Support Vector Regression for Environmental prediction

Nicolas Gilardi and Grégoire Dubois
gilardi@idiap.ch; gregoire.dubois@usa.net
Dept. des Sciences de la Terre, Université de Lausanne, Switzerland,

1. INTRODUCTION

Geostatistics are widely used to predict values of an environmental variable at unsampled locations. Ordinary kriging, which is the geostatistical function that is the most commonly used, is well known as being the best linear unbiased predictor (BLUP) [Cressie, 1993]. The predictions are however strongly dependant on the analysis and the modelling of the spatial correlation of the variable as well as on the validity of certain hypotheses that are frequently not verified. Machine learning algorithms, like Artificial Neural Networks (ANN), are promising alternatives to methods that require a priori models and/or knowledge of the data. They are able to handle large amount of data and can deal with non-linear systems. Because of these characteristics, machine learning algorithms are interesting candidates for the automatic processing of environmental data, which is necessary in emergency situations. ANN have however shown to be difficult to use for the mapping of environmental variables (rainfall, deposition of a pollutant, …) that frequently show strong local variability [De Bollivier *et al.*, 1997]. Support Vector Machine [Vapnik, 1998] are other learning algorithms which potential also worth to be explored.

2. SUPPORT VECTOR REGRESSION

Support Vector Machine (SVM) algorithms are based on the Statistical Learning theory [Vapnik, 1998] which introduces the notion of "structural risk minimisation". The underlying idea is that reducing the complexity of a model will increase its generalisation power. The SVM algorithms, initially developed for classification purposes [Boser, 1992][Burges, 1998] and extended to regression [Smola et al, 1998], are automatically minimising this structural risk. One can so expect to obtain an optimal model in term of generalisation error for a given problem. More information on SVM and related methods can be found at http://www.kernel-machines.org/.

SVR algorithm uses usually two hyper parameters:

- "ε", related to the noise that is present in the data. When the noise is low, ε is reduced in order to take a maximum of training points into account.
- "C", is called the "soft margin" parameter. It is a coefficient giving some robustness to the algorithm. Low values will have a smoothing effect, but will increase stability.

The SVR are using a "kernel function" for the estimation. For the following case study, the Gaussian radial basis function has been used. Its definition is given by

$$f(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{2\sigma^2}\right)$$

where **x**1 an d **x**2 are vectors of co-ordinates. The kernel parameter of this function is σ, the standard deviation of the Gaussian function.

3. CASE STUDY

100 daily rainfall measurements have been used as training data to predict the values of the variable at other 367 locations. The data set is the one used for the Spatial Interpolation Comparison 97 that is described in [Dubois, 1998][1]. The Root Mean Squared Errors (RMSE) and the Mean Absolute Errors (MEA) of the estimations obtained with SVR, Ordinary Kriging (OK), and the Inverse Distance Weighted function (IDW) are compared in table 1. The parameters of the last three functions have been optimised by cross-validations. The approach has been kept as simple as possible: all data have been used during the training and no anisotropy has been taken into account.

Table 1. Comparison of the results of the three methods. The true values of the 367 points are given in brackets

	Min (SIC97=0)	Max (SIC97=517)	Mean (SIC97=185)	Std Deviation (SIC97=111)	RMSE	MAE
SVR (absolute deviation)	**0** **(0)**	376.6 (140.4)	**183.6** **(1.4)**	95.9 (15.1)	59.7	43.7
OK (absolute deviation)	20.9 (20.9)	479.5 (37.5)	182.5 (2.5)	**96.5** **(14.5)**	**55.3**	38.6
IDW (absolute deviation)	16.7 (16.7)	563.4 (46.4)	**183.6** **(1.4)**	95.5 (15.5)	63	44.4
SIC97 best absolute deviation	0	3	0	0	53.1	32
SIC97 median absolute deviation	15.5	54.5	4	12	63	44
SIC97 worst absolute deviation	413	271	26	28.5	99	70.6

Although the results from SVR are worse than OK, the method can be placed in the first half of the SIC97 results, which is a good result by itself, given the extreme simplicity of the preprocessing of training data (rescaling without deformation of the input co-ordinates). SVR are therefore promising methods for the development of automatic mapping tools. Some typical machine learning problems do however remain, like, for example, the fact that the best training error over various training techniques (train/test, cross-validations, etc…) does not correspond to the best generalisation error. The most probable reason is that machine learning algorithms fail to predict data when the training data are not fully representative of the studied phenomenon (which is the case in SIC97).

4. REFERENCES

B. E. Boser, I. M. Guyon, and V. Vapnik. A Training Algorithm for optimal Margin Classifiers. In *Fifth Annual Workshop on Computational Learning Theory*, Pittsburg, 1992. ACM.

C. J. C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2):1-47, 1998.

Cressie N. (1993). *Statistics for spatial data*. John Wiley & Sons Inc., Revised Edition, 900 p.

De Bollivier M., Dubois G., Maignan M. & Kanevsky M. (1997). Modified multilayer perceptron with local constraint: Artificial Neural Networks as an emerging method in spatial data analysis. *Nuclear Instruments and Methods in Physics Research A*, 389: 226-229.

Dubois G. (1998). Spatial Interpolation Comparison 1997: Foreword and introduction. *Special issue of the Journal of Geographic Information and Decision Analysis*, 2(2): 1-10.

C. Saunders, M. O. Stitson, J. Weston, L. Bottou, B. Schölkopf, A. Smola. Support Vector Machine Reference Manual, CSD-TR-98-03, Royal Holloway, University of London, Egham, UK, 1998. Software available at http://svm.dcs.rhbnc.ac.uk/.

A. Smola and B. Schölkopf. A Tutorial on Support Vector Regression. NeuroColt2 technical Reports Series, NC2-TR-1998-030, October 1998.

V. Vapnik. Statistical Learning Theory. John Wiley & Sons, 1998.

[1] http://publish.uwo.ca/~jmalczew/gida_4.htm