Data Mining
— Techniques and Applications
《数据挖掘技术与应用》
(March 11 - July 8, 2007)

Zhou Shuigeng (周水庚)

March 11, 2007
Contacting the Instructor

- **Zhou Shuigeng (周水庚)**
  - Department of Computer Sci. & Eng.
  - Office: Room 405, CS Building (袁成英计算计机楼405室)
  - Tel: 55664298
  - Email: sgzhou@fudan.edu.cn
“Knowledge is of two kinds: we know a subject ourselves, or we know where we can find information upon it.”

Samuel Johnson (1709-1784)
Course Objectives

- Present basic concepts, methods and techniques, typical algorithms and applications of data mining
- Introduce recent achievements and evolving trends of data mining techniques and applications
Preparatory Knowledge

- Programming Language
- Data Structures
- Algorithms
- Database Management Systems
- Artificial Intelligence or Machine Learning
Textbook

- Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publisher, 2000 (高教出版社影印版)

- 中译本：《数据挖掘：概念与技术》，范明、孟小峰等译，机械工业出版社出版
Additional Reference Books

- Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, *Introduction to Data Mining*, Pearson Education, 2005 (人民邮电出版社影印出版)
- Margaret H. Dunham, *Data Mining: Introductory and Advanced Topics*, Pearson Education, 2003（清华大学出版社影印出版）
Major International Journals

- ACM Transactions on Knowledge Discovery from Data
- IEEE Transactions on Data and Knowledge Engineering
- International Journal of Data Mining and Knowledge Discovery
Major International Conferences

- Data mining conferences
  - SIGKDD
  - IEEE-ICDM, SIAM-DM
  - PAKDD, PKDD, and DaWak etc.

- Database conferences
  - PODS/SIGMOD
  - VLDB
  - IEEE-ICDE
  - EDBT, ICDT, DASFAA

- Other conferences
  - SIGIR, WWW, CIKM etc.
Network Resources

- SIGKDD Explorations
- Weka project
  - www.cs.waikato.ac.nz/~ml/index.html
- IBM Almaden Research Center Data Mining Group
  - www.almaden.ibm.com/software/disciplines/iis/
- 一个数据挖掘中文网站
  - www.dmresearch.net/index.jsp
Course Schedule - March

- **March 11:** Lecture 0/1
  - Data mining concepts, process, functions, system hierarchy and applications, research issues
- **March 18:** Lecture 2
  - Preprocessing techniques for data mining
- **March 25:** Lecture 3
  - Data warehousing and OLAP techniques
Course Schedule - April

- April 1: Lecture 4
  - Data mining primitives, languages, and system architectures
- April 8: Lecture 5
  - Association rules mining: concepts and basic algorithms
- April 15: Lecture 6
  - Association rules mining: advanced algorithms
- April 22: Lecture 7
  - Prediction and classification: decision tree and Bayesian methods
- April 29: Lecture 8
  - Prediction and classification: other methods
Course Schedule – May

- **May 6:** Lecture 9
  - Clustering analysis: basic concepts and algorithms
- **May 13:** Lecture 10
  - Clustering analysis: advanced algorithms
- **May 20:** Lecture 11
  - Outlier detection: concepts and algorithms
- **May 27:** Lecture 12
  - Spatial data mining
Course Schedule - June

- June 3: Lecture 13
  - Sequence patterns mining
- June 10: Lecture 14
  - Text and Web mining: basic concepts and major algorithms
- June 17: Lecture 15
  - Application areas and development trends
- June 24: Lecture 16
  - Data mining applications: cases study
Course Schedule - July

- July 1: No course. Preparing for the final examination
- July 8: Final exam
  - Time/room (TBD)
Course Workload

- Class attendance
- Final exam
Evaluation Scheme

- **Class attendance (20%)**
  - The students are required to attend the course

- **Final exam (80%)**
Course Homepage

  - You can download all lecture notes from the web site