



Bioinformatics Seminar

Speaker: Jadwiga Bienkowska, Serono

Title: GOTrees: Predicting GO Associations From Protein Domain
Composition Using Decision Trees

Date: Monday, 29 November 2004

Time & Location:

Refreshments: 11 am in the Theory of Computation Lab at MIT's Building 32, Stata Center Room G-575

Talk: 11:30 am the Theory of Computation Lab at MIT's Building 32, Stata Center, Room G-575

URL: <http://www-math.mit.edu/compbiosem/>

Abstract:

The Gene Ontology (GO) offers a comprehensive and standardized way to describe a protein's biological role. Proteins are annotated with GO terms based on direct or indirect experimental evidence. Term assignments are also inferred from homology and literature mining. Regardless of the type of evidence used, GO assignments are manually curated or electronic. Unfortunately, manual curation cannot keep pace with the data, available from publications and various large experimental datasets. Automated literature-based annotation methods have been developed in order to speed up the annotation. However, they only apply to proteins that have been experimentally investigated or have close homologs with sufficient and consistent annotation. One of the homology-based electronic methods for GO annotation is provided by the InterPro database. The InterPro2GO/PFAM2GO associates individual protein domains with GO terms and thus can be used to annotate the less studied proteins. However, protein classification via a single functional domain demands stringency to avoid large number of false positives. This work broadens the basic approach. We model proteins via their entire functional domain content and train individual decision tree classifiers for each GO term using known protein assignments. We demonstrate that our approach is sensitive, specific and precise, as well as fairly robust to sparse data. We have found that our method is more sensitive when compared to the InterPro2GO performance and suffers only some precision decrease. In comparison to the InterPro2GO we have improved the sensitivity by 22%, 27% and 50% for Molecular Function, Biological Process and Cellular GO terms respectively.

The seminar is co-hosted by Professor Peter Clote of Boston College's Biology and Computer Science Departments and MIT Professor of Applied Math Bonnie Berger. Professor Berger is also affiliated with CSAIL & HST.

Massachusetts Institute
of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

For General Questions, please contact kvdickey@mit.edu