Causal discovery with Point of Sales data

Peter Gmeiner

Global Data Science, GfK SE, Germany.

Abstract

GfK owns the world's largest retail panel within the tech and durable good industries. The panel consists of weekly Point of Sales (PoS) data, such as price and sales units data at store level. From PoS data and other data, GfK derives insights and indicators to generate recommendations with regards to e.g. pricing, distribution or assortment optimization of tech and durable good products. By combining PoS data and business domain knowledge, we show how causal discovery can be done by applying the method of invariant causal prediction (ICP).

Causal discovery, in essence, means to learn the actual cause and effect relations between the involved variables from data. After finding such a causal structure, one can try to further specify the function classes between those identified cause-effect pairs. Such a model could then be used to predict under intervention (predict when the underlying data generating mechanism changes) and to optimize and calculate counterfactual effects, given current and past data. In our development, we combine recent achievements in causal discovery research with PoS data structure and business domain knowledge (in the form of business rules).

The key delivery of this presentation is to show fundamental differences between a causal model and a machine learning model. We further explain the advantages of combining a causal model with a machine learning model and why causal information is key to provide explainable prescriptive analytics. Furthermore, we demonstrate how to apply ICP (for sequential data) to context-specific PoS data to achieve improved models for sales unit predictions. As a result, we obtain a model for sales units that is on the one hand derived from observed data and on the other hand driven by business knowledge. Such a refined prediction model could then be used to stabilize and support other machine learning models that can be used for generating prescriptive analytics.

Keywords: Causal Model; Causal Discovery; Machine Learning; PoS.