

Optimal Dynamic Treatment Regime by Reinforcement Learning in Clinical Medicine

Mina Song and David Han
University of Texas at San
Antonio

Abstract

Precision medicine allows personalized treatment regime for patients with distinct clinical history and characteristics. Dynamic treatment regime implements a reinforcement learning algorithm to produce the optimal personalized treatment regime in clinical medicine. The reinforcement learning method is applicable when an agent takes action in response to the changing environment over time. Q-learning is one of the popular methods to develop the optimal dynamic treatment regime by fitting linear outcome models in a recursive fashion. Despite its ease of implementation and interpretation for domain experts, Q-learning has a certain limitation due to the risk of misspecification of the linear outcome model. Recently, more robust algorithms to the model misspecification have been developed. For example, the inverse probability weighted estimator overcomes the aforementioned problem by using a nonparametric model with different weights assigned to the observed outcomes for estimating the mean outcome. On the other hand, the augmented inverse probability weighted estimator combines information from both the propensity model and the mean outcome model. The current statistical methods for producing the optimal dynamic treatment regime however allow only a binary action

space. In clinical practice, some combinations of treatment regime are required, giving rise to a multi-dimensional action space. This study develops and demonstrates a practical way to accommodate a multi-level action space, utilizing currently available computational methods for the practice of precision medicine.

Keywords: dynamic treatment regime, precision medicine, Q-learning algorithm, reinforcement learning