Small Area Estimation with Random Forests and the LASSO

Victoire Michal

McGill University

We consider random forests and LASSO methods for model-based small area estimation when the number of areas with sampled data is a small fraction of the total areas for which estimates are required. Abundant auxiliary information is available for the sampled areas, from the survey, and for all areas, from an exterior source, and the goal is to use auxiliary variables to predict the outcome of interest. We compare areal-level random forests and LASSO approaches to a frequentist forward variable selection approach and a Bayesian shrinkage method. Further, to measure the uncertainty of estimates obtained from random forests and the LASSO, we propose a modification of the split conformal procedure that relaxes the assumption of identically distributed data. We present results from simulation studies to assess the performance of the proposed split conformal procedure and to compare the four modelling approaches.

Co-authors: Jon Wakefield and Alexandra Schmidt