Biological Forum – An International Journal **7**(1): 1030-1032(2015)

ISSN No. (Print): 0975-1130
ISSN No. (Online): 2249-3239

Evaluation of Eight models in Predicting Software Weka Potential Evapotranspiration month for the next month in the Synoptic Weather station Babolsar

Seyyed Hassan Mirhashemi* and Mehdi Panahi**
**Ph.D. student of Irrigation and Drainage, Water Engineering Department, University of Zabol, IRAN*
***Assistant Professor of Water Engineering Department, College of Agriculture, University of Zanjan, IRAN*

(Corresponding author: Seyyed Hassan Mirhashemi)
(Received 12 March, 2015, Accepted 20 April, 2015)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: In this paper have been evaluated the ability of eight models of weka software to estimation "Monthly potential evapotranspiration months later," for Babolsar synoptic weather stations. That is including additive Regression, Bagging, Linear Regression, Zero, M5P, Kstar, M5Rules and REPTree. The data used in this article, are the average monthly data of Babolsar weather station, including: "average temperature", "sunshine hours", "dew point", "relative humidity", "average wind speed" and "saturation vapor pressure deficit" in forty-six-year period from 1960 to 2005 AD. Output variables used, was "Monthly potential evapotranspiration months later," as monthly basis. After introducing the weather data as mean monthly to the algorithm as input variables and monthly potential evapotranspiration months later, as the output variables, models "data mining" were evaluated using "correlation coefficient", "Root Mean Square Error" and "mean absolute error". According to the statistical indexes, Tree Bagging models have better function in estimating the monthly average temperature for the month.

Keywords: "Data mining" weka, Penman-Monteith equation, potential evapotranspiration, synoptic weather station, Babolsar

INTRODUCTION

Properly estimate evapotranspiration is crucial to water resources management, irrigation scheduling and evaluate the effect of "Land Use Change" on "efficiency and provision of crop water needs". Physical and "semi-theoretical" models proposed to modeling potential evapotranspiration, the most complex and needs many weather variables. For this reason, models and other techniques such as "multiple regression", "factor analysis", "artificial network" and "modern methods" data mining" used to estimate these quantities and determine the influence. Recently, a new data mining techniques have been successfully used in environmental sciences. Meteorological data those are measured and archives at different stations included a large volume of information and are increasing their volume over time. Accordingly, most felt the need for new methods of "data mining" of them. In some cases, a lot of variables are used that may be some of them are not measure in all weather stations. Therefore is necessary to make use of modern techniques such as data mining.

A. Definitions of data - mining

Data mining has many broad definitions. The definitions lot depends on the individual backgrounds and points of view. So we can say that the data mining is a set of methods in process of knowledge discovery that used to recognize patterns and undisclosed relationships in data. Data mining can also be said is a process recognition valid, new, useful and understandable pattern, from data. Data mining is a technique that combines hypothesis tests and derives data- discovered. In the Assuming tests, researchers can test ideas against the data to confirm or refute its validity [2]. Vandenberg and colleagues (1999) explain that discovery; the researcher draws conclusions from the data and allows the data to accept the result. The most data mining problems is solved using a combination of both methods. For example, the result may be a new hypothesis that can be tested and the test will be approved or rejected [10]. Data mining is the process of selecting, identifying and modeling large amounts of data.

In another definition, the process of selecting, exploring and modeling large data mining officials. To discover hidden relationships and achieving results clearly beneficial for the owner of the database [3]. Data mining is a process that uses various tools to analyze data, to the physical changing patterns and relationships found in different data sets [9]. The main difference between data mining and statistics, that is data mining is one approach without the default. While most conventional statistical techniques are needed to default. And statistical professionals are searching equations to match the defaults. In contrast, data mining algorithms can automatically develop these equations from information contained in the data set [5].

MATERIALS AND METHODS

Babolsar is center of Babolsar city, is located on both sides of the Babol River and post-strand of the Caspian Sea. This city is located in 21 meters below sea level , in 20 kilometers north of the Babol city and 231 km north-east of Tehran . The mean annual temperature is 16.8 centigrade and sometimes appears to 42 centigrade in the highest temperature and coldest temperature was -7 centigrade. Annual rainfall is about 886 mm. Many scientists have studied the Penman-Monteith equation to estimate ETO (Allen *et al*, 1998; De uder *et al* 1995). Jensen (1990) were analyzed the performance of 20 different methods against "ET" was measured for 11 stations located in different climatic zones of the world. Penman-Monteith method was ranked as the best method for all climatic conditions. Application of Penman-Monteith equation FAO - 56 needs the data of "sunlight", "Wind speed", "Temperature", "vapor pressure" and humidity, but all these input variables are not readily available in every location. In developing countries, the correct data collection problems and these can all climatic variables encountered in application Penman-Monteith equation FAO - 56 is considered a serious problem.

Software Weka has developed in Waikato University in New Zealand and its name has been extracted from the phrase "Waikato Environment for knowledge Analysis". Also Weka, called founder wild bird that does not fly and is found in New Zealand. The system is written in Java and has been published based on sweeping the GNU General Public License. Weka almost runs on any platform and is tested under Linux, Windows, Mac, and even a digital receptionist person. This Software is sweeping an interface to many different learning algorithms, that through it's the procedure pre- process, post - process and evaluate learning schemes on all data sets, are applicable. This environment is includes issues procedures for all standard "data mining" such as regression, classification, clustering, "exploring the association rules" and "feature selection". Considering that the data are an integral part many of the tools - data processing and visualization has been provided. All algorithms get their entry into as relational table to ARFF format. This format of the data can be generated from a read file or from a database by a query. In this study, we was used eight different models of Weka software which was include additive Regression, Bagging, Linear Regression, Zero, M5P, Kstar, M5Rules, REPTree to predict "potential evapotranspiration months later". Also used monthly meteorological data of " Babolsar synoptic weather stations" as "inputs data" that includes: "The average temperature" (c), "Sunny Hours" (h), "Dew point" (c), "Relative humidity" (percent), "Average wind speed" (meters per second) and Saturation vapor pressure deficit (mbar) in forty-six-year period from 1960 to 2005. 75% of the data as a production model using Bagging model and 25% of them was used as the test model.

