Big Data Mining and Management

Course Syllabus

Course code: COMP 4332/ RMBI 4310

Instructor: *Yangqiu Song*
Email: *yqsong@cse.ust.hk*
Telephone: *23586987*
Office: *3518 (lift 25/26)*

This is a project-oriented course. It will expose students to practical issues of large-scale and real-world data mining. Data mining is a process of extracting implicit, previously unknown, and potentially useful knowledge from data, and it is a critical task in many applications. This course will place emphasis on applications of data mining in areas such as business intelligence, which aims to uncover facts and patterns in large volumes of data for decision support. Application areas also include many other areas in science and engineering applications. This course builds on the basic knowledge gained in the introductory data-mining course and explores how to more effectively mine large volumes of real-world data and tap into large quantities of data. It will introduce new algorithms that can more effectively find hidden and profitable data patterns and knowledge. Working on real-world data sets, students will experience all steps of a data-mining project.

Course Learning Outcomes

- Students will understand issues related to real-world data mining.

- Students will master tools and skills for large-scale data mining projects.

- Students will gain experience on recent topics in business intelligence and social media mining.

Course Prerequisites:
- Statistics and Probability

- Machine Learning/Pattern Recognition/Data Mining

- Programming

Text Books:
- Introduction to Data Mining, by Pang-Ning Tan, Michael Steinbach and Vipin Kumar. Addison-Wesley. [DM]

- Community Detection and Mining in Social Media, by Lei Tang and Huan Liu. Morgan & Claypool Publishers.

- Social Network Data Analytics, by Charu C. Aggarwal, Springer

- Introduction to Data Mining with Case Studies, Second Edition, G.K. Gupta. PHI Learning, 2011. [DM-CS]

- Luna Dong, Divesh Srivastava: Big Data Integration. Morgan & Claypool, 2015. [BI]

- Bing Liu, Web Data Mining, Springer, 2011[BI]

Performance Evaluation: In general, the earned grade in the course will be based on the calculated total points according to the following schedule:

Activity or Task	Max Point Value
Assignment/projects/	50%
Presentations	10%
Final exam	40%
Total	100

Tentative Schedule:

Topics	Content
1	Introduction
2	New Applications of Big Data Mining
3	Introduction to Deep Learning
4	Text Mining in Recommendation
5	Graph Mining in Recommendation
6	Heterogeneous Information Mining
7	More Applications
8	Conclusions