Computer Science 5083: Knowledge Discovery and Data Mining

Instructor: Dr. McGovern

Spring 2009

1 Introduction

In today's information-rich world, finding automatic ways to sift through the wealth of information and to identify the salient patterns is critical. The process of knowledge discovery is about making sense of data that is too complex for humans to understand. Knowledge discovery is a interactive process between a human and a computer. Data mining techniques highlight salient patterns in the data. These patterns are examined by the human to provide feedback. Data mining refines the patterns and the cycle continues until the human is satisfied.

By the end of the semester, you will be able to:

- Recognize a knowledge discovery/data mining task

- Choose an appropriate knowledge representation (static vs dynamic, discrete vs continuous, propositional vs relational)

- Identify the elements of a successful knowledge discovery algorithm

- Apply current knowledge discovery/data mining algorithms to new situations

- Create new and successful knowledge discovery/data mining algorithms

- Evaluate and analyze the performance of new approaches

This will be a seminar style class where we will read and discuss current papers. In addition, we will have a class-wide project that we will work on and discuss.

2   General Information

Class time: TR 3-4:15

Class location: Felgar 334

Prerequisites: CS 4033/5033 or CS 6033 or permission of the instructor. Note that prior programming experience is assumed. Also, I assume a mathematical background and an understanding of statistics and probability.

Course description: This course is application oriented and gives students the tools to develop and apply appropriate data mining and knowledge discovery techniques to real world problems.

Instructor: Dr. McGovern

- *Office:* EL 144A
- *Phone:* 325-5427 (voice mail available)
- *URL for class:* http://learn.ou.edu
- *Personal URL:* http://www.cs.ou.edu/∼amy
- *Email:* amcgovern@ou.edu
- *Office hours:* Tuesday: 10-11:30, Thursday 1:30-2:45. Also by appointment. Also available via AIM at *dramymcgovern.*

3   Evaluation

This is a small seminar class where everyone is expected to be actively keeping up with the readings, project, and participating in class. To stimulate class discussion, you will be asked to write a short critical summary of each paper. In addition, we will all participate in a class project (topic chosen in the first week). Your final grade will be calculated from these items (project: 50 %, summaries 25%, participation 25%).

Slack days: You have two slack days for reading summaries which will entitle you to skip the summary for that reading. Please note that you should not skip the reading entirely as you will end up behind!

Projects: There will be a class-wide project. In 2007, we focused on the netflix prize data where the goal (for $1,000,000$) was to improve the netflix predictions by 10%. The project for 2009 will be chosen in January 2009 based on the available real-world datasets but it will again be chosen from a real-world task.

Desire2Learn Grade Summary: Desire2Learn has a grade book that is used to store the raw data that is used to calculate your course grade. It is the responsibility of each student in this class to check their grades on Desire2Learn after each project or homework is returned. If an error is found, bring the grading document to me or the TA, and we will correct Desire2Learn.

4   Course Policies

The following set of rules will help keep us all on the same page all semester and help to ensure fair treatment for all students.

Academic Misconduct: Academic misconduct hurts everyone but particularly the student who does not learn the material. All work submitted for an individual grade should be the work of that single individual and not her friends. It is fine to ask a fellow student for help as long as that help does not consist of copying any computer code, or solutions to other assignments. Students working on joint projects may certainly help one another and are expected to share code within the project group. However, they may not share beyond the group.

1. Collaboration is encouraged for homework and projects. For the projects, you will work within your groups. For the homework, you may form study groups so long as each homework is in your own words. Write your study partners' names on your homework when you turn it in.

2. Do not show another student (or group) a copy of your projects or homework before the submission deadline. The penalties for permitting your work to be copied are the same as the penalties for copying someone elses work.

3. Make sure that your computer account is properly protected. Use a good password, and do not give your friends access to your account or your computer system. Do not leave printouts or thumb drives around a laboratory where others might access them.

Upon the first documented occurrence of academic misconduct, I will report it to the Campus Judicial Coordinator. The procedure to be followed is documented in the University of Oklahoma Academic Misconduct Code[1]. In the unlikely event that I elect to admonish the student, the appeals process is described in `http://www.ou.edu/provost/integrity-rights/`.

---

[1]http://www.ou.edu/studentcode

Project code: Your project code and writeups must be written exclusively by you or your group. **Use of any downloaded code or code taken from a book (whether documented or undocumented) is considered academic misconduct and will be treated as such.** Exceptions from this policy (such as a project that builds on an existing open-source project) may be granted but you MUST speak with me first.

Classroom Conduct: Disruptions of class will not be permitted. Examples of disruptive behavior include:

- Allowing a cell phone or pager to repeatedly beep audibly.
- Playing music or computer games during class in such a way that they are visible or audible to other class members.
- Exhibiting erratic or irrational behavior.
- Behavior that distracts the class from the subject matter or discussion.
- Making physical or verbal threats to a faculty member, teaching assistant, or class member.
- Refusal to comply with faculty direction.

In the case of disruptive behavior, I may ask that you leave the classroom and may charge you with a violation of the Student Code of Responsibilities and Conduct.

Class Web Page: Login to the Desire2Learn website using your 4+4 (first four letters of your last name followed by the last four digits of your student number), using your standard OU password. If you have difficulty logging in, call 325-HELP. This software provides a number of useful features, including a list of assignments and announcements, an electronic mailing list, newsgroups, and grade book. All handouts are available from Desire2Learn. You should check the site daily. When I update the site, I will post an announcement telling you what has been added and where it is located.You are responsible for things posted on the site with a 24 hour delay.

Class Email Alias: Urgent announcements will be sent through email. It is your responsibility to:

- Have your university supplied email account properly forwarded to the location where you read email.
- Make sure that your email address in Desire2Learn is correct, and forwards email to the place where you read it. I'll send out a test message during the first week

of class. If you do not receive this message, it is your responsibility to get the problem resolved immediately.

- Have your email program set up properly so that replying to your email will work correctly the first time. You can send email to yourself and reply to yourself to test this.

If you need assistance in accomplishing any of these tasks, contact 325-HELP.

Newsgroups and Email: The newsgroup on Desire2Learn should be the primary method of communication, outside of class. This allows everyone in the class to benefit from the answer to your question. If you email me a question of general interest, I may post your question and my answer to the newsgroup. Matters of personal interest should be directed to email instead of to the newsgroup, e.g. informing me of an extended personal illness. Posting guidelines for the newsgroup are available on Desire2Learn.

Religious Holidays: It is the policy of the University to excuse the absences of students that result from religious observances and to provide without penalty for the rescheduling of examinations and additional required classwork that may fall on religious holidays.

Incompletes: The grade of I is intended for the rare circumstance when a student who has been successful in a class has an unexpected event occur shortly before the end of the class. I will not consider giving a student a grade of I unless the following three conditions have been met. 1. It is within two weeks of the end of the semester. 2. The student has a grade of C or better in the class. 3. The reason that the student cannot complete the class is properly documented and compelling.

Accommodation of Disabilities: The University of Oklahoma is committed to providing reasonable accommodation for all students with disabilities. Students with disabilities who require accommodations in this course are requested to speak with the professor as early in the semester as possible. Students with disabilities must be registered with the Office of Disability Services prior to receiving accommodations in this course. The Office of Disability Services is located in Goddard Health Center, Suite 166, phone 405/325-3852 or TDD only 405/325-4173.