Large Scale Machine Learning for Query Document Matching in Web Search

Hang Li, Huawei Labs

In this talk, I will introduce our recent work on large scale machine learning for query document matching in web search. I will start my talk by pointing out that query-document mismatch is one of the biggest challenges in web search. I will then describe the necessity of employing machine learning techniques to overcome the challenge. Next, I will specifically explain two approaches, namely topic modeling and projection into latent space. In the former approach queries and documents are matched in the topic space learned from documents, and in the latter approach queries and documents are matched in the latent space learned from click-through data. Both approaches need to be applied to very large scale data sets. I will describe in details about the learning techniques which we have developed for enhancing the scalability.

Joint work with Wei Wu, Quan Wang, Jun Xu, and Zhengdong Lv.