ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 4, Issue 11, November 2017

Brief Survey of Different Techniques for Prediction of Groundwater Level and Groundwater Quality

[1] Ajinkya P. Chatur, [2] Prof. R. V. Mante,[3] Amol R. Dhakne
[1] Student, Department of CSE, Government College of Engineering, Amravati
[2] Assistant Professor, Department of CSE, Amravati
[3] Assistant Professor, Department of Computer Engineering, DYPIEMR, Akurdi, Pune

Abstract— **Data analysis is a stream which uses concepts from statistics and computer science to clean the data and then to find some useful information out of it. In order, to access useful information, various algorithms of machine learning can be used. Such algorithms can further be used for prediction of future data to some extent. So this paper is focused on survey of different methods and algorithms to analyse the patterns of groundwater level. After the analysis of their methods and their accuracies, it is concluded that CBA method can best describe the trends in groundwater and various factors affecting it.**

Keywords: **CBA, NB, Naïve bayes, KNN, FNN**

INTRODUCTION

Water is considered as one of the most important elements, not only for human life, but also for all types of life on the planet. The provision of water represents one of the main problems in many countries. Jordan is one of these countries and is recognized as one of the water-poorest countries in the world. According to [1], “Jordan suffers from water scarcity, which poses a threat that would affect all sectors that depend on the availability of water for the sustainability of their activities for their development and prosperity”. Moreover, Jordanian citizens only have access to around 1,000 cubic meter per year, while American citizen use more than 9,000 cubic per year [2]. According to [1], groundwater is the main water resource in Jordan. Moreover, the groundwater resource is the only water supply available in many parts of the country. In Jordan, there are 12 main basins, which are comprised of a set of aquifers. The most important aquifers are the Amman/Wadi Elsir, Basal and Ram. Countries like Jordan need to exploit every available water resource in order to provide water for their citizens. One of these methods is to use data mining techniques to predict and find groundwater. Many methods have been developed to predict groundwater areas. For example, [3] proposed a method to check if the ground water from selected areas are potable or not. As multivariate is present, Principal Component Analysis with data mining techniques using JRIP rules were employed for classifying the ground water. The authors conclude that data mining techniques can be employed. For quicker classification of water potability. Reference [4] presents two models to predict groundwater levels in an unconfined shallow aquifer in the Searsville basin, part of the Jasper Ridge Biological Preserve. Linear regression does a good job of predicting groundwater levels in the summer, when water levels are low, while the neural network does a good job of predicting groundwater levels in winter, when water levels are high. This result supports the combination of linear regression and neural networks for predicting hydrologic response up to one year in advance. Moreover, [5][6] developed different methods with varying techniques that aim to detect groundwater areas.According to the abovementioned paragraphs, it is important to discover new areas of groundwater with minimum resources and cost. Four well-known data mining techniques have been used to predict groundwater areas in Jordan. In order to discover new groundwater sites, Jordanian groundwater experts identified seven features that used as input parameters for the four data mining techniques. The groundwater features are: Elevation, Valleys, Slope of the earth's surface, the annual long-term average of rainfall, the annual long-term average of temperatures, Geological outcrop and the faults of earth surface.

2. DIFFERENT METHODS FOR PREDICTION OF GROUND WATER LEVEL

This section discuses four well-known data mining techniques: CBA, SVMs, NB and KNN. CBA is the first approach that integrates the association rule task with the classification data mining task. Also, SVMs is considered as one of the most accurate data mining algorithms that performs classification by building an N-dimensional hyperplane that optimally splits the data into two classes. Moreover, NB is a simple probabilistic algorithm based on Baye's theorem. Finally, KNN is a statistical data mining algorithm, which has been intensively studied in pattern recognition over five decades. The next four subsections discuss the four algorithms that we consider

2.1 CBA

Reference [7] considered CBA as one of first method works that presented the utilization of association rule in classification. The CBA algorithm works through three steps: rule generation, model building and prediction. In the first step and according to [8], the algorithm employs the Apriori algorithm to find the frequent rules among training data features. Any rule is called a frequent rule if it has support greater than or equal to the inputted minimum support. For the second step, the frequent rules are sorted according to certain criteria. Then, some of the frequent rules are pruned because these rules may be redundant, conflict, or noise. The remaining rules are sorted according to confidence and support, and inserted into CBA model, this step is called, model building. Finally, to predict a new test instance with the suitable class, the first rule in the set of ranked rules that matches the test instance predicts it to the class label of the matched rule. According to [7]; [9], if no rule matches to this instance, it assigns the default class.

