

Guest Editor's Introduction to the Special Issue: Machine Learning for Bioinformatics—Part 2

Charles X. Ling, William Stafford Noble, and Qiang Yang

THIS issue of the *IEEE/ACM Transactions on Computational Biology and Bioinformatics* comprises the second half of our special issue on Machine Learning for Bioinformatics. In total, more than 50 papers were submitted to the special issue, of which 13 were accepted. The six papers included in this issue cover a broad range of topics, from mass spectrometry, to gene expression, to gene networks, etc:

- The paper "Analyzing Gene Expression Time-Courses" by Alexander Schliep, Ivan G. Costa, Christine Steinhoff, and Alexander Schönhuth proposes a robust and efficient method to analyze gene expression time-course data using a mixture of hidden Markov models. Experiments on biological and simulated data show its suitability and effectiveness compared to previous methods.
- The paper "Combining Sequence and Time Series Expression Data to Learn Transcriptional Modules" by Anshul Kundaje, Manuel Middendorf, Feng Gao, Chris Wiggins, and Christina Leslie presents a generative probabilistic model for combining regulatory sequence and time series expression data for learning transcriptional modules in organisms such as yeast. Experiments show that the method finds modules and associated binding site motifs with a functionally coherent and biologically plausible set of regulated genes.
- The paper "Associative Clustering for Exploring Dependencies between Functional Genomics Data Sets" by Samuel Kaski, Janne Nikkilä, Janne Sinkkonen, Leio Lahti, Juha E.A. Knuutila, and Christophe Roos introduces a new method, called Associate Clustering (AC), that discovers statistical dependencies between paired, co-occurring patterns of multiple data sets. The method is applied and validated in two functional genomics studies.
- The paper "Predicting Molecular Formulas of Fragment Ions with Isotope Patterns in Tandem Mass Spectra" by Jingfen Zhang, Wen Gao, Jinjin Cai, Simin He, Rong Zeng, and Runsheng Chen proposes

a new approach, called Fragment Ion Formula Prediction (FFP), to predict elemental component formulas of fragment ions based on isotope patterns in tandem mass spectra. The method consists of multiple steps to search and generate candidate formulas and to filter out invalid and improbable formulas.

- The paper "Discovering Gene Networks with a Neural-Genetic Hybrid" by Edward Keedwell and Ajit Narayanan describes a neural-genetic hybrid model to discover gene networks from temporal gene expression data. Experiments on real-world gene-expression data sets confirm the effectiveness of the proposed method.
- In the paper "The Applicability of Recurrent Neural Networks for Biological Sequence Analysis," John Hawkins and Michael Bodén empirically demonstrate that recurrent neural networks have an architectural bias that favors their application to sequence analysis tasks. Compared to standard feed-forward networks, the recurrent architecture lends the network increased sensitivity to inexact patterns of a sequential nature. Hawkins and Boden demonstrate this bias with a case study involving the prediction of protein subcellular localization.

Both the quality and quantity of submissions to this special issue illustrate the important role that machine learning plays in current bioinformatics research. As our analytical methods improve and the quantity and diversity of biological data grow, we believe that machine learning's influence in this domain will increase as well.

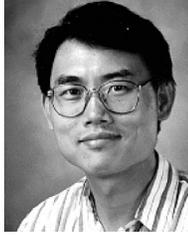
- C.X. Ling is with the Department of Computer Science, The University of Western Ontario, Ontario, London, ON N6A 5B7 Canada. E-mail: cling@csd.uwo.ca.
- W.S. Noble is with the Department of Genome Sciences, University of Washington, Box 357730, 1705 NE Pacific St #K-357, Seattle WA 98195-7730. E-mail: noble@gs.washington.edu.
- Q. Yang is with the Department of Computer Science, Hong Kong UST, Hong Kong. E-mail: qyang@cs.ust.hk.

For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org.



Charles X. Ling received the Msc and PhD degrees from the Department of Computer Science at the University of Pennsylvania in 1987 and 1989, respectively. Since then, he has been a faculty member in computer science at the University of Western Ontario. See <http://www.csd.uwo.ca/faculty/cling> for more information. His main research areas include machine learning (theory, algorithms, and its applications in Internet, e-business, and bioinformatics). He

has published more than 80 research papers in journals (such as *Machine Learning*, the *Journal of Artificial Intelligence Research*, and *IEEE Transactions on Knowledge and Data Engineering*) and international conferences (such as IJCAI, and ICML).



Qiang Yang received the bachelor's degree from Peking University in 1982 and the PhD degree from the University of Maryland, College Park in 1989. He was a faculty member at the University of Waterloo and Simon Fraser University in Canada between 1989 and 2001. At Simon Fraser University, he held an NSERC Industrial Chair from 1995 to 1999. He is currently a faculty member at the Hong Kong University of Science and Technology. He is

interested in machine learning and data mining, AI planning, and casebased reasoning. He has published two books and more than 100 research articles in conferences and journals. He was the conference chair of the 2001 International Conference on Case-Based Reasoning, a program cochair for the 2000 Canadian AI Conference, and a tutorial cochair of AAAI 2005 Conference. He has been a guest editor for the *IEEE Transactions on Knowledge and Data Engineering*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *IEEE Intelligent Systems*, *Computational Intelligence Journal*, and *Applied Intelligence Journal*. He is also on the editorial board of the *IEEE Transactions on Knowledge and Data Engineering*, the *Knowledge and Information Intelligence Journal*, and the *Web Intelligence Journal*.



William Stafford Noble received the PhD degree in computer science and cognitive science from the University of California, San Diego in 1998, where he studied with Charles Elkan. He then spent one year as a Sloan/DOE postdoctoral fellow with David Haussler at the University of California, Santa Cruz. From 1999 to 2002, he was an assistant professor in of the Department of Computer Science at Columbia University, with a joint appointment at the

Columbia Genome Center. In 2002, he joined the faculty in the Department of Genome Sciences at the University of Washington, with adjunct appointments in the Department of Computer Science and Engineering and in the Division of Medicine. He is the recipient of a US National Science Foundation CAREER award and is a Sloan Research Fellow.

► **For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.**