OR/STAT 719
Computational Models for Probabilistic Reasoning

Prof. Paulo C. G. Costa, PhD
Department of Systems Engineering and Operations Research
George Mason University

Course Description

Spring 2010

Intelligent computer agents must perform goal-directed action in complex, uncertain, and dynamic environments. Agents tasked with problems of any complexity must have methods for handling uncertainty. This course examines theory and methods for building computationally efficient software agents that reason, act and learn in environments characterized by noisy and uncertain information. The course covers methods based on graphical probability and decision models. The course studies approaches to representing knowledge about uncertain phenomena, drawing inferences about uncertain phenomena, and planning and acting under uncertainty. Theory, practical tools, and hands-on experience are provided. Students learn graph theoretic concepts for representing conditional independencies among a set of uncertain hypotheses. Students study exact and approximate methods for updating probabilities to incorporate new information. Practical model building experience is provided. Probabilistic and decision theoretic approaches to major areas of artificial intelligence such as knowledge representation, machine learning, data mining, case-based reasoning, planning, and temporal reasoning are discussed. Students apply what they learn to a semester project of their own choosing.

Class Details

Prerequisites

* *The listed prerequisites are STAT 692 or SYST/STAT 664 or permission of instructor. Students are expected to have a strong grounding in calculus-based probability theory, to have graduate-level mathematical sophistication, to be competent at building mathematical models and deriving conclusions from the models, to be comfortable performing statistical analysis of a data set and deriving conclusions from the analysis, and to be comfortable with software tools for mathematical modeling such as Matlab, S+, or Mathematica.*

Classes

** Room 134 of the Innovation Hall.*
** Mondays, from 4:30 p.m. to 7:10 p.m.*

Office hours

** Room 2227 of the Engineering Building.*
** Mondays and Thursdays, from 3:00 p.m. to 4:00 p.m., or by appointment.*

Administrative

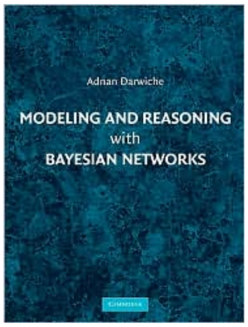
** Registration deadline: Feb 02.*
** Drop without Tuition Penalty Deadline: Feb 02.*
** Drop with Tuition Penalty dates: Feb 03 to Feb 19.*
** Final Drop deadline: Feb 19.*

Logistics and Expected Behavior

1. All communication, file exchanges, and submissions must be done using the Blackboard System! Yet, you are welcomed to use my email directly as a backup or to address aspects not related with the course content.
2. You are allowed to enter or leave at any time, provided you do your best to avoid disrupting the activity going on.
3. Please make sure you have your cell phone, pagers, pda, etc., in silent mode. Should you see yourself in extreme need to answer an incoming call, just leave the room to do so.
4. Attendance in class is very important. Information will be presented that will not necessarily be in the book that will show up on the midterm and final.
5. Students are permitted to work together on assignments, but your write-up must be your own. Assignments are intended to provide practical, hands-on experience with the ideas presented in the course. Students will have at least 2 weeks to complete each assignment. Late assignments receive half credit. The take-home exam must be done individually without collaboration.
6. Academic Policy: All academic policies as given in the Honor System and code will be strictly followed. Visit URL: http://catalog.gmu.edu/content.php?catoid=5&navoid=104

Exercise planning, be proactive and do your best to stay ahead of schedule.

Textbook and Software



Modeling and Reasoning with Bayesian Networks, Adnan Darwiche. Cambridge University Press; 1 edition (April 6, 2009. 548 pp.
ISBN-10: 0521884381.
ISBN-13: 9780521884389.

Students are encouraged to also refer to the following resources:

- Learning Bayesian Networks (Paperback), Richard E. Neapolitan. Prentice Hall; illustrated edition (April 6, 2003). ISBN-10: 0130125342.
- Bayesian Networks and Decision Graphs, Finn V. Jensen, Duxbury Press, 1997. ISBN-10: 0534516920.
- Causality: Models, Reasoning and Inference, Judea Pearl. