Navtej Bhatt *et al*, International Journal of Computer Science and Mobile Applications,
Vol.3 Issue. 8, August- 2015, pg. 14-18 **ISSN: 2321-8363**

# A Survey based Research for Data Mining Techniques to Forecast Water Demand in Irrigation

**Mr. Navtej Bhatt[1], Dr. P.V.Virparia[2]**

[1]*Assistant Professor, MCA Department, Faculty of IT & Computer Science, Parul University, Vadodara*
*bhattnavtej@gmail.com*
[2]*Professor & Director, Department of Computer Science, Sardar Patel University, Vallabh Vidyanagar*
*pvvirparia@yahoo.com*

## Abstract
Data mining is an emerging technology in the recent decade. There are several application areas in which data mining is used widely. It is also useful in agriculture for various purposes. In this paper our focus is on the various techniques used to predict/forecast the water demand for agriculture purpose. The techniques discussed in the paper are Artificial Neural Network (ANN), Decision Tree (DT), Support Vector Machine (SVM), Evapotranspiration (ETc) and Systematically developed Forest (SysFor). Prediction of water demand is a very important factor that remains to be solved based on available data. The prediction of water demand can be done with the various data mining techniques.

***Keywords:*** Data Mining, Water Demand Forecasting, Decision Tree, ANN

## 1. Introduction
Water scarcity is a major issue in the developed countries that affects the agriculture. As the scarcity of the water increases, the demand for managing available water resources becomes crucial [4].
Due to change in the climate, growth of population and heavy demand of water for domestic and industrial use, it is necessary to utilize the available water resources. Since all the existing water resources are fully utilized and drawing of more water is impracticable, the best alternative is to increase the water productivity [1].
Studies report that, the water delivered for irrigation is not always efficiently used for crop production, on an average 25% of water is wasted due to inefficient water management practices[2][3].
The agriculture scientists, government, farmers are trying to put extra efforts and techniques to improve the productivity of the water. New methods, techniques and tools are available but still a very few farmers are actually using this for better production. As a result, there is a need to develop a Decision Support System (DSS) for effective management and utilization of water during on farm irrigation process. This system on agriculture will improve the performance on agriculture which in turn will be helpful for sustainable agriculture and better crop productivity [5]. This DSS model will helpful to improve water management and maximize water productivity.
Data mining can be used for predicting the future trends of agricultural processes. Recently, data mining techniques are increasingly being applied in the field of hydrology for developing models to predict various hydrological attributes such as rainfall, pan evapotranspiration, flood forecasting, weather forecasting etc. [6].
Data mining uses a number of analytical tools such as decision trees, neural networks, fuzzy logic, rough sets, and genetic algorithms to perform classification, prediction, clustering, summarisation, and optimization. The most common tasks among these are classification and prediction [7].
To develop an effective irrigation water demand forecasting model using data mining techniques adequate historical data for the attributes having high influence on water usage are required [2].

 

## 2. Literature Survey

Many researchers have work in the direction of Water demand forecasting in order to enhance the productivity of crop. Some of them have approached through Data Mining algorithms and Traditional Statistical methods. Here, we have gone through the research work carried out in this area and highlighted their outcomes.

Khan et. al.(2012), have analysed various data mining techniques like SysFor, ANN, Decision Tree, SVM and Traditional ETc method. They had compared the prediction accuracies of the various techniques with the actual water consumed by the various crops [4].

Pulido-Calvo et. al. (2003), have proposed the use of linear multiple regression univariate time series models and Computational Neural Network (CNN) in the development of the prediction model for water demand forecasting [9].

Pulido-Calvo et. al. (2009), have developed a hybrid methodology by combining CNN, fuzzy logic and genetic algorithm. The model was used to develop policies on irrigation water consumptions. They also derived that this type of model will have much better statistics and error measures than of traditional neural network models [8].

Shrinavasan Raju K. and Nagesh Kumar D. (2004) have presented Genetic Algorithm based model for evolving optimum cropping pattern. The objective of the model was to maximize net benefit through calculating water demand and cropping pattern [10].

Bingxiang, Z. (2012) has presented a method for feature extraction of irrigation water requirement based on kernel canonical correlation analysis. By this first he had extracted non-linear features then based on Least Square Support Vector Machine (LS-SVM) regression modelling, prediction of irrigation water requirement was implemented [11].

Fang X. and Deshan T. have used LS-SVM method which can efficiently handle non-linear factors that influenced the irrigation water. They have proposed forecasting model of irrigation water requirement based on it. The model was also compared with Back Propagation Artificial Neural Network (BPANN) forecasting [12].

## 3. Methods Identified

Through the survey of various research papers some methods are identified that can be used for water demand forecasting in agriculture. All the methods are well known and well established.

### 3.1 Decision Tree (DT) analysis

Decision trees are data classification tool used to find the hidden relationship between classifying and classifier attributes. Any attribute of the data set that contains the value, which are possible outcomes of the record are classifying attributes. In other words, a decision tree explores patterns also known as logic rules from any data set [13]. By using the rules generated by a decision tree the relationship between the attributes of a dataset can be extracted. Each rule represents a unique path from the root node to each leaf of the tree [4].

### 3.2 Support Vector Machine (SVM)

Support Vector Machine is a set of supervised learning methods. It is used for classification, regression and outliers detection. SVM can be used for maximizing any function to given dataset. This algorithm works with the decision plane, the soft margin, the kernel function and the decision boundaries. A decision plane is one that separates between a set of objects having different class memberships [14].

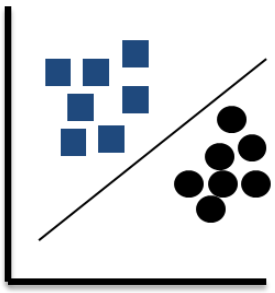

**Figure 1:** Support Vector Machine

 

Figure 1 exhibits the basic concept of support vector machine. The objects belong either to class Filled Square or Filled Circle. From Figure 1, it can be seen that the decision plane separates the two different objects from the dataset with decision boundary. Any new object can be classified by its margin to the decision boundary and based on that it will be classified.

### 3.3 Systematically Developed Forest of Multiple Trees (SysFor)

SysFor is a multiple tree building technique based on the concept of gain ratio. The technique was developed by Islam & Giggins in 2011[15]. The purpose of building multiple trees is to gain better knowledge through the extraction of multiple patterns.

First of all, a set of good attributed and their separation points will identified based on the gain ration and separation value defined by user. Islam & Giggins (2011) says, a numerical attribute can be chosen more than once within the set of good attributes, if it has higher gain ratios with different split points that are not close to each other. After the set of good attributes are selected and if the a user defined number of trees are larger than the size of the good attributes, then SysFor builds the tree by considering a good attribute as the root attribute and then it will generate trees an many as number of good attributes otherwise it will build user defined number of trees. The process continues until user defined number of trees are build/generated [4].

At last, after generation of all trees, the leaf with the highest accuracy is determined based on maximum number records with same class values to total number of records and finally the majority class value of the lead is chosen as the predicted class value of the record [4].

### 3.4 Evapotranspiration (ETc) Method

The crop evapotranspiration under standard condition (ETc), is the evapotranspiration from disease-free, well-fertilized crops, grown in large fields, under optimum soil water conditions, and achieving full production under the given climate conditions. The formula to calculate Evapotranspiration is as below [16]:

**Equation 1:**

$$ETc = Kc \times ETo$$

Where:

ETc = Crop evapotranspiration (mm/day)
ETo= Reference Crop evapotranspiration (mm/day)
Kc = Crop coefficient

ETc is traditional commonly used technique to estimate water demand. The crop coefficient method was developed for the agriculture users to calculate ETc which helps them in making irrigation management decisions [4].

### 3.5 Artificial Neural Network

Artificial Neural Network (ANN) is a data processing and classification model that is inspired by the biological neural network. ANN learns the non-linear relationships, trends and patterns from training dataset and uses the knowledge for predicting the class values of unseen datasets [4]. Following figure shows the architecture of most popular multilayer perception ANN.

 

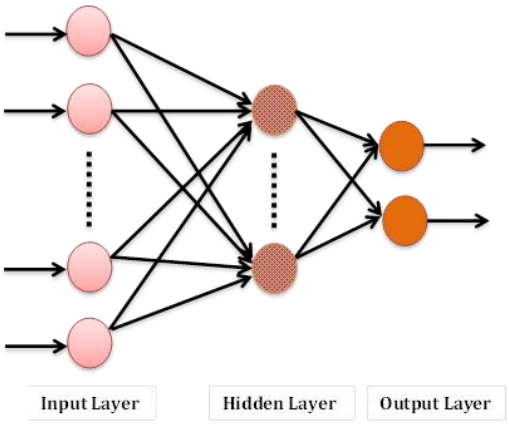


**Figure 2:** Artificial Neural Network

In a multilayer perception network there is an input layer, an output layer and one or more hidden layers. These layers extract patterns from a dataset and use the learned patterns to predict class values of new records. The nodes in the input layer pass the processed information to the computational nodes in a forward direction. The hidden layer is also responsible for resolving the non linearity between the input and output attributes of the data set [4].

