify areas where patterns in data have the possible of providing value.

2. **Techniques to renovate difficulty into Information**: for this function, the created results need to be tacit in order to make the righteous cycle successful. Several pitfalls can obstruct with the ability to use the results of data mining. Some of the pitfalls are bad data formats, confusing data fields, and lack of functionality. In addition, identifying the right source of data is crucial to the results of the analysis, as well as

bringing the right data together on the computing system used for analysis.

3. **Performing of the information**: where the results from data mining are acted upon then fed into the measurement stage.

4. **Evaluate the outcome:** this measurement provides the feedback for continuously improving results. These measurements make the righteous cycle of data mining *righteous.* Even though the value of measurement and continuous improvement is widely acknowledged, it is usually given less

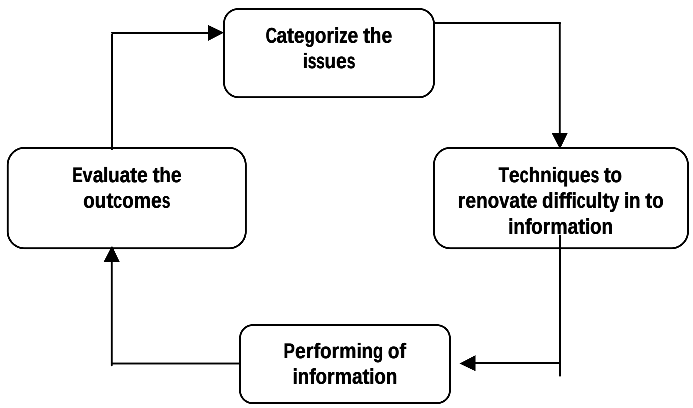


Fig 1. **Data Mining Lifecycle**

Data Mining Techniques in Water Resource Department

There are different data mining techniques available:

Decision Tree

Decision trees are an approach of representing a sequence of rules that lead to a set or value. As a result, they are used for directed data mining, mainly classification. One of the main rewards of decision trees is that the model is quite reasonable since it takes the form of explicit rules. This allows the evaluation of results and the identification of key attributes in the process. It consisting of nodes and branches organized in the form of a tree such that, every interior non-leaf node is labeled with ideals of the attributes. The branches coming out from an inner node are labeled with ideals of the attributes in that node. Each node is labeled with a rank (a worth of the goal characteristic). Tree based models which include classification and regression trees, are the common implementation of induction modeling. The decision tree can be built from the very small data set (Table 1). In this table each row corresponds to an enduring record. We will refer to a row as a data instance. The data set contains four predictor attributes, HDUG, work area, type of data, quality parameters

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.2, Pages : 04 - 07 (2014)
Special Issue of ICACE 2014 - Held on March 10, 2014 in Hotel Sandesh The Prince, Mysore, India

Table 1: Data set used to build decision tree of

Key HDUG's	Area of work	Type of data	Water quality parameter
Fisheries	Fish culture	Metrological/sediment river flow data	Hardness of water, Water chemistry level
Agriculture	Crop productio n	Metrological/sediment river flow data	W.Q for agriculture yield.
PHE	Drinking water	Metrological/ Ground water data	Hardness of water, Water chemistry level, Biological parameters
Irrigation	Irrigation water	Historical data, Daily rainfall data	Hardness of water, Water chemistry level
PCB	Quality control	Rainfall/ climatic data	Maximum WQ parameters
Hydropowe r	Power generation	Metrological data, Historical data	Water chemistry level

Decision tree can be used to classify a need of water data of the above data set given in the Table 1.

Classification techniques in data need: The objective of the classification is to assign a class to find previously unseen records as accurately as possible. If there is a collection of records (called as training set). The motive is to find a classification model for class attributes, where a test set is used to determine the accuracy of the mode. The known figures set are separated into guidance and test sets. The training set used to build the model and test set is used to validate it]. Classification process consists of training set that are analyzed by a classification algorithms and the classifier or learner. Model is represented in the structure of classification rules. Test data are used in the classification rules to estimate the accuracy. The beginner model is represented in the form of classification rules, decision trees or mathematical formulae. The result briefly indicates the scenario of the

1. Data requirement where 91 % HDUG's required current/ present data, 66 % HDUG's collectively require real time & time interval data and there are only 42.30 % HDUG's who require periodic data/ time span data.

2. AS far as **Hydrological data** need is concern there is 75 % HDUG's require meteorological data and water quality data where only 25 % users require sediment data.

3. **Meteorological data** where 83.34 % HDUG's required rainfall data and 74 % users require climatic data.

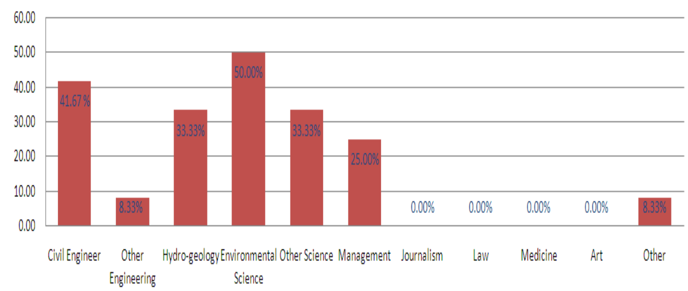
4. As far as **River flow data need** is concern there is 66.67 % users require Annual runoff data, 58.34 % users require monthly runoff data and only 33 % HDUG's require River water level data.

