

Department of Computer Science and Engineering

Presents the Faculty Candidate

Jing Gao, University of Illinois at Urbana Champaign

Exploring the Power of Heterogeneous Information Sources

Nowadays, a vast ocean of data is collected from trillions of connected devices everyday. Useful knowledge is usually buried in multiple genres of data, which are from different sources, in different formats, and with different types of representation. Many interesting patterns cannot be extracted from a single data collection, but have to be discovered from the integrative analysis of all heterogeneous data sources available. Although many algorithms have been developed to analyze multiple information sources, real applications continuously pose new challenges: Data can be gigantic, noisy, unreliable, dynamically evolving, highly imbalanced, and heterogeneous. Meanwhile, users provide limited feedback, have growing privacy concerns, and ask for actionable knowledge. In this talk, I will discuss my thesis work on exploring the power of multiple heterogeneous information sources in challenging learning scenarios. I will present two perspectives of learning from multiple sources, i.e., exploring their similarities (knowledge integration) or their differences (inconsistency detection). First, for knowledge integration, I proposed a graph based consensus maximization framework to combine multiple supervised and unsupervised models, which greatly improves classification accuracy. Second, I developed approaches based on probabilistic models and spectral embedding techniques to detect objects performing inconsistently across multiple sources as a new type of outliers. I will show the effectiveness of these general learning techniques with a few sample applications in social networks, Internet, multimedia, and cyber-security.

Jing Gao is a PhD candidate in the Data Mining Research Group in Department of Computer Science at University of Illinois at Urbana Champaign. She received M.E. and B.E. from the Department of Computer Science and Technology at Harbin Institute of Technology in China. She is broadly interested in data and information analysis with a focus on data mining and machine learning. In particular, her dissertation research is on discovering and integrating knowledge from multiple heterogeneous information sources. She is a recipient of the IBM PhD fellowship. Please refer to her homepage for more information: <http://www.ews.uiuc.edu/~jinggao3>.

Thursday, March 10, 2011

3:30 - 4:30 PM

Bell 224 - University at Buffalo – North Campus