Geophysical Research Abstracts, Vol. 9, 09665, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-09665

Tree-like machine learning models in hydrologic forecasting: optimality and expert knowledge

M. Siek, D.P. Solomatine

UNESCO-IHE Institute for Water Education (m.siek@unesco-ihe.org, d.solomatine@unesco-ihe.org)

Hydrologic forecasting is normally based on physically-based (process) models based on equations describing the water flow. Data-driven models use methods of machine learning (e.g., neural networks, support vector machines, etc.) and built using the historical data describing flows and hydrological loads. Modular models include sub-models each of which is responsible for a particular hydrological condition, and they are typically more accurate than the overall global models. Many algorithms for allocating such regions to local models typically do this in automatic fashion, and this is not well-accepted by experts. In forecasting natural processes domain experts want to bring in more knowledge into such allocation, and to have certain control over the choice of models. Another problem is that modular models often use so-called greedy algorithms which are sub-optimal and can be improved.

This paper presents an approache to make modular models more “expert-friendly”. An issue of including a domain expert into the modelling process is discussed, and updated M5flex algorithms in the class of model trees (piece-wise linear modular regression models) is presented. Another algorithm, which is also a modification to M5 algorithm, allows for building more accurate M5 model trees. Comparison of the algorithms based on modular local modeling to the more traditional “global” learning models on a number of benchmark tests and river flow forecasting problems shows their higher accuracy and the transparency of the resulting models.