

LSSR 2013: Abstract

Title: Random Forest for Environmental Data Mining

Authors: Michael Leuenberger, Carmen D. Vega Orozco and Mikhail Kanevski

Affiliation: Center for Research on Terrestrial Environment (CRET), Faculty of Geosciences and Environment, University of Lausanne, 1015 Lausanne, Switzerland.

Environmental phenomena don't lie in a simple two or three dimensional scale. Most of them (e.g. forest fires, landslide, permafrost, etc.) lie in a high dimensional space and require powerful nonlinear methods to explain or visualize their real "shape". Using anthropogenic and geographical feature data along with the random forest algorithm, this study aims at highlighting factors that most influence the occurrences of such phenomena and at identifying areas under risk.

Developed by L. Breiman and A. Cutler, the Random Forests (RF) algorithm provides an ensemble of classification and regression trees. By a pseudo-random variable selection for each split node, this method grows a variety of decision trees which return different results, and thus by a committee system, returns a value that has a better accuracy than other machine learning methods. This algorithm incorporates directly measurement of the variable's importance which is used to display the factors affecting the phenomenon under study. Dealing with this parameter, several models can be fitted, and thus, a prediction can be made throughout validity domains.

In the study of forest fires in Canton Ticino, comprehensive RF analysis was carried out in order to 1) understand the importance of anthropogenic and topographic features, 2) to assess the predictability of forest fires using environmental variables, and 3) to compare RF with other machine learning algorithms for this particular case study.

Risk maps were plotted by estimating the burned area taking into account the environmental variables and according to the most relevant selected variables.

Key words: Random Forest, classification and regression trees, environmental data mining, risk maps.

Acknowledgements

This work was partly supported by the SNFS Project No. 200021-140658, "Analysis and Modelling of Space-Time Patterns in Complex Regions".

References

- Breiman L., 2001, Random Forest, Machine Learning, vol. 45-1: 5-32.