Distributed Bayesian Inference for Varying Coefficient Spatiotemporal Models

Cheng Li

National University of Singapore
E-mail: stalic@nus.edu.sg

Abstract: Bayesian varying coefficient models based on Gaussian processes are popular in many disciplines because they balance flexibility and interpretability. Markov chain Monte Carlo (MCMC) methods are available to fit these models, but they are inefficient even for moderately large data. Motivated by the task of modeling massive spatiotemporal data, we develop a divide-and-conquer Bayesian method for fitting spatiotemporal varying coefficient models based on multiple output Gaussian processes. Our method partitions the space-time tuples into a large number of overlapping subsets, obtains MCMC samples of parameters and predictions in parallel across the subsets, and combines the subset MCMC samples into an approximate full data posterior. By tuning the stochastic approximation in subset posteriors, we show theoretically that the combined posterior distribution can converge at an optimal rate towards the true underlying surface, and we provide guidance for choosing the number of subsets depending on the analytic properties of Gaussian processes. To improve the efficiency of MCMC sampling, we further develop a new data augmentation scheme based on parameter expansion. We demonstrate the excellent empirical performance of our method across diverse simulations and a real data application to the temperature and precipitation data in the U.S.A.