"AutoBayes: Automatic Synthesis of Data Analysis Programs"

by
Bernd Fischer
Automated Software Engineering Group
USRA/RIACS, NASA Ames Research Center

Friday, January 23, 2004
Baskin Engineering Room 330
12:30pm-2:00pm

Abstract:

AutoBayes is a fully automatic synthesis system for statistical data analysis programs. Its input is a concise description of a data analysis problem in the form of a statistical model; its output is optimized and fully documented C/C++ code which can be linked dynamically into the Matlab and Octave environments. AutoBayes synthesizes code in a schema-based process. A schema is a parameterized code template with associated semantic constraints which define and restrict the template's applicability. The schema parameters are instantiated in a problem-specific way during synthesis as AutoBayes checks the constraints against the original model or, recursively, against emerging subproblems. AutoBayes' schema library contains problem decomposition operators (which are justified by theorems in a formal logic in the domain of Bayesian networks) as well as machine learning algorithms (e.g., EM, k-Means) and numeric optimization methods (e.g., Nelder-Mead simplex, conjugate gradient). AutoBayes augments this schema-based approach by symbolic computation to derive closed-form solutions whenever possible. This is a major advantage over other statistical data analysis systems which use numerical approximations even in cases where closed-form solutions exist.

AutoBayes has been applied to a number of textbook and application problems, including clustering (using EM), changepoint detection, and software reliability estimation. In the talk, I will discuss some of the schemas and the derivation process in more detail and demonstrate the system "live".

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Anyone needing special arrangements to accommodate a disability is encouraged to call Ma Xiong, at Baskin School of Engineering, (831) 459-5745.