

Learning Individualized Treatment Rules from Electronic Health Records

Yuanjia Wang

Columbia University

E-mail: yw2016@cumc.columbia.edu

Abstract: To address substantial heterogeneity in patient response to treatment of chronic disorders and achieve the promise of precision medicine, individualized treatment rules (ITRs) are estimated to tailor treatments according to patient-specific characteristics. Randomized controlled trials (RCTs) provide gold standard data for learning ITRs not subject to confounding bias. However, RCTs are often conducted under stringent inclusion/exclusion criteria, and participants in RCTs may not reflect the general patient population. Thus, ITRs learned from RCTs lack generalizability to the broader real world patient population. Real world databases such as electronic health records (EHRs) provide new resources as complements to RCTs to facilitate evidence-based research for personalized medicine. However, to ensure the validity of ITRs learned from EHRs, a number of challenges including confounding bias and selection bias must be addressed. In this work, we propose a matching-based machine learning method to estimate optimal individualized treatment rules from EHRs using interpretable features extracted from EHR documentation of medications and ICD diagnoses codes. We use a latent Dirichlet allocation (LDA) model to extract latent topics and weights as features for learning ITRs. Our method achieves confounding reduction in observational studies through matching treated and untreated individuals and improves treatment optimization by augmenting feature space with clinically meaningful LDA-based features. We apply the method to EHR data collected at New York Presbyterian Hospital clinical data warehouse in studying optimal second-line treatment for type 2 diabetes (T2D) patients. We use cross validation to show that ITRs outperforms uniform treatment strategies (i.e., assigning same treatment to all individuals), and including topic modeling features leads to more reduction of post-treatment complications.