

A Conversation with Dr. Haifeng Wang

Haifeng Wang

Baidu Inc.

Beijing, China

wanghaifeng@baidu.com

1. Please describe your expertise and contribution to KDD.

My group mainly works on data mining applications in Internet products including the search engine. We study two types of data, Web data and user logs. First, the Web data includes different entities (e.g. websites and web pages), edges between entities (e.g. hyper links), and content (e.g. text and rich-media). Second, user logs contain various user behavior information produced by users when they are using the search engine or other Internet products. These two types of data have different properties, but are correlated and complementary. We build a complete view of the data, mine the most valuable knowledge from the data, and improve our various products, e.g. the Baidu search engine.

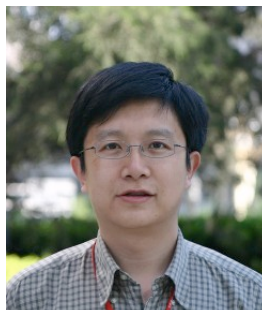
2. Please share with us your view on the future of KDD both in China and the world.

The Web provides a broad data mining application platform, and poses new challenges for the data mining community.

- Multi-source data mining: the data source can be of websites, web pages, links, content, logs, user generated content (UGC), social networks, Hidden Web, etc.
- Heterogeneous data mining: we have to mine knowledge from heterogeneous data formats of hyper links, logs, text and rich-media.

- Quality control in data mining: we aim to identify high quality information and filter out the other, in order to discover the most valuable knowledge.
- Infrastructure support for large-scale data mining: we have to design or customize some infrastructure to support our large-scale data collection and processing.
- Adaptive and large-scale data mining algorithms: we need to adapt existing data mining algorithms to new data and big data.
- User-oriented data mining: our designed data mining algorithms shall response swiftly and serve as an important component in our product to satisfy the users' requirement.

About the author:



Haifeng Wang, Chief Scientist (Basic Technology) at Baidu Inc., and ACL Vice President. <http://ir.hit.edu.cn/~wanghaifeng/>