How do We Incorporate a Recommendation Framework into the Web Search Engines?

Tatsumi Kobayashi, Yahoo Japan Cooperation

In recent years the development of web search engines has been becoming a field of large-scale data science, which includes various broad ideas from fundamental science to technology. We are now facing a new era of enormous amount of contents on the web, and web search engines are expected not only to return appropriate results for any types of user requests, but also to handle various contents such as normal web pages, multimedia (images, movies), UGC (User Generated Content), etc. The challenge is how to incorporate rapidly expanding UGC into search engine's relevancy technology because of the nature of UGC, such as frequent content generation, sensitivity to world trends, variety of people's opinion included, etc. In last decade Natural Language Processing and Machine Learning have been mainly contributing to progress of search engines as a core technology, however, the situation is now going to change. We need more innovative approach/theory/technology to tackle increasing demands for search engines.

In my talk I will focus on a user behavior adaptation framework as a candidate of future technologies, and discuss the impact and possibility against a traditional search engine model which pursues relevancy between a query and a document. More specifically I will review a state-of-the-art technology in user preference estimation, advanced information extraction and semantic structuring technology from contents, and research on matching/ranking algorithms between user preference and contents. Then I will emphasize an importance of a new search metric to measure a search result page quality with more analytic way linked to a user's real demand. The other point I would like to emphasize is necessity of large-scale data mining and grid computing as an infrastructure to support the proposed user behavior adaptation framework. Lastly I will try to present possible technical directions of search engines for the next step. The claim here is that the next generation search engine would be obviously required to understand user's true need and provide an optimal recommendation as a general information delivery system.

Keywords: search metrics, large-scale data mining, grid computing