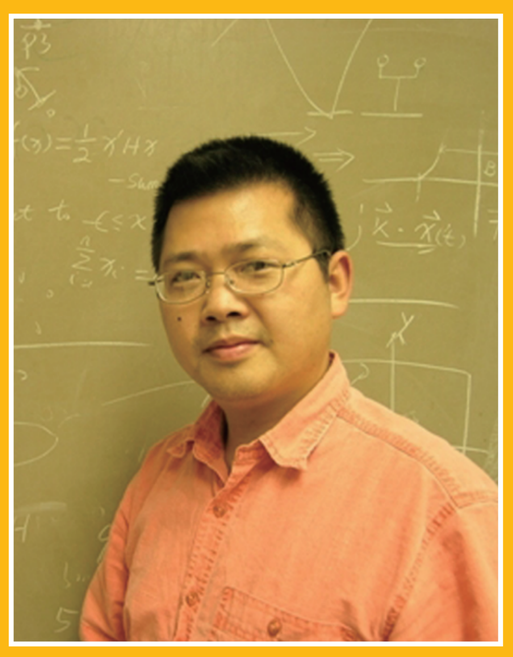
清华大学信息技术研究院系列学术报告4

Locate Potential Support Vectors for Faster Sequential Minimal Optimization



Dr. Hansheng Lei

received his B.S. (07/1998), M.S. (07/2001) and Ph.D. (02/2006) from the Ocean University of China, the University of Science and Technology of China, and the State University of New York at Buffalo respectively, all in computer science. He joined the University of Texas at Brownsville (UTB) as an assistant professor in January 2006. His current research interests lie in distributed machine learning, computer vision and data mining in a High Performance Computing (HPC) environment. He has published over twenty articles in peer-reviewed conferences/ journals. He has several ongoing projects funded by NSF and U.S. Department of Education with a total amount of nearly $3.5M. He has served as a panelist for a variety of funding agencies including NSF, Department of Education and Department of Commerce.

The presentation will serve as a brief report on the latest work of Dr. Lei and his group on Support Vector Machine (SVM). The presentation will also introduce the modern computing infrastructure being built at UTB and several industrial projects led by Dr. Lei.

As a state-of-the-art classifier, SVM has been applied in many areas including Machine Learning, Artificial Intelligence and Data Mining. The training of SVM is essentially an optimization process. Currently, most fast SVM training algorithms are based on Sequential Minimal Optimization (SMO), which sequentially selects a pair of training points at each iteration for joint optimization. SMO uses one outer loop through the entire training set looking for a point which violates the Karush–Kuhn–Tucker (KKT) conditions, and one inner loop to look for a point which heuristically maximizes the increase of the optimization objective function. While there is limited room for improvement in the inner loop, the outer loop can be reduced significantly if one can locate a small subset of training points that potentially contains a large portion of Support Vectors (SVs).

We present a Fisher Discriminant Analysis (FDA) based approach to locate potential SVs in a small core subset so that the SVM training can be focused on a promising small region. The method iteratively trains the classifier on the core subset and tunes it using the remaining subset. Experimental results show that the approach significantly speeds up SMO training by reducing the number of kernel evaluations and the number of iterations as well.

Title:Locate Potential Support Vectors for Faster Sequential Minimal Optimization

Time:2011-5-30 15:00 pm

Speaker: Hansheng Lei, Ph. D.

Assistant Professor

Computer and Information Sciences Department

The University of Texas at Brownsville (UTB)

Place: FIT Building 1-415

Organizer: Research Institute of Information Technology (RIIT), Tsinghua University