Enumeration of Most General Why-Not Explanations

Evgeny Sherkhonov

University of Amsterdam

Abstract

A large number of databases are a result of extraction, curation and integration of different data sources. Consequently, it is essential to provide data consumers with mechanisms and tools that allow them to gain understanding of the data. One such mechanism is providing meaningful explanations to why data is present or missing from the output of a query.

A framework for explaining why data is *missing* from the query output has recently been proposed in [2]. It exploits ontologies in order to provide *high-level* and meaningful explanations (formalized in the notion of a *most general explanation* (MGE) w.r.t. an ontology in [2]). The central algorithmic problem in [2] is to generate *one* most general explanation, if it exists. However, in practice generating just one explanation might not be satisfactory. Preferably, a data consumer should have the ability to explore the entire space of possible explanations.

This talk is about the problem of enumerating (i.e., generating) *all* most general explanations. Typically, in a setting with a large (exponential) number of output solutions, "efficient" enumeration algorithms are those that run with *polynomial delay.* We show that enumerating all MGEs is at least as hard as a prominent problem in complexity theory – dualization of a monotone Boolean function [1], which is still unknown to be solved in polynomial time in the size of both the input and the output. In particular, existence of an "efficient" algorithm enumerating all MGEs would settle this open problem. However, we provide an enumeration algorithm that runs with polynomial delay for a large class of inputs.

References

[1] Thomas Eiter, Kazuhisa Makino, and Georg Gottlob. Computational aspects of monotone dualization: A brief survey. *Discrete Applied Mathematics*, 156(11):2035–2049, 2008.

[2] Balder ten Cate, Cristina Civili, Evgeny Sherkhonov, and Wang-Chiew Tan. High-level why-not explanations using ontologies. In *PODS 2015*, 2015.