University at Buffalo
The State University of New York

Department of Computer Science and Engineering

Presents

Yijun Sun
Department of Microbiology and Immunology
Center of Excellence in Bioinformatics and Life Sciences
SUNY Buffalo

Feature selection is a fundamental problem in machine learning and closely related to the issue of the curse of dimensionality. We recently developed a new feature-selection algorithm that addresses several major issues of prior work. The basic idea is to decompose an arbitrarily complex nonlinear problem into a set of locally linear ones through local learning, and then learn feature relevance globally within the large margin framework. The proposed algorithm is based on well-established machine learning and numerical analysis techniques, without making any assumptions about the underlying data distribution. Our experiments showed that the algorithm achieves a close-to-optimum accuracy in the presence of millions of irrelevant features, to a point far beyond that needed in most data-analysis settings one may currently encounter in biological research. By using computational learning theory, our analysis suggests that the algorithm has a logarithmical sample complexity with respect to the number of features. Following the same line of research, we further proved that it is possible to develop a classification algorithm with a generalization error bound that grows only logarithmically with respect to the input data dimensionality for arbitrary data distribution. This result may have significant theoretical implications in learning theory considering that a large part of machine-learning research is focused on developing learning algorithms that behave gracefully when faced with the curse of dimensionality.

Yijun Sun received two B.S. degrees in electrical and mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1995, and the M.S. and Ph.D. degrees in electrical engineering from the University of Florida, Gainesville, USA, in 2003 and 2004, respectively. From 2005 to 2012, he was an Assistant Scientist at the Interdisciplinary Center for Biotechnology Research and an affiliated faculty member at the Department of Electrical and Computer Engineering at the University of Florida. He is now an Assistant Professor in bioinformatics at the Department of Microbiology and Immunology and the Center of Excellence in Bioinformatics and Life Sciences at SUNY Buffalo. His research interests are on machine learning, bioinformatics, and their applications to cancer informatics and microbial ecology. He is a co-recipient of the 2005 IEEE M. Barry Carlton Best Transactions Paper Award. One of his papers is selected as the Spotlight Paper in the September 2010 issue of the TPAMI journal. His research is supported by National Science Foundation, Bankhead-Coley Cancer Research Program, and Susan Komen Breast Cancer Foundation, and his work on metagenomics and feature selection has been used by more than 200 research institutes worldwide.

Thursday, November 1, 2012

3:30 - 4:45 PM

University at Buffalo – North Campus – Davis 113A

Refreshments for attendees after the talk.   This talk is free and open to the public

For more information, please email **cse-dept@cse.buffalo.edu** or contact (716) 645-3180