Geophysical Research Abstracts
Vol. 12, EGU2010-14920-1, 2010
EGU General Assembly 2010

Ensembles of extremely randomized trees and feature ranking for streamflow prediction

Andrea Castelletti (1,2) and Stefano Galelli (1)

(1) Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milano, Italy (castelle@elet.polimi.it), (2) Centre for Water Research, University of Western Australia, Crawley, Australia

Accurate and reliable stream-flow predictions are an important input to water resources planning and management processes, which heavily depend upon the availability of water (e.g. river basin planning, optimal reservoir operation, irrigation system management). Hydrological processes are extremely complex, combining high non-linearity and spatial-temporal variability. The prediction of hydrological variables is therefore a challenging task, very often complicated by lack of data and/or the presence of outliers. Usually, data-driven modelling provides a good balance between model accuracy and complexity, which are ultimately critical to the adoption of optimization-based approaches. While neural networks have been widely used in hydrological modelling (e.g. Govindaraju and Rao, 2000), tree-based model is a relatively unexplored methodology (Solomatine and Dual, 2003; Solomatine and Xue, 2004; Iorgulescu and Beven, 2004; Stravs and Brilly, 2007). In this paper a new data-driven modelling approach based on Ensembles of Extremely Randomized Trees (ETs; Geurts et al., 2006) is proposed for stream-flow prediction using different hydro-meteorological predictors. By randomizing the tree construction process and merging a forest of diversified trees to predict the output, ETs alleviate the well-known poor generalization property of traditional standalone decision tress (e.g. CART), thus avoid over fitting the training data. Input to the model are selected using a tree-based feature ranking algorithm, which ranks the candidate predictors (e.g. precipitation and evaporation at different stations, linear combinations thereof) according to their contribution in explaining the variance of an underlying ETs-based model of the stream-flow process. The approach is applied in the Red river basin (Vietnam), a sub-tropical catchment characterized by extremely variable weather conditions, where strong precipitations significantly contribute to the high [U+FB02]ow. Results shown that combining ETs and ranking techniques provides good performance, compared to other data-driven methods (e.g. neural networks or ARX models).

References

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63 (1), 3-42.

Govindaraju, R. S. , Rao, A. R., 2000. Artificial Neural Network in Hydrology. Kluwer, Dordrecht, The Netherlands.

Iorgulescu, I., Beven, K., 2004. Nonparametric direct mapping of rainfall-runoff relationships: An alternative approach to data analysis and modeling?. Water Resources Research 40 (8), W08403.

Solomatine, D., Dulal, K., 2003. Model trees as an alternative to neural networks in rainfall-runoff modelling. Hydrological Sciences 48 (3), 399-411.

Solomatine, D., Xue, Y., 2004. M5 model trees compared to neural networks: application to flood forecasting in the upper reach of the huai river in china. ASCE Journal of Hydrologic Engineering 9 (6), 491-501.

Stravs, L., Brilly, M., 2007. Development of a low-flow forecasting model using the m5 machine learning method. Hydrological Sciences 52 (3), 466-477.