Abstract

Explaining Actual Causation via Reasoning about Actions and Change
Emily C. LeBlanc
Santiago Ontañón, Ph.D. and Marcello Balduccini, Ph.D.

The goal of this research is to investigate and demonstrate the suitability of action languages and answer set programming (ASP) to design and realize a novel framework for explaining *actual causation*. Actual causation is a broad term that encompasses all possible antecedents that have played a meaningful role in producing a consequence. Attempts to characterize reasoning about actual causation have largely pursued counterfactual analysis of a scenario, inspired by the intuition that if X caused Y, then not Y if not X. However, it has been widely documented that the counterfactual criteria alone is problematic and fails to recognize causation in a number of straightforward cases. Departing from a counterfactual reasoning approach, our framework favors reasoning about the underlying causal mechanisms of the scenario itself in order to explain how an outcome of interest came to be. The framework leverages techniques from Reasoning about Actions and Change to support reasoning about domains that change over time in response to a sequence of events. The action language $\mathcal{AL}$ enables us to represent a scenario in terms of the evolution of the state of the world over the course of the scenario's events. $\mathcal{AL}$ lends itself naturally to an automated translation in Answer Set Programming (ASP), using which, reasoning tasks of considerable complexity can be specified and executed. In this dissertation, we present a theoretical framework for reasoning about actual causation and demonstrate that the framework enables reasoning about traditionally challenging examples of actual cause. We also present a sound and complete implementation of the theoretical framework in ASP, along with a collection of empirical studies that evaluate and analyze the framework's performance on a number of novel and challenging problems.