Training of the model can be carried out either by supervised learning or unsupervised learning. The ANN model is based on the principle of error minimization. During the training process first the weights of the neurons computed which are generally interconnection between layers and used to store the knowledge gained.

The number of layers and nodes in each layer of the network will be decided by the user, based on the nature of the dataset. Literature study says that in hydrological modelling most ANNs are trained with single hidden layer.

## 4. Conclusion

The paper has narrated the actual scenario in agriculture for the sake of water demand and supply chain mechanism. It has been concluded that data mining can play an effective role to address this problem. As a result of literature survey, we have identified the application of data mining in this field through decision tree induction, Support Vector Machine, forest of multiple trees, Evapotranspiration, Artificial Neural Network etc. methods. A model discovery to predict water demand will be the ultimate and expected outcome of the future work.

## References

[1]Shaw M., 2003, Writing good software engineering research papers, *In Proceedings of 25th International Conference on Software Engineering*, pp.726-736

[1] Khan, M., Islam, M. Z., Hafeez, M., 2011, Irrigation Water Demand Forecasting – A Data Pre-Processing and Data Mining Approach based on Spatio-Temporal Data, *In Proceedings of 9th Australasian Data Mining Conference (AusDm11), CRPIT*, 121, pp.183-194

[2] Smith, M., 2000, The application of climatic data for planning and management of sustainable rainfed and irrigation crop production, *Agricultural and Forest Meteorology*, 102, pp.99-108

[3] FAO, 1994 *Water for Life. World Food Day* 1994, Rome.

[4] Khan, M., Islam, M. Z., Hafeez, M.,2012,Evaluating the Performance of Several Data Mining Methods for Predicting Irrigation Water Requirement, *In Proceedings of the Tenth Australasian Data Mining Conference (AusDM 2012), CRIPT, 134*, pp.199-207

[5] Bhatt, N., Virparia, P.,A Decision Support model for better Crop Productivity through Irrigation Water in Saurashtra Region, *International Journal on Recent and Innovation Trends in Computing and Communication, 2(3)*, pp.545-548

 

[6] Pulido-Calvo, I., Roldan, J., Lopez-Luque, R. And Gutierrez-Estrada, J.C., 2003: Demand Forecasting for Irrigation Water Distribution Systems, *Journal of Irrigation and Drainage Engineering, 129(6)*, pp.422-431
[7] Zurada, J. and Lonial, S., 2005, Comparison of the Performance of Several Data Mining Methods for bad Debt Recovery In The Health Industry, *The Journal of Applied Business Research, 21*,pp. 37-54
[8] Pulido-Calvo, I. and Gutierrez-Estrada, J.C., 2009, Improved irrigation water demand forecasting using soft-computing hybrid model, *Biosystems Engineering, 102*, pp.202-218
[9] Pulido-Calvo, I., Roldan, J., Lopez-Luque, R. and Gutierrez-Estrada, J.C., 2003, Demand Forecasting for Irrigation Water Distribution Systems, *Journal of Irrigation and Drainage Engineering, 129(6)*, pp.422-431
[10] Raju Shrinavasan, K. and Nagesh Kumar, D., 2004, Irrigation Planning using Genetic Algorithms, *Water Resources Management, 18*, pp.163–176
[11] Bingxiang, Z., 2012, Research on Model of Predicting Irrigation Water Requirement Based on Kernel Method, *in proceedings of 2nd International Conference on Electronic & Mechanical Engineering and Information Technology (EMEIT-2012), Atlantis Press*, pp.1338-1341
[12] Fang, X. and Deshan, T., 2010, Forecasting Irrigation Water Requirement Based on Least Squares Support Vector Machine, *Journal of Anhui Agriculture Science, 38*, pp.10273-10275
[13] Islam, M. Z., 2010, EXPLORE: A Novel Decision Tree Classification Algorithm, *the 27th International Information Systems Conference, British National Conference on Databases*,
[14] DELL, STATSOFT, *http://www.statsoft.com/textbook/support-vector-machines accessed on 25/07/2015*.
[15] Islam, M. Z. and Giggins, H., 2011, Knowledge Discovery through SysFor: A Systematically Developed Forest of Multiple Decision Trees, *In Proceedings of the 9th Australasian Data Mining Conference (AusDM 11), CRPIT, 121*, pp.195-204
[16] Savva, A. and Frenken, K., 2002, Crop water Requirement and Irrigation Scheduling, *Irrigation Manual, Module 4*, Chapter 1, Page 2

 