Geophysical Research Abstracts
Vol. 14, EGU2012-6683-1, 2012
EGU General Assembly 2012

Forward Greedy ANN input selection in a stacked framework with Adaboost.RT - A streamflow forecasting case study exploiting radar rainfall estimates

D. Brochero (1,2), F. Anctil (1), and C. Gagné (2)

(1) Université Laval, Chaire de recherche EDS en prévisions et actions hydrologiques, Department of Civil and Water Engineering, Quebec, Canada , (2) Université Laval, Computer Vision and Systems Laboratory (CVSL), Department of Electrical Engineering and Computer Engineering, Quebec, Canada

In input selection (or feature selection), modellers are interested in identifying k of the d dimensions that provide the most information. In hydrology, this problem is particularly relevant when dealing with temporally and spatially distributed data such as radar rainfall estimates or meteorological ensemble forecasts.

The most common approaches for input determination of artifitial neural networks (ANN) in water resources are cross-correlation, heuristics, embedding window analysis (chaos theory), and sensitivity analyses. We resorted here to Forward Greedy Selection (FGS), a sensitivity analysis, for identifying the inputs that maximize the performance of ANN forecasting. It consists of a pool of ANNs with different structures, initial weights, and training data subsets. The stacked ANN model was setup through the joint use of stop training and a special type of boosting for regression known as AdaBoost.RT.

Several ANN are then used in series, each one exploiting, with incremental probability, data with relative estimation error higher than a pre-set threshold value. The global estimate is then obtained from the aggregation of the estimates of the models (here the median value).

Two schemes are compared here, which differ in their input type. The first scheme looks at lagged radar rainfall estimates averaged over entire catchment (the average scenario), while the second scheme deals with the spatial variation fields of the radar rainfall estimates (the distributed scenario).

Results lead to three major findings. First, stacked ANN response outperforms the best single ANN (in the same way as many others reports). Second, a positive gain in the test subset of around 20%, when compared to the average scenario, is observed in the distributed scenario. However, the most important result from the selecting process is the final structure of the inputs, for the distributed scenario clearly outlines the areas with the greatest impact on forecasting in terms of the estimated radar precipitation and the forecast horizon. Thus, this research facilitates interpretability of the results under a downward approach, in which the zones of influence of rainfall at different forecasting horizons help understanding the pattern of the hydrologic response at the event scale. Third, the input selection is slightly different between experiments due to the active principle of diversity, defined as hydrological model complementarities addressing different aspects of the forecast.

Finally the following guidelines favoured efficient stacked ANNs: i) design different combiners, ii) base the stack preprocess on probabilistic score representations oriented toward the bias of the ensemble (e.g. the ignorance score), and iii) evaluate more powerful multi-criterion selection algorithms such as multiobjective evolutionary algorithms (MOEA).