

Principles of Knowledge Discovery in Databases

Fall 1999

Dr. Osmar R. Zaiane



University of Alberta



Class and Office Hours

Class:

Mondays, Wednesdays and Fridays from 10:00 to 10:50

Office Hours:

Tuesdays from 11:00 to 11:55

Course Requirements

- Understand the basic concepts of database systems
- Understand the basic concepts of artificial intelligence and machine learning
- Be able to develop applications in C/C++ or Java

Course Objectives

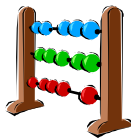
To provide an introduction to knowledge discovery in databases and complex data repositories, and to present basic concepts relevant to real data mining applications, as well as reveal important research issues germane to the knowledge discovery domain and advanced mining applications.



Students will understand the fundamental concepts underlying knowledge discovery in databases and gain hands-on experience with implementation of some data mining algorithms applied to real world cases.

Evaluation and Grading

There is no final exam for this course, but there are assignments, presentations, a midterm and a project. I will be evaluating all these activities out of 100% and give a final grade based on the evaluation of the activities.



- Assignments (? 2) 10%
- Midterm 30%
- Project 35%
 - Quality of presentation + quality of report + quality of demo
- Class presentations 25%
 - Quality of presentation + quality of slides + peer evaluation

More About Evaluation

Re-examination.

None, except as per regulation.

Collaboration.

Collaborate on assignments; do not merely copy.



Notes and Textbook

Course home page:

<http://www.cs.ualberta.ca/~zaiane/courses/cmput690/>

There is also a newsgroup for the course.



Textbook:















Data Mining: Concepts and Techniques

Jiawei Han and Micheline Kamber

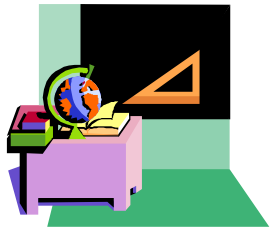
Morgan Kaufmann Publisher, 2000 (in the works)



Legend of Link Icons

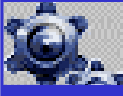

-  Link to an HTML page (default)
-  Link to a text file
-  Link to a page to be displayed in a new browser window
-  Link to a Portable Document Format (PDF) file
-  Link to a Postscript (PS) file
-  Link to slides presentation
-  Link to a compressed (gz or zip) file
-  Link to an image
-  Link to a video
-  Link to a C/C++ program listing
-  Link to a Java class file
-  Link to a data file
-  Link to password protected html page
-  Link to a page under construction

On-line Resources



- Course notes
- Course slides
- Web links
- Glossary
- Student submitted resources
- Student spaces
- U-Chat
- Newsgroup
- Frequently asked questions

Projects

	Choice	Deliverables
	Implement data mining project	Project proposal + 10' proposal presentation + project demo + project report
	Write survey paper (or research paper)	Survey proposal + 10' proposal presentation + paper presentation + survey paper (20-30 pages)

Examples of survey topics:

- Web usage mining
- Knowledge discovery from unstructured or semi-structured data on the WWW
- Text mining
- Data mining from non-traditional databases (OODB/deductive DB).
- Spatial data mining
- Multimedia data mining
- Clustering
- Classification
- Association rule mining
- Datacube construction
- Datawarehousing

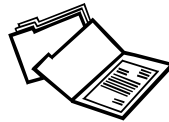
Examples of data mining projects will be posted on the course web site.



More About Projects

Either for the implementation project or the survey paper, students should write a project proposal (1 or 2 pages).

- project topic;
- implementation choices;
- approach;
- schedule.



All projects are demonstrated at the end of the semester.

Implementations: C/C++ or Java,

OS: Linux, Window NT/98 , or other systems.





Course Schedule

(Tentative, subject to changes)

There are 14 weeks from Sept. 10th to Dec. 8th.

Distribution of papers list **week 2**
Paper presentation (schedule will be on-line)

Assignment 1 distribution **week 3**
Assignment 1 due **week 4**
Assignment 2 distribution **week 5**
Assignment 2 due **week 8**

Midterm **week 8 or 9**

Project proposals due **week 5**
Project reports and survey papers due **week 13**

Course Content

- Introduction to Data Mining
- Data warehousing and OLAP
- Data cleaning
- Data mining operations
- Data summarization
- Association analysis
- Classification and prediction
- Clustering
- Web Mining
- Similarity Search
- *Other topics if time permits*