2.2 SVMs

SVMs developed by [10], are a supervised data mining algorithm which can be used for either classification or regression challenges. However, it is frequently used in classification problems. SVMs have become the method of choice to solve difficult classification problems in a wide range of application domains. SVMs built on the fundamental of minimization of structural risk. In linear classification, SVMs produce a hyper plane that splits the training data into two groups with maximum-margin. A maximum-margin hyper plane is a hyper plane which separates two of points and is at equal distance from the two. Mathematically, SVMs learn the sign function f (x) = sng (wx + b), where w is a weighted vector in Rn. SVMs find the hyper plane y = wx + b by separating the space Rn into two half-spaces with the maximum-margin. Linear SVMs can be generalized for nonlinear problems. To do so, the data is mapped into another space H and we perform the linear SVMs algorithm over this new space.

2.3 NB

The NB is a simple and effective algorithm that utilizes the likelihoods of each feature belonging to each class value to do a prediction step [11]. NB streamlines the computation of likelihoods by supposing that the likelihood of each feature belonging to a given class value is independent of all other features. The likelihood of a class value given a value of a feature is named the conditional likelihood. By multiplying the conditional likelihoods together for each attribute for a given class value, we have a likelihood of a data object belonging to that class. To do a prediction we can compute the likelihoods of the object belonging to each class and select the class value with the highest likelihood.

2.4 kNN

The KNN algorithm [11] is simple. Based on training and test data, the KNN finds the kNNs of the training data, and uses classes of the k-neighbors to assign the class of the test instance. The scores of similarity of each neighbor instance to the test instance are used as a weight of the classes of the neighbor instance. When several kNNs share a class, then the pre-neighbor weights of that class should be added together, and the result of the added weights should be used as the likelihood score of that class with regard to the test instance. In order to find a ranked list for the test instance, the scores of the candidates' classes should be sorted.

3. SOME MORE TECHNIQUES FOR PREDICTION OF GROUND WATER LEVEL AND QUALITY:

3.1 ANN (Artificial Neural Network)

ANN is a classification model which is grouped by interconnected nodes. It can be viewed as a circular node which is represented as an artificial neuron that reveals the output of one neuron to the input of another. The ANN model is helpful in revealing the unexposed interrelationships in the classical information, therefore expiditing the prediction idea, envision and forecasting of water quality. Based on their performance metrics, we use various formulas which have been illustrated in [12].ANN model is definite and systematic enough to make important and relevant decisions regarding data usage.

3.2 Naïve Bayes

Naïve Bayes is a classification technique which is based on probability theories which entirely demonstrate the characteristics of water quality assessment[13]. Bayes model is easy to use for very large datasets. In other terms, a Naive Bayes assumed that the value of a distinct feature does not related to the presence or absence of any other feature, given in the class variable. It undergoes through following steps: 1. Extract, clean and classify the water quality. 2. Remove large punctuations and split them. 3. Counting Tokens and calculating the probability. This probability is called as posterior probability which is calculated by the formula described in [14]. 4. Adding the probabilities and then wrapping up.

3.3 Decision Tree

Decision tree is one of the predictive modeling technique used in data mining. It aids to divide the larger dataset into smaller dataset indicating a parent-child relationship. Each internal node defined as inner node is labeled with an input feature. The inner nodes which exhibit many types of attribute test, bifurcations exhibit the test outcomes and leaf nodes particularly exhibit the category of a specific type[15]. Decision tree can handle both numerical and categorical data. It is well suited with large datasets. Higher accuracy in decision tree classification technique depicts that the technique can simulate. It is able to optimize variety of input data such as nominal, numeric and textual. It is a successful supervised learning approach which has the capability of extracting the information from vast amount of data based on decision rules.

3.4 FNN (Fuzzy Neural Network)

ANN is basically associated with the neurons having the capability of storage and processing for the information. FNN is chosen as a algorithm for data mining introducing the artificial neural networks. It describes the integration of fuzzy Fuzzy logic with the neural network. Fuzzy neural network algorithm which deals with the prediction process is composed of five layers named as: input layer, hidden layer, fuzzification layer, fuzzy reasoning layer and reconciliation fuzzy layer. Zhu has explained the structure of FNN [16]:

- The input layer is the main index that affects the teaching quality as it is the input as well as feedback of fuzzy neural network.

- The fuzzification layer is the activity of computing membership function value from input parameters that belongs to fuzzy set.

- The fuzzy reasoning layer is the fundamental part of fuzzy neural network that is basically employed for simulating the operation of fuzzy relational mapping.

