

Computational Biology and the Analysis of Heterogeneous Data

Friday, December 9, and Saturday, December 10, 2005

Christina Leslie, *Columbia University*

William Stafford Noble, *University of Washington*

Gunnar Rätsch, *Friedrich Miescher Laboratory of the Max Planck Society*

Simon Kasif, *Boston University*

Koji Tsuda, *AIST Computational Biology Research Center*

<http://www.fml.tuebingen.mpg.de/nipscompbio>

The field of computational biology has seen dramatic growth over the past few years, both in terms of available data, scientific questions and challenges for learning and inference. These new types of scientific and clinical problems require the development of novel supervised and unsupervised learning approaches.

In particular, the field is characterized by a diversity of heterogeneous data. The human genome sequence is accompanied by real-valued gene expression data, functional annotation of genes, genotyping information, a graph of interacting proteins, a set of equations describing the dynamics of a system, localization of proteins in a cell, a phylogenetic tree relating species, natural language text in the form of papers describing experiments, partial models that provide priors, and numerous other data sources.

The goal of this workshop is to present emerging problems and machine learning techniques in computational biology, with a particular emphasis on methods for computational learning from heterogeneous data. The workshop includes invited and submitted talks from experts in the fields of biology, bioinformatics and machine learning. The topics range from case studies of particular biological problems to novel learning approaches in computational biology.

Program Committee

- Pierre Baldi, UC Irvine
- Kristin Bennett, Rensselaer Polytechnic Institute
- Nello Cristianini, UC Davis
- Gal Chechik, Stanford University
- Eleazar Eskin, UC San Diego
- Nir Friedman, Hebrew University and Harvard
- Alexander Hartemink, Duke University
- Michael I. Jordan, UC Berkeley
- Tommi Jaakkola, Computer Science and Artificial Intelligence Laboratory, MIT
- Michal Linial, The Hebrew University of Jerusalem
- Klaus-Robert Müller, Fraunhofer FIRST
- Alexander Schliep, Max Planck Institute for Molecular Genetics
- Bernhard Schölkopf, Max Planck Institute for Biological Cybernetics
- Eran Segal, Stanford University
- Jean-Philippe Vert, Ecole des Mines de Paris

Computational Biology and the Analysis of Heterogeneous Data

Friday December 9, 2005

Organizers: Christina Leslie & William S. Noble & Gunnar Rätsch & Simon Kasif & Koji Tsuda

Morning session: 7:30am–10:45am

- 7:30am *Welcome*
- 7:35am **Machine learning for HIV vaccine design**, *David Heckerman (Microsoft Research)*
- 8:15am **Separation of overlapping subpopulations by mutual information**,
Sean O'Rourke, Gal Chechik and Eleazar Eskin
- 8:40am **Analysis of tuberculosis patient trends using *Mycobacterium tuberculosis* complex strain genotype families**,
Inna Vitol, Jeff Driscoll, Natalia Kurepina, and Kristin Bennett
- 9:05am *Coffee break*
- 9:25am **The Biozon system for complex analysis of heterogeneous interrelated biological data and discovery of emergent structures**, *Golan Yona (Cornell University)*
- 10:05am **Using hierarchies to improve classifier predictions in genomics**,
Zafer Barutcuoglu, Robert E. Schapire and Olga G. Troyanskaya
- 10:30am **A novel algorithm for maximum likelihood haplotyping on general pedigrees**,
Kees Albers and Bert Kappen

Afternoon session: 3:30pm–6:30pm

- 3:30pm **Revising our view of the mammalian genome using mega microarrays and probabilistic inference**, *Brendan Frey (University of Toronto)*
- 4:10pm **Mixture model estimation with constraints: analysis of time-course gene expression with heterogeneous data**, *Ivan G. Costa and Alexander Schliep*
- 4:25pm **Untangling biological networks using graph priors**,
Quaid Morris, Vladimir Jojic and Brendan J. Frey
- 4:50pm *Coffee break*
- 5:10pm **Comparative sequence analysis of gene regulation: examples from alternative splicing**, *Uwe Ohler (Duke University)*
- 5:50pm **Genome-wide discovery of modulators of transcriptional interactions in human B lymphocytes**, *Kai Wang, Ilya Nemenman, Nilanjana Banerjee, and Andrea Califano*
- 6:05pm **Accurate prediction of alternative splicing events**,
Gunnar Rätsch, Sören Sonnenburg, Cheng Soon Ong, and Bernhard Schölkopf

Computational Biology and the Analysis of Heterogeneous Data

Saturday December 10, 2005

Organizers: Christina Leslie & William S. Noble & Gunnar Rätsch & Simon Kasif & Koji Tsuda

Morning session: 7:30am–10:30am

- 7:35am **Similarity and kernels in chemoinformatics and drug discovery,**
Pierre Baldi (University of California at Irvine)
- 8:15am **Protein function prediction via faster graph kernels,**
Karsten M. Borgwardt, S.V.N. Vishwanathan, Nic Schraudolph, and Hans-Peter Kriegel
- 8:40am **Beyond pairwise classification and clustering using hypergraphs,**
Dengyong Zhou, Jiayuan Huang and Bernhard Schölkopf
- 9:05am *Coffee break*
- 9:25am **A game-theoretic approach to protein-DNA binding,** *Luis Ortiz (MIT)*
- 10:05am **Information theoretic genotype-phenotype association reveals intrinsic modularity in genetic networks,**
Noam Slonim, Olivier Elemento and Saeed Tavazoie

Afternoon session: 3:30pm–6:30pm

- 3:30pm **Explaining vertebrate gene expression,** *Tim Hughes (University of Toronto)*
- 4:10pm **A mixture of experts approach for protein-protein interaction prediction,**
Yanjun Qi, Judith Klein-Seetharaman and Ziv Bar-Joseph
- 4:25pm **Discovery of biological networks from diverse functional genomic data,**
Chad L. Myers and Olga G. Troyanskaya
- 4:50pm *Coffee break*
- 5:10pm **The abuse of the mode and an ensemble alternative,**
Charles Lawrence (Brown University)
- 5:50pm **How reproducible are protein expression estimates?,**
Bernd Fischer, Volker Roth and Joachim M. Buhmann
- 6:05pm **Filling missing components in yeast metabolic pathways using heterogeneous data,** *Gal Chechik, Aviv Regev and Daphne Koller*
- 6:30pm *Closing remarks*