Towards Web Search Engine Scale Data Mining

Jian Pei
School of Computing Science
Simon Fraser University
8888 University Drive, Burnaby, BC CANADA V5A 1S6
Email: jpei@cs.sfu.ca

Abstract

Data mining is one of the most critical driving technologies behind Web search engines. Web search engine scale data mining posts many grand challenges, ranging from efficiency and scalability to diversity and adaptability. In this talk, I will review our recent effort on mining a very large amount of data accumulated in one of the major commercial search engines. Particularly, we tackle the problem of context-aware search and query suggestion by employing statistical models. Moreover, we construct a very large statistical model (millions of states) from a very large amount of data (billions of sessions) by distributed data mining. I will also introduce some of our recent initiatives in Web mining.