Geophysical Research Abstracts, Vol. 9, 01321, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-01321

Support Vector Model Selection for Environmental Mapping

A. Pozdnoukhov, M. Kanevski

Institute of Geomatics and Analysis of Risk, UNI Lausanne, Switzerland
(Alexei.Pozdnoukhov@unil.ch)

Support Vector algorithms for regression, classification and novelty detection became a widely used tool in data-driven modeling. They have gained this popularity due to the number of reasons. Being independent on the dimensionality of the feature space, they develop robust non-linear models with remarkable generalization abilities. They are based on the solid background of Statistical Learning Theory and implement the Structural Risk Minimization principle. According to this inductive learning principle, the optimal predictive model for the given data modeling task has to be built by finding the trade-off between the model complexity and its fit to training data. This trade-off is controlled by the hyper-parameters of the Support Vector algorithms, including the type and parameters of its kernel function. The choice of the latter is very important, since the Support Vector model is the kernel expansion – the linear combination of kernels. Support Vector algorithms were found to provide promising results in environmental prediction modeling.

The choice of optimal parameters is often the key point of the model development. Given some prior information, it is possible to tune beforehand and fix the parameters of the mapping algorithm. The resulting model is highly dependent on the hyper-parameters, which are closely linked to such properties as characteristic spatial scale and small-scale data variability. The optimal parameters of the model can be considered as the characteristics of the modeled natural process. In real-time monitoring under the changing or emergency conditions, it may be required to adapt the parameters of the SVM according to the newly observed data. The significant changes in optimal values of the parameters may indicate the changes in the observed process, such as emergency fallouts, newly emerged conditions, extreme events, etc. The specific approaches have to be developed in order to make SVMs to be the fully appropriate tool

for detection and automatic spatial mapping.

The contribution of this paper is twofold. First, it presents a review of the automatic methods for selecting the multiple parameters of Support Vector Machine. The applicability of these methods in real-world conditions is discussed. Next, some useful heuristics (specific to spatial data modeling) and empirical findings to fix the near-optimal values of the parameters are presented. Gaussian Radial Basis Functions is a common choice of the kernels in spatial data modeling. They are considered in more details. The parameter of the kernel is related to the characteristic spatial scale of the process which generates the data. The use of exploratory variography is elaborated in this context. The other issues which are discussed are the spatial characteristics of the monitoring network and data variability.

The developed method is validated on the real data on radioactive fallout of Cs^{137} due to Chernobyl accident, and Swiss meteorological data on precipitation in Alpine regions.