Symmetry in Probabilistic Databases

Guy Van den Broeck

Computer Science Department University of California, Los Angeles guyvdb@cs.ucla.edu

Abstract. Researchers in databases, AI, and machine learning, have all proposed representations of probability distributions over relational databases (possible worlds). In a tuple-independent probabilistic database, the possible worlds all have distinct probabilities, because the tuple probabilities are distinct. In AI and machine learning, however, one typically learns highly symmetric distributions, where large numbers of symmetric databases get assigned identical probability. This symmetry helps with generalizing from data. In this talk I discuss what happens to standard database notions of data and combined complexity when considering AIstyle symmetric probabilistic databases. The question proves to be a fertile ground for database theory, with interesting connections to counting complexity and 0-1 laws.