Most Probable Explanations for Probabilistic Database Queries (Extended Abstract)⋆

İsmail İlkan Ceylan[1], Stefan Borgwardt[1], and Thomas Lukasiewicz[2]

[1] Faculty of Computer Science, Technische Universität Dresden, Germany
`firstname.lastname@tu-dresden.de`
[2] Department of Computer Science, University of Oxford, UK
`thomas.lukasiewicz@cs.ox.ac.uk`

Probabilistic databases (PDBs) have been widely studied in the literature, as they form the foundations of large-scale probabilistic knowledge bases like NELL and Google's Knowledge Vault. In particular, *probabilistic query evaluation* has been investigated intensively as a central inference mechanism. However, despite its power, query evaluation alone cannot extract all the relevant information expressed in PDBs. Inspired by the *maximal posterior probability* computations in Probabilistic Graphical Models (PGMs) [3], we investigate the problem of finding *most probable explanations* for database queries to exploit the potential of such large databases to their full extent. The *most probable database* [2] is the (classical) database with the largest probability that satisfies a given query. Intuitively, the query defines constraints on the data, and the goal is to find the most probable database that satisfies these constraints. We also introduce a more intricate notion, called *most probable hypothesis*, which is only a *partial* database satisfying the query. The most probable hypothesis contains only facts that contribute to the satisfaction of the query, which allows to more precisely pinpoint the most probable explanations for the query. We study the complexity of the corresponding decision problems for a variety of database query languages. In particular, we also consider *ontology-mediated queries (OMQs)*, which enrich UCQs with the power of Datalog$^{\pm}$ ontologies. They allow us to query PDBs in a more advanced manner [1]. We show that the complexity of these problems changes significantly with the ontology languages and the complexity-theoretic assumptions. Our results provide tight complexity bounds for a multitude of Datalog$^{\pm}$ languages (which cover some Horn Description Logics).

References

1. Borgwardt, S., Ceylan, İ.İ., Lukasiewicz, T.: Ontology-mediated queries for probabilistic databases. In: Proc. AAAI (2017)
2. Gribkoff, E., Van den Broeck, G., Suciu, D.: The most probable database problem. In: Proc. BUDA (2014)
3. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. MIT Press (2009)