Geophysical Research Abstracts
Vol. 16, EGU2014-4206, 2014
EGU General Assembly 2014

Multivariate Mapping of Environmental Data Using Extreme Learning Machines

Michael Leuenberger and Mikhail Kanevski

Institute of Earth Surface Dynamics (IDYST), University of Lausanne, Switzerland (Michael.Leuenberger@unil.ch, Mikhail.Kanevski@unil.ch)

In most real cases environmental data are multivariate, highly variable at several spatio-temporal scales, and are generated by nonlinear and complex phenomena. Mapping - spatial predictions of such data, is a challenging problem. Machine learning algorithms, being universal nonlinear tools, have demonstrated their efficiency in modelling of environmental spatial and space-time data (Kanevski et al. 2009). Recently, a new approach in machine learning - Extreme Learning Machine (ELM), has gained a great popularity.

ELM is a fast and powerful approach being a part of the machine learning algorithm category. Developed by G.-B. Huang et al. (2006), it follows the structure of a multilayer perceptron (MLP) with one single-hidden layer feedforward neural networks (SLFNs). The learning step of classical artificial neural networks, like MLP, deals with the optimization of weights and biases by using gradient-based learning algorithm (e.g. back-propagation algorithm). Opposed to this optimization phase, which can fall into local minima, ELM generates randomly the weights between the input layer and the hidden layer and also the biases in the hidden layer. By this initialization, it optimizes just the weight vector between the hidden layer and the output layer in a single way. The main advantage of this algorithm is the speed of the learning step. In a theoretical context and by growing the number of hidden nodes, the algorithm can learn any set of training data with zero error. To avoid overfitting, cross-validation method or "true validation" (by randomly splitting data into training, validation and testing subsets) are recommended in order to find an optimal number of neurons. With its universal property and solid theoretical basis, ELM is a good machine learning algorithm which can push the field forward.

The present research deals with an extension of ELM to multivariate output modelling and application of ELM to the real data case study - pollution of the sediments in Lake Geneva. The basic idea is to model several pollutants together taking into account complex dependencies between them. The original data set consists of 200 measurements on 8 pollutants. Pairwise analysis of the variables using correlation matrix shows different relationships between them: linear correlations, nonlinear correlations, no correlations. The use of different combination of pollutants helps to understand the behavior of ELM and to perform detailed methodological analysis. Besides this real data case study, simulated patterns (generated by adding shuffled data) were analyzed as well.

The methodology proposed and preliminary results obtained are very promising and establish a solid basis for more challenging case studies on high dimensional and multivariate data relating to environmental risks and natural hazards.

References

Huang G.-B., Zhu Q.-Y., and Siew C.-K. 2006, Extreme learning machine: theory and applications, Neurocomputing, vol. 70: 489-501.

Kanevski M., Pozdnoukhov A., and Timonin V. 2009, Machine Learning for Spatial Environmental Data. Theory, Applications and Software, EPLF Press, Lausanne.