

Combining Discrete Algorithmic and Probabilistic Approaches in Data Mining

Heikki Mannila

Nokia Research Center and Helsinki University of Technology

Abstract. Data mining research has approached the problems of analyzing large data sets in two ways. Simplifying a lot, the approaches can be characterized as follows. The database approach has concentrated on figuring out what types of summaries can be computed fast, and then finding ways of using those summaries. The model-based approach has focused on first finding useful model classes and then fast ways of fitting those models. In this talk I discuss some examples of both and describe some recent developments which try to combine the two approaches.