

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
SPECIAL SEMINAR

Host: Trevor Hastie

2:15 p.m., Friday, November 5, 1999
Sequoia Hall Rm. 200

Edwin P. D. Pednault
Mathematical Sciences Department
IBM T. J. Watson Research Center
Yorktown Heights, New York

A Machine Learning Approach to Insurance Risk Modeling

The IBM DecisionEdge for Insurance product announced in November 1998 includes software for constructing tree-based models of insurance risk for personal lines property and casualty insurance. The algorithms that are employed differ from other tree-based modeling techniques, such as CART, CHAID, C4.5, SPRINT, etc., in that insurance risks are modeled using joint Poisson/log-normal statistical models in the leaves of the trees that are constructed, and actuarial credibility constraints (i.e., confidence bounds on parameter estimates) are used to guide the tree building process. The joint Poisson/log-normal models take into account insurance-specific concepts such as earned exposure (i.e., the amount of time that a policy is actually in force during a given time interval), and the status of each claim (i.e., whether a claim is closed and paid versus open with final action pending). Neither of these concepts are accounted for in standard tree-based algorithms. Our method is thus tailored to the specific statistical properties of insurance data, and it has so far yielded risk models that are more predictive of actual risks than those obtained using conventional tree-based algorithms.

Edwin Pednault joined IBM in 1996, where he is a Research Staff Member in the Data Abstraction Research Group. From 1986 to 1995, he was a Member of Technical Staff at AT&T Bell Laboratories. Edwin received a Ph.D. in electrical engineering from Stanford University in 1987, a M.S. in computer science from Stanford in 1981, and a B.Eng. in electrical engineering from McGill University in 1979. His current research interests center on statistical learning theory and its application to automated predictive modeling.