Geophysical Research Abstracts
Vol. 12, EGU2010-2915-1, 2010
EGU General Assembly 2010

Environmental Monitoring Networks Optimization Using Advanced Active Learning Algorithms

Mikhail Kanevski, Michele Volpi, and Loris Copa
Institute of Geomatics and Analysis of Risk, University of Lausanne, Switzerland (Mikhail.Kanevski@unil.ch)

The problem of environmental monitoring networks optimization (MNO) belongs to one of the basic and fundamental tasks in spatio-temporal data collection, analysis, and modeling. There are several approaches to this problem, which can be considered as a design or redesign of monitoring network by applying some optimization criteria. The most developed and widespread methods are based on geostatistics (family of kriging models, conditional stochastic simulations). In geostatistics the variance is mainly used as an optimization criterion which has some advantages and drawbacks. In the present research we study an application of advanced techniques following from the statistical learning theory (SLT) - support vector machines (SVM) and the optimization of monitoring networks when dealing with a classification problem (data are discrete values/classes: hydrogeological units, soil types, pollution decision levels, etc.) is considered. SVM is a universal nonlinear modeling tool for classification problems in high dimensional spaces. The SVM solution is maximizing the decision boundary between classes and has a good generalization property for noisy data. The sparse solution of SVM is based on support vectors – data which contribute to the solution with nonzero weights. Fundamentally the MNO for classification problems can be considered as a task of selecting new measurement points which increase the quality of spatial classification and reduce the testing error (error on new independent measurements). In SLT this is a typical problem of active learning – a selection of the new unlabelled points which efficiently reduce the testing error. A classical approach (margin sampling) to active learning is to sample the points closest to the classification boundary. This solution is suboptimal when points (or generally the dataset) are redundant for the same class.
In the present research we propose and study two new advanced methods of active learning adapted to the solution of MNO problem: 1) hierarchical top-down clustering in an input space in order to remove redundancy when data are clustered, and 2) a general method (independent on classifier) which gives posterior probabilities that can be used to define the classifier confidence and corresponding proposals for new measurement points. The basic ideas and procedures are explained by applying simulated data sets.
The real case study deals with the analysis and mapping of soil types, which is a multi-class classification problem. Maps of soil types are important for the analysis and 3D modeling of heavy metals migration in soil and prediction risk mapping. The results obtained demonstrate the high quality of SVM mapping and efficiency of monitoring network optimization by using active learning approaches.
The research was partly supported by SNSF projects No. 200021-126505 and 200020-121835.