



Bayesian Large-scale Multiple Testing for Dependent Data

Xia Wang*

University of Cincinnati, Cincinnati, USA – xia.wang@uc.edu

Ali Shojaie

University of Washington, Seattle, USA – ashojaie@uw.edu

Jian (Frank) Zou

Worcester Polytechnic Institute, Worcester, USA – jzou@wpi.edu

The study aims to construct optimal and flexible multiple hypotheses testing procedure for temporally or spatially dependent data. Ignoring dependence among hypotheses tests may lead to loss of efficiency and bias in decision. Misspecification in the alternative distributions can also cause problem. In this study, we apply Bayesian method to overcome these two challenges in large-scale multiple testing. The observed data is assumed to be generated from an underlying two-state hidden Markov model (HMM). The testing algorithm using Bayesian methods optimizes the false negative rate (FNR) while controlling the false discovery rate (FDR). The Dirichlet prior on the alternative distributions overcomes the potential pitfalls in distribution misspecification. We show the advantages of the proposed model by comparison to a group of other methods in both simulation studies and real data applications.

Keywords: false discovery rate; false negative rate; Hidden Markov Model; nonparametric Bayesian.