Geophysical Research Abstracts
Vol. 20, EGU2018-4405, 2018
EGU General Assembly 2018

Analysis of high dimensional environmental data using local fractality concept and machine learning

Mikhail Kanevski and Mohamed Laib
University of Lausanne, Institute of Earth Surface Dynamics, Lausanne, Switzerland (Mikhail.Kanevski@unil.ch)

In many cases, complex phenomena like environmental risks, natural hazards, and renewable resources assessments are considered in high dimensional feature spaces [1]. Therefore, an exploratory analysis and visualization of high dimensional data is an important topic in environmental data driven modelling. In the present research, new tools for the exploratory data analysis are proposed. They are based on a local, even pointwise, study of local Growth Curves (GC), usually used to estimate, so-called sandbox fractal dimension. The computation of GC is simple: each point (Xi) of a high dimensional data set is considered as a center and the number of points N(Xi, R) falling into the hypersphere with a radius R is counted. Then, the radius is changed and a local GC – N(Xi,R) is produced. The same procedure is performed for all points and a collection of GC is constructed [2]. These curves characterize how local density and local fractal dimension are changing in a space. In general, GC form a set of functional data, which can be studied using different techniques. In the present research, we use (1) local fractal dimension measure (slope of the curve in the log-log coordinates: log(N) versus log(R)), to study local intrinsic properties of the space and (2) machine learning clustering algorithms to reveal characteristic groups in curves.
Case studies considered deal with real and simulated data. Simulated data are used for better understanding of the tools. The real case study consists of a 13-dimensional input space used for the prediction of monthly wind speed [3]. The analysis includes estimation of global fractal dimension (intrinsic dimension of the manifold); calibration of the local calculations using global analysis (optimization of the parameters); mapping of local fractality; clustering of growth curves using different techniques; analysis and visualization of GC using functional boxplots. The first results are promising and help in deeper understanding of high dimensional modelling space and reveals interesting patterns.
The R code used for the research is available at GitHub server [4]. The code uses efficient R libraries to process big data and to perform parallel fast computing and analysis.

References

[1] Kanevski M., Pozdnoukhov A. and Timonin V. Machine Learning of Spatial Environmental Data. Theory, Applications and Software. EPFL Press, 2009.
[2] Kanevski M. and Pereira M. Local fractality: The Case of Forest Fires in Portugal. Physica A, vol. 479, 2017.
[3] Robert, S., Foresti, L., and Kanevski, M. (2013). Spatial prediction of monthly wind speeds in complex terrain with adaptive general regression neural networks. International Journal of Climatology, 33(7):1793-1804.
[4] Mohamed Laib, Mikhail Kanevski, "FractalTools", Fractality-based tools, (2018) URL: https://github.com/mlaib/FractalTools