CS 4701 Artificial Intelligence
Columbia University, Summer 2009

This course will review the main topics of artificial intelligence and will emphasize the uncertain knowledge, and learning components. Participants will have an opportunity to develop a major project using one of the main methods of artificial intelligence in their area of interest.

Topics covered include (number on the left of each topic corresponds to chapter of the Russell and Norvig's book):

Introduction and Logic:
1 Introduction to Artificial Intelligence
2 Intelligent Agents
7 Logical Agents
8 First-Order Logic

Problem Solving:
3 Solving Problems by Searching
4 Informed Search and Exploration
6 Adversarial search (games)

Uncertain Knowledge and Reasoning:
13 Uncertainty
14 Probabilistic Reasoning
15 Probabilistic Reasoning Over Time
Learning:
18 Learning from Observations
20 Statistical Learning Methods
21 Reinforcement Learning

- Instructor

German Creamer
633 Seeley W. Mudd Building

Office Hours: Tuesday/Thursday 9.10 pm - 9.40 pm or by appointment
825 Seeley W. Mudd Building
ggc14@columbia.edu

- Teaching Assistant:

Office Hours:
TA room (122A Mudd)

- Class schedule

Tuesdays/Thursday, 6:00 - 9:10 PM, 633 Seeley W. Mudd Building

- Requirements

Data structures and algorithms course (CS 3137).

- Textbook and software

The main textbook that we will use in this class is: *Artificial Intelligence: A modern approach* by Stuart Russell and Peter Norvig 2nd. edition

This will be available at Columbia University bookstore.

Additional Reference:

Ian Witten and Eibe Frank, Data Mining, (1st. ed. Academic Press, 2000. or better 2nd. ed. Morgan Kaufmann, 2005). This is a good reference book to study applied techniques of artificial intelligence and data mining. Most of the concepts explored in this book are implemented in the software WEKA.

Project and software

A major project will be developed during the course. You are not required to use LISP for this project. You can use WEKA and/or another language such as C++ or Java. I recommend that you use the software WEKA developed by the authors of the above book at the University of Waikato in New Zealand. You can download it for free in the above website where you have a lot of support documentation including the chapter 8 of the Witten and Frank's book.

- Communication

Feel free to let us know what you find good, interesting, or problematic about the course. In case that you have questions, you are welcome to visit the instructor or TA during their office hours or communicate by email.

Please do not use the email as a substitute for office hours visit. Do not expect an immediate answer of your email questions. We will try to answer as soon as possible, however if you have a complicated question, it is better that you ask it during our office hours.

- Class Attendance

You are responsible for all material presented in the class lectures, recitations, and so forth. Some material will diverge from the textbooks and we have important class discussions, so regular attendance is important.

- Grading Scheme

Grading is determined as follows:

  - 40% - Homeworks / Project
  - 22% - Midterm
  - 32% - Final
  - 6% - Class Participation

There are two assigned homeworks. You can skip two or you will receive the highest grade of two out of four homeworks.

- Homework Policies

Late Policy: Zero credit for anything handed in after it is due without explicit approval of the instructor. Homework is due at the beginning of class on the due date.

Cooperation on Homework: Collaboration on solutions, sharing or copying of solutions is not allowed. Of course, no cooperation is allowed during exams. This policy will be strictly enforced. See the Computer Science department's Policies and Procedures Regarding Academic Honesty for details.

Re-grades: If you dispute the grade received for an assignment, you must submit, in writing (not email), your detailed and clearly stated argument for what you believe is incorrect and why. This must be submitted by the beginning of the next class after the assignment was returned. (For example, if the assignment was returned to the class on Tuesday, your write-up is due by beginning of class on Thursday). Requests for re-grade after the beginning of class will not be accepted. A written response will be provided by the next class indicating your final score. (By Tuesday for the prior example). Be aware that requests of re-grade of a specific problem can result in a regrade of the entire assignment. This re-grade and written response is final; no additional re-grades or debate for that assignment.

Final Project

Build a decision support system (DSS) with these characteristics:

- Can be programmed using WEKA and/or using a functional language such as LISP or using an object-oriented language such as C++ or JAVA. You can also use Perl. I recommend using WEKA because of the limited amount of time available.
- Can be model-driven or data-driven using any modern AI technique that you consider appropriate for the problem studied
- You must present a final report. Your write-up should be in ASCII plain text format (.txt) or Postscipt (.ps) or the Adobe Portable Document Format (.pdf).

Please do not submit Microsoft Office documents, LaTeX source code, or something more exotic. We will not be able to read it. LaTeX is preferred and highly recommended, but it is not mandatory. You can use any document editing software you wish, as long as the final product is in .ps or .pdf. Even if you do not use LaTeX to prepare your document, you can use LaTeX notation to mark up complicated mathematical expressions, for example.

Your project can be in your current research/interest area, but make sure to show what novel work you have done for it for this course, you can't just submit previous work.

Projects are done in teams of 3 people.

Project Proposal

The project proposal should be typed, not hand writen, and should include:

1. Problem definition, what is it that this project will do and what problems it will solve.
2. How are you planning to solve the problem (a high level description), which search strategies or algorithms are you considering, which ones do you think are appropriate for this problem domain and why?
3. Programming language to be used.
4. How would one test the performance of the algorithm to be used?
5. What is the data to be used?

Remeber to include the names of all team members on the proposal.

Look at the case distributed in class from appendix D, Vasant Dhar and Roger Stein. Seven methods for transforming corporate data into business intelligence. Upper Saddle River: Prentice Hall. 1997.

Project Report

The final report must include at least the following sections:

- Organization or area of application
- Problem
- Solution
- Results
- Lessons learned

This final report may have one of the two following formats according to your orientation:

1. Industry (biomedical, finance etc.): case study (see appendix D, Vasant Dhar and Roger Stein. Seven methods for transforming corporate data into business intelligence. Upper Saddle River: Prentice Hall. 1997.)
2. Academic: academic paper presented in conferences such as "Innovative Applications of Artificial Intelligence Conference"

   If you are planning to present your project as an academic paper, look at the following link which describes the format that you should follow (in general terms): Academic

This article explains the text that you can use to compare different learning algorithms: Dietterich, T. G., (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Computation, 10 (7) 1895-1924. Postscript preprint. (Revised December 30, 1997).

Files that you must include in your final submission:

1. project report
2. source code
3. readme file explaining how to operate your system
4. output file showing some results of your program.

We will evaluate all the projects under the same standards. We will necessarily be judging three person projects differently from two-person projects (differently from individual projects), but we will not apply a tougher standard - instead we are looking for a more extensive investigation of effort proportional to the size of the team. Final writeup size guideline is roughly 5-7 pages for two person team and 10-12 pages for teams of three.

Teams are responsible for managing themselves and distributing their workload. We really do not want to deal with your team politics.

Project Demo

Each project will have one mandatory demonstration/discussion at the completion of the project. This will involve an informal 10-minute presentation of your project to the class followed by a discussion of your work and a demonstration of the project. The purpose of this presentation is two-fold: 1) it allows you to highlight portions of your project that may not be adequately shown in your write-up and 2) it allows the instructor to understand how deeply you understand the material and your own work. Details for the Presentations will be provided as the course progresses.

Some Resources

You can also look at the following links for ideas or tutorials that could be useful for your project:

http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/ai-repository/ai/areas/0.html This is a depository about different AI software packages.

http://satirist.org/learn-game/links/tutorial.html Tutorials for machine learning methods (not all links work).

http://www1.cs.columbia.edu/~jebara/4771/PROJECT.htm This link has good project and format suggestions for machine learning papers.