

Computation and **I**nformatics in **B**iology and **M**edicine
Training Program

Mark Craven

Assistant Professor
Department of Biostatistics and Medical Informatics
and Computer Sciences Department
University of Wisconsin-Madison

***Data Mining in Molecular Biology Domains:
Results of KDD Cup 2002***

KDD Cup is an annual data mining competition held in conjunction with the International Conference on Knowledge Discovery and Data Mining (KDD). Along with Alex Yeh of MITRE Inc., I served as co-chair of KDD Cup this year. The competition included two tasks that involved data mining in molecular biology domains. The first task focused on constructing models that can assist genome annotators by automatically extracting information from scientific articles. The second task focused on learning models that characterize the behavior of individual genes in a high-throughput, experimental setting. The data used in the second task was generated by Guang Yao and Prof. Chris Bradfield of UW's McArdle Lab. I will describe the two tasks, and summarize the results and some of the lessons learned in KDD Cup this year.

**Tuesday, November 5th
4:00 p.m.**

Room 1111
Biotechnology Center/Genetics
425 Henry Mall