Forward variable selection for sparse ultra-high dimensional varying coefficient models

Ming-Yen Cheng*
National Taiwan University, Taipei, Taiwan - cheng@math.ntu.edu.tw

Toshio Honda
Hitotsubashi University, Tokyo, Japan - t.honda@r.hit-u.ac.jp

Jin-TIng Zhang
National University of Singapore, Singapore - stazjt@nus.edu.sg

Varying coefficient models have numerous applications in a wide scope of scientific areas. While enjoying nice interpretability, they also allow flexibility in modeling dynamic impacts of the covariates. But, in the new era of big data, it is challenging to select the relevant variables when there are a large number of candidates. Recently several work are focused on this important problem based on sparsity assumptions; they are subject to some limitations, however. We introduce an appealing forward variable selection procedure. It selects important variables sequentially according to a sum of squares criterion, and it employs an EBIC- or BIC-based stopping rule. Clearly it is simple to implement and fast to compute, and it possesses many other desirable properties from both theoretical and numerical viewpoints. We establish rigorous selection consistency results when either EBIC or BIC is used as the stopping criterion, under some mild regularity conditions. Notably, unlike existing methods, an extra screening step is not required to ensure selection consistency. Even if the regularity conditions fail to hold, our procedure is still useful as an effective screening procedure in a less restrictive setup. We carried out simulation and empirical studies to show the efficacy and usefulness of our procedure.

Keywords: B-spline; EBIC; independence screening; marginal model; semi-varying coefficient models; sub-Gaussion error; structure identification.