

数学与系统科学研究院

计算数学所学术报告

报告人: **Prof. Jieping Ye**

( *Arizona State University, US* )

报告题目:

**Sparse Screening for Exact Data  
Reduction**

邀请人: 戴彧虹、吕召松

报告时间: **2014 年 6 月 26 日 (周四)**

**上午 10:00-11:00**

报告地点: **科技综合楼三层 311**

**计算数学所报告厅**

## **Abstract:**

Recent technological innovations have enabled data collection of unprecedented size and complexity. Examples include web text data, social media data, gene expression images, neuroimages, and genome-wide association study (GWAS) data. Such data have incredible potential to address complex scientific and societal questions, however analysis of these data poses major challenges for the scientists. As an emerging and powerful tool for analyzing massive collections of data, data reduction in terms of the number of variables and/or the number of samples has attracted tremendous attentions in the past few years, and has achieved great success in a broad range of applications. The intuition of data reduction is based on the observation that many real-world data with complex structures and billions of variables and/or samples can usually be well explained by a few most relevant explanatory features and/or samples. Most existing methods for data reduction are based on sampling or random projection, and the final model based on the reduced data is an approximation of the true (original) model. In this talk, I will present fundamentally different approaches for data reduction in that there is no approximation in the model, that is, the final model constructed from the reduced data is identical to the original model constructed from the complete data. Finally, I will use several real world examples to demonstrate the potential of exact data reduction for analyzing large-scale data.

## **Bio:**

Jieping Ye is an Associate Professor of Computer Science and Engineering at the Arizona State University. He is a core faculty member of the Bio-design Institute at ASU. He received his Ph.D. degree in Computer Science from University of Minnesota, Twin Cities in 2005. His research interests include machine learning, data mining, and biomedical informatics. He has served as Senior Program Committee/Area Chair/Program Committee Vice Chair of many conferences including NIPS, KDD, IJCAI, ICDM, SDM, ACML, and PAKDD. He serves as an Action/Associate Editor of Data Mining and Knowledge Discovery, IEEE Transactions on Knowledge and Data Engineering, and IEEE Transactions on Pattern Analysis and Machine Intelligence. He won the NSF CAREER Award in 2010. His papers have been selected for the outstanding student paper at ICML in 2004, the KDD best research paper honorable mention in 2010, the KDD best research paper nomination in 2011 and 2012, the SDM best research paper runner up in 2013, and the KDD best research paper runner up in 2013.

**欢迎大家参加!**