ABSTRACT:
Dynamic treatment regimes (DTRs) are sequential decision rules for individual patients that can adapt over time to an evolving illness. The ultimate goal is to accommodate heterogeneity among patients and find the DTR which will produce the best long term outcome if implemented. In this work, we introduce two new statistical learning methods for estimating the optimal DTR, termed backward outcome weighted learning (BOWL), and simultaneous outcome weighted learning (SOWL). These approaches convert individualized treatment selection into an either sequential or simultaneous classification problem, and thus can be widely applicable by modifying existing machine learning techniques. Furthermore, the proposed methods directly maximize the expected long-term outcomes and are thus fundamentally different from popular regression-based methods, for example Q-learning, which attempt such maximization only indirectly. We prove that the resulting rules are consistent, and provide finite sample bounds for the errors using the estimated rules. Simulation results suggest the proposed methods produce superior DTRs compared with Q-learning especially in small samples. We illustrate the methods using data from a clinical trial for smoking cessation.