Capturing Knowledge Patterns in Entity Resolution

Dr. Qing Wang
Australian National University

Entity resolution is concerned with deciding whether two representations of entities refer to the same real-world object. It is one of the major impediments affecting data quality provided by information systems. The difficulty of this problem has been widely acknowledged by various research communities and industry practitioners. State-of-the-art approaches to entity resolution mostly favor similarity-based methods. In this talk, I will present a simple yet expressive framework that can support knowledge-based entity resolution. Knowledge patterns, as the building blocks of the framework, have the capability of capturing knowledge about different entities at an arbitrary level of abstraction. From a logical point of view, the expressive power of the framework is equivalent to a fragment of first-order logic including conjunction, disjunction and a certain form of negation. Thus, the question of how efficiently redundancy can be eliminated from knowledge patterns naturally arises. I will discuss the containment problem of knowledge patterns and introduce a decision procedure for determining containment and equivalence between knowledge patterns. It turns out that the containment problem of knowledge patterns is NP-complete w.r.t. expression complexity but in PTIME w.r.t. data complexity. Finally, I will introduce a mechanism of optimizing knowledge patterns, which can lead to more efficient knowledge management and thus improve accuracy of entity resolution over time.