5. **Ground water level** where 68.75 % users require seasonal GW level, 33.67 % users require monthly GW level and only 16.76 % HDUG's require daily GW level data.

6. As far as **Type of data** is concern there is 85 % users require Hydro-metrological data, 71.75 % users require Surface data, 68 % HDUG's require Ground water data & 92 % users require water quality data as shown in table below:

S.No.	Data Type	Data requirement as per users (%)
1	Data requirement	Current/present data: 91% Real time data: 66 % Time interval data: 42.30
2	Hydrological data	Metrological data: 75 % sediment data: 25%
3	Meteorological data	Rainfall data: 83.34% Climatic data: 74 %
4	River Flow data	Annual runoff data: 66.67 % Monthly runoff data: 58.34 % River water level: 33 %
5	Ground water level	Seasonal GW level: 68.75 % Monthly GW level 33.67 % Daily GW level: 16.76 %
6	Category of data	Hydro-metrological data: 85 % surface water data: 71.75 % Ground water data: 68 % Water quality data: 92%

Table 3 : HDUG wise Data require (%)



Discipline of water data users

Prediction (Require WQ parameters for fish production)

WQ Parameter	Range	Result (if below)
Dissolved Oxygen	0 to 1.5 mg	lethal especially if exposed for long periods
	1.4 – 5 mg	fish survive, but reduced feed intake higher FCRs, slow growth, stress, and increased susceptibility to disease results
Temperatur e	**Below 15C**	Growth stops and death occurs at extremes
	15 to 26 C	Reduced feed intake and growth rates. Higher FCRs. Fish more stressed at lower

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.2, Pages : 04 - 07 (2014)
Special Issue of ICACE 2014 - Held on March 10, 2014 in Hotel Sandesh The Prince, Mysore, India

		temperatures, therefore, more susceptible to disease
pH	Below 4	acid death point
	4 – 6.	Survive but stressed, slow growth, reduced feed intake, higher FCRs
Alkalinity and Hardness	Extreme fluctuations in pond pH levels during the day which is stressful to Fish are under physiological Stress the fish.	
Ammonia Nitrogen	Not more 0.3 – 2 mg	Fish are happiest when there is no or little ammonia in water The fish succumb more to attacks by trematodes and other parasites

Issues & challenges

The requirement of water for various uses and the exact quality of water along with all the minerals & other essential component is considered as a significant yet intricate task that needs to be carried out precisely & efficiently. The automation of the same would be highly beneficial. Decisions are often made based on scientist, intuition and laboratory experts, experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive information/test cost which affect the quality of services provided to HDUG's. Data mining have the potential to generate a knowledge environment which can help to significantly improve the quality of decisions.

RESULT & DECISION

With the current rapid increase in the amount of HDUG's data being collected electronically in water uses and the widespread availability of cheap and reliable computing equipment, many researchers have already started, or are eager to start, exploring these data. The outcome obtained by data mining, in particular from the subfield of machine learning, may not only be oppressed to recover the worth of water uses by implement particular change to water uses policies but can also be used as a basis for the structure of computer-based decision support systems. We present a case study of application of data mining and analyze data base of water resources for different HDUG's. The concept of Classification method has been applied in the study of various sources of water. Meteorological data & water quality has emerged as an opportune need for data mining technology for a number of factors. Water resource administers would like to know how to improve outcomes as much as possible.

After preliminary results were analyzed, the program projected that per year yield of crop/ fish/ can be improved through this information as well as disease and other water generated problems in human & animal can be prevented effectively. These information's will also help local bodies & other public health related departments to make effective and easy planning for availability of drinking water. There is still, however, much that can be done. Through the use of data mining algorithms it was possible to verify the improvement of quality. Future work includes the Collecting information about levels of water availability & uses. The proposed work can be further enhanced and expended for the automation of water availability & uses with details. Real data from water resource department needs to collected and all the available techniques will compare for the optimum accuracy.

REFERENCES

Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI Press/The MIT Press, 1996.

Ester M., Wittmann R.: "Incremental Generalization for Mining in a Data Warehousing Environment", Proc. 6th Int. Conf. on Extending Database Technology, Valencia, Spain, 1998, in: Lecture Notes in Computer Science, Vol. 1377, Springer, 1998, pp. 135-152.

Fayyad U., Piatetsky-Shapiro G., and Smyth P.: "Knowledge Discovery and Data Mining: Towards a Unifying Framework", Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, Portland, OR, 1996, pp. 82-88.

_ http://www.engpaper.com/data-mining-research-papers-2012-107.htm

_ Role of data mining & dat warehousing in E-governance

Dr. Kishori Lal Bansal, Satish Sood

o K. Jain and R. C. Dubes, "Algorithms for Clustering Data", Prentice Hall, 1988.

NISG, PMI, Grant Thornton India (2011), "Project Management in E-Governance: Issues & Challenges in navigating to the New Normal", available at www.nisg.org/docs/539_Report.pdf

Mission Mode Projects", Ministry of Communication & Information Technology, Government of India, available at http://www.mit.gov.in/content/mission-mode-projects

_E-Governance in India indiaegovernance.blogspot.com/

_ National e-Governance Plan - Government: National Portal of India india.gov.in › Government

_Department of water resource Madhya Pradesh: portal mp.gov.in/wrd.