Bayesian Ensemble Methods for Survival Prediction in Gene Expression Data

Kim-Anh Do[*], U.T. M.D. Anderson Cancer Center

Abstract

We propose a Bayesian ensemble method for survival prediction in high-dimensional gene expression data. We specify a fully Bayesian hierarchical approach based on an ensemble "sum-of-trees" model and illustrate our method using three popular survival models. Our nonparametric method incorporates both additive and interaction effects between genes, which results in high predictive accuracy compared to other methods. In addition, our method provides model-free variable selection of important prognostic markers based on controlling the false discovery rates; thus providing a unified procedure to select relevant genes and predict survivor functions. We assess the performance of our method on several simulated and real microarray data sets. We show that our method selects genes potentially related to the development of the disease as well as yields predictive performance that is very competitive to many other existing methods. This is joint work with Vinicius Bonato, Veera Baladandayuthapani, Bradley Broom, Ken Aldape, and Eric Sulman.

Keywords: Survival prediction; Bayesian; Ensemble method; Gene expression; High-dimensional data.

---

[*] Presenting author