Abstract

This research addresses causal inference and modeling over event streams where data have high throughput and are unbounded. The availability of large amount of data along with the high data throughput present several new challenges related to causal modeling, such as the need for fast causal inference operations while ensuring consistent and valid results. There is no existing work specifically for such a streaming environment. We meet the challenges by introducing a time-centric causal inference strategy that leverages temporal precedence information to decrease the number of conditional independence tests required to establish the dependencies between the variables in a causal network. Dependency and temporal precedence of cause over effect are the two properties of a causal relationship. We also present the Temporal Network Inference algorithm to model the temporal precedence relations into a temporal network. Then, we propose the Fast Causal Network Inference algorithm for faster learning of causal network using the temporal network. Experiments using synthetic and real datasets demonstrate the efficacy of the proposed algorithms.