IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 2, Ver. V (Mar-Apr. 2016), PP 01-02
www.iosrjournals.org

Comparison of Various Data-Driven Modelling Techniques for Inflow Analysis

Manali Pawar, Samveda Mohite, Rushikesh Deshmukh, Prof. Nivedita Bhirud
Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune-48.Savitribai Phule Pune University Pune

***Abstract*:** *Inflow Analysis of a reservoir is an important factor as water is becoming a scarce resource and hence management of water was essential. There are many traditional methods of analysing the reservoir inflow but which were very tedious and time consuming and hence we approach the soft computing data-driven techniques. In this paper comparisons of different data-driven techniques, namely, Artificial Neural Networks (ANN) and M5 Model tree (MT) is given, which gave better results with relative ease. Here past data viz. the rainfall-runoff and the inflow of the rain gauge stations surrounding the Chaskaman Reservoir, is collected, studied and analysed using the data-driven techniques and prediction is made. It was found that M5 Model tree technique performed reasonably well and gave more accurate results than the other techniques.*

Keywords*: Inflow analysis, artificial neural networks, M5 model trees, rainfall runoff*

I. Introduction

Management of reservoirs is essential as water is becoming a limited resource and also it provides flood and drought control safety measures and guidance of reservoir planning and management. Forecasting of reservoir inflow is rather an important factor for giving optimal sharing of water supplies to various contending agencies namely for domestic, irrigation, industry etc. The previous hydrological methods were complex and problematic having great temporal and spatial inconsistency and required more time for analysing the inflow. It took into account many factors such as geomorphologic, climatic, catchment, etc of whose data was difficult to collect and analyse. Hence in place of this the data-driven techniques were applied which used the attributes like direct rainfall-runoff and the inflow from the rain gauge stations. There are nearly six rain gauge stations surrounding the Chaskaman Reservoir in Bhima basin of Maharashtra, they are Wada, Kude, Goregaon, Bhorgiri, Bhimashankar and Adharshingi. So the data from these rain gauge stations is collected and mapped with the direct rainfall-runoff using linear regression method to get an M5 model algorithm.

Due to high complexities of the traditional methods which had many disadvantages researchers moved on to machine learning techniques which were comparatively simpler and gave more precise inflow forecast. The techniques like ANN and Model tree gave promising ways in hydrological prediction (Londhe & Charhate, (2010)). Out of which Model trees performed the best.

II. Data-Driven Techniques

Data-driven techniques are the computational techniques for the hydrological modelling of a system which gives relationship between inputs and outputs of the system which results into machine learning algorithms based on the training data set. The model generated is then further used to determine the unseen data. Few of the data-driven techniques are listed below.

A. Artificial Neural Networks (ANN's)

Artificial neural networks is evolved from the biological neural networks of the human brain and this process (ANN) was introduced into machine learning which was used to model the training data set. This architecture is basically used for supervised training data set. ANN gives a solution as an intricate, non-linear function designed from many attribute such that the calculated output is adjusted using back propagation algorithm to the expected values. The method involves basic three layers, input layer, and an output layer and in between them is the hidden layer. There are mathematical terms like weights and bias that are associated with the input layer. Firstly, the weights are multiplied by the data from the input layer, later the product of which is added by a bias to give the hidden layer data set. Then the output is obtained through a transfer function applied at the hidden layer. This process is recursively applied until the targeted output is reached through the practice called as back propagation. To meet the difference between the output generated and the targeted output an error function is generated. Once this network is obtained similar process is applied for unseen data.

DOI: 10.9790/0661-1802050102

B. M5 Model Trees

M5 model tree is the data-driven technique based on the concept of decision trees that is applied for classification problems. Into machine learning, when input attributes is used to determine or predict a class or define a model then model trees methodology is applied using the M5 algorithm.

