Computer Science Seminar

Mining and Using Contextual Information from Large-Scale Web Search Logs

Paul Bennet
Microsoft

Abstract: Information retrieval has made significant progress in returning relevant results for a single query. However, much search activity is conducted within a much richer context of a current task focus, recent search activities as well as longer-term preferences. For example, our ability to accurately interpret the current query can be informed by knowledge of the web pages a searcher was viewing when initiating the search or recent actions of the searcher such as queries issued, results clicked, and pages viewed. We develop a framework that enables representation of a broad variety of context including the searcher's long-term interests, recent activity, current focus, and other user characteristics. We then demonstrate how that can be used to improve the quality of search results. We describe recent progress on three key challenges in this domain: mining contextual signals from large scale logs; understanding and modeling the combination of short-term and long-term behavior; and learning a more robust model that mitigates the risk of applying the contextual model when a simpler model would suffice.

This talk will present joint work with Filip Radlinski, Lidan Wang, Ryen White, Kevyn Collins-Thompson, Wei Chu, Susan Dumais, Peter Bailey, Emine Yilmaz, Fedor Borisyuk, and Xiaoyuan Cui.

Bio:

Paul Bennett is a Researcher in the Context, Learning and User Experience for Search (CLUES) group at Microsoft Research where he works on using machine learning technology to improve information access and retrieval. His recent research has focused on classification-enhanced and contextual information retrieval, pairwise preferences, human computation, and text classification while his previous work focused primarily on ensemble methods, active learning, and obtaining reliable probability estimates, but also extended to machine translation, recommender systems, and knowledge bases. He completed his dissertation on combining text classifiers using reliability indicators in 2006 at Carnegie Mellon where he was advised by Profs. Jaime Carbonell and John Lafferty.

Monday, November 5, 2012, 4:00 pm
Mathematics and Science Center: W303

Mathematics and Computer Science
Emory University