Wright State University

CORE Scholar

Computer Science & Engineering Syllabi

College of Engineering & Computer Science

Fall 2012

CS 7720: Data Mining

Guozhu Dong
Wright State University - Main Campus, guozhu.dong@wright.edu

Follow this and additional works at: https://corescholar.libraries.wright.edu/cecs_syllabi

Part of the Computer Engineering Commons, and the Computer Sciences Commons

CS 7720 Data Mining
Fall Quarter, 2012

Description: This course studies the fundamental concepts, issues, and techniques of data mining. Topics include basics of data, data preprocessing, feature selection/extraction, frequent pattern and association/correlation mining, classification, clustering, outlier analysis, OLAP/OLAM, contrast mining, applications, etc.

Prerequisite: CS 6700 (CS 605) (Introduction to Database Systems), or CS 6850 (CS 609) (Introduction to AI), or equivalent, or consent of the instructor. Implicitly implied by the above, graduate student level programming skills are also required.

Instructor: Dr. Guozhu Dong. Joshi 383.

Phone & Email: (937)-775-5066, guozhu.dong@wright.edu

Class details: 7:40-9:00 MW, Joshi 193

Office hours: 4:30-5:25, T Th. Use e-mail for short questions.

Text Book: Data Mining: Concepts and Techniques, 3rd edition. J. Han, M. Kamber, and Jian Pei. Morgan Kaufmann.

Reference texts: Introduction to Data Mining, P.-N. Tan, M. Steinbach, V. Kumar, Addison Wesley.

Data Mining: Practical Machine Learning Tools and Techniques, I.H. Witten and E. Frank, Morgan Kaufmann.

Principles of Data Mining. D. Hand, H. Mannila, and P. Smith. MIT Press.

Contrast Data Mining: Concepts, Algorithms, and Applications. G. Dong and J. Bailey, Chapman & Hall/CRC (Data Mining and Knowledge Discovery Series).

Sequence Data Mining. G. Dong and J. Pei. Springer.

Coverage of Book Chapters: We plan to focus on the following chapters of the text book:
Chapter 1. Introduction
Chapter 2. Getting to Know Your Data
Chapter 3. Preprocessing
Chapter 6. Mining Frequent Patterns, Associations and Correlations: Concepts and Methods
Chapter 8. Classification: Basic Concepts
Chapter 10. Cluster Analysis: Basic Concepts and Methods
Chapter 12. Outlier Analysis
Chapter 4. Data Warehousing and On-Line Analytical Processing
Chapter 5. Data Cube Technology
Chapter 13. Trends and Research Frontiers in Data Mining
Please note that we do not plan to cover all the chapters above completely.
In addition to the chapters above, we also plan to present some results on contrast data mining (from the Contrast Data Mining book), and give a brief discussion on the three advanced chapters of the text book (Chapters 7, 9, 11) if time permits; students who are seriously considering doing research in data mining should study these chapters.

Resources: Slides will be available on pilot.

Students may find this webpage useful: http://www.kdnuggets.com/, especially its pointers to datasets.

Many Java programs for data mining are available at www.cs.waikato.ac.nz/ml/weka, which you may want to install and experiment with.

Grading: Homeworks: 10%, Midterm: 25%; Final: 35%; Projects 30%.

Final grade: A=[90,100], B=[80,90), C=[70,80), D=[60,70), F=[0,60).

The instructor may curve the final grades in such a way that they deviate from these standards at his discretion.

The projects require extensive programming. Submissions that do not compile or that do not address project requirements will receive zero or very low marks.

Handouts: Handouts, and other course materials may be distributed in class. It is the students' responsibility to collect them.

Important dates:

Midterm exam: class time, Monday, 10/15/12 (Week 8).

Final exam: 8:00-10:00pm, Wednesday, 12/12/12.