

COMPUTER SCIENCE
SEMINAR

MCDB: The Monte Carlo Database System

Chris Jermaine
Rice University

Abstract: Analysts working with large data sets often use statistical models to "guess" at unknown, inaccurate, or missing information associated with the data stored in a database. For example, an analyst for a manufacturer may wish to know, "What would my profits have been if I'd increased my margins by 5% naturally depends upon the extent to which the higher prices would have affected each customer's demand, which is undoubtedly guessed via the application of some statistical model. In this talk, I'll describe MCDB, which is a prototype database system that is designed for just such a scenario. MCDB allows an analyst to attach arbitrary stochastic models to the database data in order to "guess" the values for unknown or inaccurate data, such as each customer's unseen demand function. These stochastic models are used to produce multiple possible database instances in Monte Carlo fashion (a.k.a. "possible worlds"), and the underlying database query is run over each instance. In this way, fine-grained stochastic models become first-class citizens within the database. MCDB can be used for diverse tasks such as risk assessment and large-scale, data-driven simulation.

Friday, February 11, 2011, 3:00 pm
Mathematics and Science Center: W301

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY