CS 685: Special Topics in Data Mining

Syllabus

With the unprecedented rate at which data is being collected today in almost all fields of human endeavor, there is an emerging economic and scientific need to extract useful information from it. Data mining is the process of automatic discovery of patterns, changes, associations and anomalies in massive databases, and is a highly inter-disciplinary field representing the confluence of several disciplines, including database systems, data warehousing, machine learning, statistics, algorithms, data visualization, and high-performance computing. This seminar will provide an introductory survey of the main topics (including and not limited to classification, regression, clustering, association rules, trend detection, feature selection, similarity search, data cleaning, privacy and security issues, and etc.) in data mining and knowledge discovery as well as a wide spectrum of data mining applications such as biomedical informatics, bioinformatics, financial market study, image processing, network monitoring, social service analysis.

For each topic, a few most related research papers will be selected as the major teaching material. Students are expected to read the assigned paper before each class and to participate in the discussion during or after the presentation.

Jinze Liu	**Meeting Time**: TR 2:00PM-3:15PM
liuj@cs.uky.edu	**Meeting Place**: MMRB 243
http://www.cs.uky.edu/~liuj	**Office Hours**: By appointment
(859) 257 – 3101	**Office:** 235 James F. Hardymon Building

Prerequisite:

Some background in algorithms, data structures, statistics, machine learning, artificial intelligence, and database systems is helpful.

Book:

Mining of Massive Data, by Anand Rajaraman, Jeff Ullman and Jure Leskovec. The book can be accessed freely online (http://i.stanford.edu/~ullman/mmds.html#latest)

Other References:

1). Data Mining --- Concepts and techniques, by Han and Kamber, Morgan Kaufmann, 2006. (ISBN:1-55860-901-6)
2). Principles of Data Mining, by Hand, Mannila, and Smyth, MIT Press, 2001. (ISBN:0-262-08290-X)
3). Introduction to Data Mining, by Tan, Steinbach, and Kumar, Addison Wesley, 2006. (ISBN:0-321-32136-7)

4). The Elements of Statistical Learning --- Data Mining, Inference, and Prediction, by Hastie, Tibshirani, and Friedman, Springer, 2001. (ISBN:0-387-95284-5)
5). Pattern Recognition and Machine Learning, by Christopher M. Bishop, 2006.

Grading

Each student in CS685 will be expected to *present once in class* and *lead the discussion* following his/her presentations. Students are also expected to finish *a project* on selected topics. There will be 4-6 *assignments* and 1 *mid-term exam*.

Homeworks	30%
Exam	20%
Presentation	20%
Project	30%

Expected Outcome

At the end of class, the students are expected to

1. Get a basic understanding of a variety of data mining techniques including: association rule mining, classification, clustering and graph mining in terms of algorithms and modeling.
2. Learn the interdisciplinary nature and forefront research topics in data mining.
3. Be able to determine whether an application can be solved by data mining techniques, if so, which techniques should be applied to.
4. Develop a novel algorithm to solve a general data mining problem or apply one or more data mining techniques to a particular dataset, through course project.

Academic Conduct Expectations

Students are expected to complete all assignments independently. Honest and ethical behaviors are always expected. There will be no tolerance for plagiarism or other academic misconduct. The minimum punishment is an E grade that cannot be removed by the repeat option. You may read U.K. Student Rights and Responsibilities at http://www.uky.edu/StudentAffairs/Code for a detailed description.