Novel Techniques in Datalog Optimization (Abstract)

Reinhard Pichler

TU Wien, Institut für Logic and Computation, Favoritenstraße 9, A-1040 Wien, Austria

Abstract

Datalog has been specifically designed for processing recursive queries. As such it seems the perfect fit for today's data analytics applications, which typically require some kind of iteration or recursion. However, basic analytical tasks such as computing shortest paths, betweenness centrality, or connected components in a graph as well as solving optimization problems by gradient descent or other methods crucially depend on aggregation. For Datalog, this poses new challenges due to the loss of the usual monotonicity properties in case of aggregation.

In this talk, I will present recent work in which

- we have studied a generalization of Datalog that allows for recursive computations over general semirings (with classical Datalog corresponding to the special case of the Boolean semiring),

- we have analyzed the convergence of this generalization of Datalog, and

- we have introduced a powerful new optimization technique that covers known optimizations such as magic-set rewriting as well as new ones.

This talk is mainly based on the following papers:

1. Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu, Yisu Remy Wang: Convergence of Datalog over (Pre-) Semirings. PODS 2022.

2. Yisu Remy Wang, Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu: Optimizing Recursive Queries with Progam Synthesis. SIGMOD Conference 2022.