- The reconciliation fuzzy layer is the output layer which means "the distribution of value" that is depicted as by "certainty value ".

The fuzzy rules as well as membership functions which can be stated by using neural network (import) and neural network that is generated to maintain the use as fuzzy inference[16]. Eventually, fuzzy rules and membership functions are usually extricated from the neural network (Export) in association compared to that it would be helpful to describe the actual neural network's central rendering and also the operation.

3.5 Back Propogation Neural Network(BPNN)

ANN consists of interconnected processing units. Each unit is known as neuron. Each neuron will receive an input from another neuron. Weights are assigned to each neuron. These kinds of weights regulate the nature as well as strength and power of the significance involving the interlocked neurons. The respective signals named as indicators tend to be refined from each and every input and then further processed via a weighted sum to the inputs. The BPNN algorithm criteria looks for the error with the method called as steepest descent. The united weights are modified by simply moving on the way to the negative gradient of the energy function by providing emphasis at each and every iteration for evaluating the network performance. Various performance metrics are used for calculating the network error based on specific formulas. This algorithm follows four major steps:

1. Feed forward computation.
2. Applying back propogation at the output layer.
3. Applying back propogation to the hidden layer.
4. Weights updation.

This algorithm will continue its processing until the value of error function becomes too small.

4. CONCLUSION

This paper gives brief survey on various methods of predicting the ground water level such as such CBA, SVM, NB and kNN along side with Artificial Neural Networks, Naive Bayes, Decision Tree, FNN and BPNN. These methods helps to access groundwater level and quality of groundwater. Researchers can propose new more effective methods to predict groundwater level and quality accurately by considering all these methods. Also, in all of these methods, it has been found that CBA outperforms other methods in terms of their performances.

REFERENCES

[1] Jordan Ministry of Water and Irrigation – Reports 2013-2016. http://www.mwi.gov.jo/sites/enus/SitePages/MWI%20BGR/Reports.asp x

[2] Nortcliff A, Carr G, Potter RB, Darmame K. (2008) Jordan's Water Resources: Challenges for the Future. Geographical Paper No. 185, The University of Reading.

[3] Karthik, D., & Vijayarekha, K. (2014). Multivariate Data Mining Techniques for Assessing Water Potability. Rasayan Journal of Chemistry, 7 (3):256-259.

[4] Maatta, S. (2011). Predicting groundwater levels using linear regression and neural networks, CS229 final project, December 15, 2011.

[5] Al Kuisi, M., El-Naqa, A., & Hammouri, N. (2006). Vulnerability mapping of shallow groundwater aquifer using SINTACS model in the Jordan Valley area, Jordan. Environmental Geology, 50(5), 651-667.

[6] Karthik, D., Vijayarekha, K. & Abirami S. (2015). Classifying ground water quality using data mining technique for Thanjavur district, Tamilnadu, India. Journal of Chemical and Pharmaceutical Research, 7(3):1724-1727.

[7] Liu B., Hsu W. and Ma Y. (1998). Integrating classification and association rule mining. Proceedings of the KDD, (pp. 80-86). New York, NY.

[8] Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215, pp. 487-499).

[9] Antonie M. and Zaiane O. (2002). Text Document Categorization by Term Association, Proceedings of the IEEE International Conference on Data Mining (ICDM '2002), (pp.19-26), Maebashi City, Japan.

[10] Vapnik V. (1995). The Nature of Statistical Learning Theory, chapter 5. Springer-Verlag, New York.

[11] Hadi, W., Thabtah, F., ALHawari, S., & Ababneh, J. (2008). Naive Bayesian and k-nearest neighbour to categorize Arabic text data. In Proceedings of the European Simulation and Modelling Conference. Le Havre, France (pp. 196-200).

[12] S. Palani, S. Liong, P. Tkalich, "An ANN application for water quality forecasting", Marine Pollution Bulletin 56 (2008) 1586–1597.

[13] L. Chaunqi, W. 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 (ICISE2009).

[14] J. Lu, T. Huang, “Data Mining on Forecast Raw Water Quality from Online Monitoring Station Based on Decision-making Tree”, 2009 Fifth International Joint Conference on INC, IMS and IDC.

[15] Z. Xing, Q. Fu, D. Liu, “Water Quality Evaluation by the Fuzzy Comprehensive Evaluation based on EW Method”, 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD).

[16] C. Zhu, Z. Hao, “Fuzzy Neural Network Model and its Application in Water Quality Evaluation”, 2009 International Conference on Environmental Science and Information Application Technology