

A Tree-Based Scan Statistic for Database Studies with Time-to-Event Outcomes

Massimiliano Russo

Abstract:

Tree-based scan statistics (TBSSs) are popular machine learning methods to conduct disproportionality analyses in database studies. They simultaneously scan for thousands of hierarchically related outcomes to detect potential signals of harm from drugs and vaccines while controlling for multiplicity. In pharmacoepidemiology, TBSSs have been extensively used to mine claims data and detect potential adverse events of drugs. However, current implementations of TBSSs do not allow for comparative safety evaluation with time-to-event outcomes, which are available in most database studies. Explicitly accounting for person time in the data analysis can potentially improve the power to detect signals compared to methods that only use the number of events. We propose and compare three novel TBSS methods that analyze time-to-event data. The first assumes proportional Hazard Rates (HRs) for each node of the hierarchy and uses a permutation scheme for inference. The second builds on exponential survival models for the terminal nodes of the hierarchy, implying a constant HR for each node. It uses a parametric bootstrap for inference. The third approach uses robust asymptotic approximations of the HRs to build an approximate parametric bootstrap. We compare the proposed methods with standard TBSSs in various simulation scenarios and a real database study that compares exposure to newly prescribed canagliflozin with a dipeptidyl peptidase 4 inhibitor in adults affected by type 2 diabetes