M5 model tree algorithm divides the input space into subspaces and builds in a linear regression model for each of its sub-areas. MT uses same technique that is used in forming of decision trees, and it has linear regression functions at leaf nodes unlike that in decision trees. Model trees as ‘classifiers’ are more precise than the simple decision trees as they forecast continues numeric attributes at leaf nodes. Model trees learn powerfully and can handle tasks with high dimensionality which could range up to hundreds of attributes. Another advantage over regression trees is that, model trees are much smaller in size, their decision structure is clear and regression functions do not occupy many variables.

Construction for M5 Model Trees

Consider a collection of training data set T. Each training set is further divided into subsets and each subset is given a fixed set of attributes which are related to the target value. These attributes may be numeric or discrete. The target values of each training subset are then related to the values of other attributes. The significance of this model is thus analysed by the accuracy with which it can forecast the target values of unseen subsets.

The very first step is to create a decision tree. Collection of set T is either associated with a leaf node or some test is selected that could split T into subsets that are equivalent to the test outcomes and same process is applied recursively. The information collected in the M5 tree is measured by the Standard Deviation method prior and post to the test and then the training data set T is split based on the result of the test. The splitting of T is based on the Standard Deviation Reduction formula.

$$\Delta error = SD(T) - \sum_i \frac{|T_i|}{|T|} sd(T_i)$$

Where, T_i denotes subset of training data sets analogous to the i^{th} outcome of the test. Now consider the deviation sd(T_i) of target values of subsets in T_i as a measure of error. The expected error reduction is thus calculated and the maximised error reduction is chosen by M5.

After the splitting of the training dataset is done, pruning process is applied. Pruning is generally applied to improve capacity of the tree to generalize the unseen data using the linear regression at the leaf nodes progressing towards the root node. At each internal node, the estimated error of that node and the estimated error of the sub-tree below are compared and the node with lower estimated error is chosen by the M5 tree. Then the sub-tree is pruned if it does not improvise the performance of the tree.

Finally the tree undergoes the smoothing procedure that is used to forecast future values. When value of a subset is predicted by a model tree, it is compared with the values at the internal nodes and adjusted to reflect the forecasted values along the path from the root to the particular leaf node.

III. Conclusion

Traditional methods were more complicated and hence we approached the soft computing data-driven machine learning techniques that are the ANN and MT techniques which are instance based learning techniques. The model tree divides the input space into subspaces and then builds in a piece-wise linear model whereas, ANN builds in a non-linear model. In ANN, we need to find out the best topology, all the hidden layers and number of neurons in each hidden layer. This all depends on the trial and error method, hence it is very time consuming and thus ANN is not as transparent as MT. On the contrary, MT is non-parametric and more convenient. It is more understandable and uses simple rules. It was found that M5 Model tree technique performed reasonably well and gave more accurate results than ANN where model setting is easy, training the data set is fast and results are in linear equation format.

References

[1]. Comparison of data-driven modelling techniques for river flow forecasting, Shreenivas Londhe; Shrikant Charhate, Department of Civil Engineering, Vishwakarma Institute of Information Technology, Survey no.2/3/4, Kondhwa (Bk), Pune, MH, India b Department of Civil Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, MH, India.

[2]. Neural networks and M5 model trees in modelling water level–discharge relationship Bhattacharya, D.P. Solomatine, Department of Hydro informatics and Knowledge Management, UNESCO-IHE Institute for Water Education, P.O. Box 3015, 2601 DA Delft, The Netherlands.

[3]. Model trees as an alternative to neural networks in rainfall–runoff modelling, DIMITRI P. SOLOMATINE, International Institute for Infrastructural, Hydraulic and Environmental Engineering (IHE), PO Box 3015, NL-2601 DA Delft, The Netherlands, KHADA N. DULAL, Department of Hydrology and Meteorology, PO Box 406, Kathmandu, Nepal.

[4]. A Simple Regression Based Heuristic for Learning Model Trees Celine Vens and Hendrik Blockeel K.U.Leuven - Department of Computer Science.