Challenges in Managing and Mining Large, Heterogeneous Data

Haibo Hu[1], Haixun Wang[2], and Baihua Zheng[3]

[1] Department of Computer Science, Hong Kong Baptist University,
Kowloon Tong, Hong Kong SAR, China
haibo@comp.hkbu.edu.hk
[2] Microsoft Research Asia
Beijing Sigma Center, Hai Dian District, Beijing, China 100190
haixunw@microsoft.com
[3] School of Information Systems, Singapore Management University,
80 Stanford Rd, Singapore 178902
bhzheng@smu.edu.sg

Abstract. Success in various application domains including sensor networks, social networks, and multimedia, has ushered in a new era of information explosion. Despite the diversity of these domains, data acquired by applications in these domains are often voluminous, heterogeneous and containing much uncertainty. They share several common characteristics, which impose new challenges to storing, integrating, and processing these data, especially in the context of data outsourcing and cloud computing.

Some challenges include the following. First, autonomous data acquisition gives rise to privacy and security issues. Therefore, data management and mining must be elastic and privacy-conscious. Second, data is often dynamic and the trend in the data is often unpredictable. This calls for efficient incremental or cumulative algorithms for data management and mining. Load balancing and other real-time technologies are also indispensable for the task. Third, data repositories are distributed. Thus, gathering, coordinating, and integrating heterogeneous data in data management and mining will face unprecedented challenges.

This panel session gives researchers of different background and expertise an opportunity to address these challenging issues together. The main topics of this panel session target the themes in the interdisciplinary domains spreading across database, web, wireless data management, social networking, multimedia, and data outsourcing.