

Machine learning from big data on land use

Hichem Omrani, Amin Tayyebi and Nicholas T. Longford

Abstract

Big data has received a lot of attention in the last few years. Dealing with a big dataset is a complex task, challenging analysts, data managers and their organisations. Supervised classification of land use is one of these challenges. In this application, the data is used for model calibration and standard methods are difficult to calibrate with very large data. Computing time is an associated problem. We propose a clustering trick to deal with this problem. In essence, the dataset is partitioned into smaller subsets which are then analysed separately. We evaluate the proposed approach by well-established criteria. The experiments are based on land-use data of the Grand Duchy of Luxembourg. They show good results.

O. Hichem (Corresponding author) • A. Tayyebi and N.T. Longford
Department of Urban Development and Mobility, LISER, Luxembourg
Email: hichem.omrani@liser.lu

T. Amin
Center for Conservation Biology, University of California-Riverside, USA
Email: amin.tayyebi@gmail.com

N.T. Longford
NTL, Statistics Research and Consulting, Barcelona, Spain
Email: sntlnick@sntl.co.uk