To predict monthly potential evapotranspiration next month used six variables: "Sunny Hours" (h), "Dew point" (c), "Relative humidity" (percent), "Average wind speed" (meters per second) and Saturation vapor pressure deficit (mbar) and "Average temperature"

That was considered all as monthly basis month after month as the input data and the monthly potential evapotranspiration next month as the output data.

RESULTS

To calculate "monthly potential evapotranspiration next month" was used average monthly data series of Babolsar station.

Values "monthly potential evapotranspiration next month" estimated from the eight models were compared with monthly potential evapotranspiration next month", calculated by Penman-Monteith equation by "correlation coefficient", "Root Mean Square Error" and "mean absolute error".

As can be seen in Table 1 Bagging tree model with a correlation coefficient of 0.9075, RMSE of 2.2041 MAE of 2.8977 and the appropriate model was selected to estimate the "Monthly potential evapotranspiration next months". To determine the most important factor for modeling "average monthly potential evapotranspiration for the next month" via Bagging tree model were compared by changing the input data and using the statistical parameters.

Which contains the "correlation coefficient", "Root Mean Square Error" and "mean absolute error" when compared third row were include five meteorological parameters have most "regression" and less "square root error" and mean "absolute error".

Table 1: Comparison ten models weka software with three statistical indices.

Statistical indices / Models	R	MAE	RMSE
Additive Regression	0.8706	2.6619	3.4081
Bagging	0.9075	2.2041	2.8977
Kstar	0.8996	2.2402	3.0244
Linear Regression	0.8989	2.4429	3.0235
M5P	0.8966	2.4507	3.0568
M5Rules	0.8963	2.4534	3.0613
REP Tree	0.8865	2.4873	3.2064
Zero	-0.1187	6.1874	6.9107

As a result, five parameters were used have the greatest impact in the function tree model Bagging. Then the six parameters are located in the first row and are including the parameter "average monthly temperature", "sun hours", "dew point", "relative humidity", "wind speed" and "Saturation vapor pressure deficit" and the five parameters are located in second row are including parameters' average monthly temperature "," wind speed "," relative humidity ": "Saturation vapor pressure deficit" and "Sunny Hours"are the second and third ranks respectively in the positive impact Bagging model to proper functioning in the estimating average monthly evapotranspiration for the next month.

Table 2: The combination of input parameters to estimate the monthly potential evapotranspiration for the next month, using Bagging model.

Statistical indices / Combination of input parameters	R	MAE	RMSE
T,n,w,RH,dwe,e	0.9075	2.2041	2.8977
T,n,w,RH,e	0.9074	2.2085	2.8993
T,n,RH,dwe,e	0.8997	2.2748	3.0123
T,n,w,e	0.9077	2.2079	2.8947
T,n,w,RH	0.9032	2.2395	2.961
T,n,w	0.8975	2.3135	3.0427
T,n	0.8877	2.4186	3.1781
T,w	0.8606	2.8299	3.5149
n,w	0.8345	2.8919	3.8049

In Table 2 "dew point" (c) "relative humidity" (percent), "Sunny Hours" (h), "Saturation vapor pressure deficit" (mbar), "Wind speed" (m/s), "the average monthly temperature '(C) respectively is shown as dwe, RH, n, e, W, T.

CONCLUSION

From this study it can be concluded that Techniques of "data mining" such Weka software models can be used to estimate evapotranspiration potential of the next month.

Bagging model with an estimate of monthly potential evapotranspiration for the next month is shown that can have a high capacity to estimating meteorological parameters.

This model can be used to estimate "potential evapotranspiration" used in a variety of stations that are deficient in recorded meteorological parameters.

It was concluded from Table 2: Sensitivity to model weather Bagging enter the six parameters, including "average temperature" (c), "sunny hours" (h), "dew point temperature" (c), "The average relative humidity" (%), " The average speed wind "(meters per second) and" saturation vapor pressure deficit "(mbar) as "input variables" have the best performance, relative to the composition of the other parameters in Table 2.

REFERENCES

Mamashkani, A. Nazemi, A.R. Introduction to "data mining", Neishabour branch of Islamic Azad University Press, 1388, pp. 456.

PH Cabena., Stadler R., Verhees J., and Zanasi (1998). Discovering data mining: From concept to implementation, IMB, New Jersey, 195 pp.

P Giudici. (2003). Applied data Mining: statistical methods for business and industry. Wily, London . pp. 364.

JR Quinlan. (1992). Learning with continuous classes. Proceeding of Australian Joint Conference on Artificial Intelligence. World Scientific Press: Singapore; 343-348.

T Crows Corporation, (1999). Introduction to data mining and knowledge discovery, third ed., Postmac, MD. Available at: www.twocrows.com, (April 29, 2000).

H Vanderberg., Sogard P. and Motoroni S. (1999). MineSetTM 3.0 Enterprise Edition Tutorial for Windows, Doc. No. 007-4006-001, Silicon Graph.

ME Jensen., Burman RD. and Allen RG. (1990). Evapotranspiration and irrigation water requirements. ASCE Manual and Reports on Engineering Practice No. 70. ASCE: New York.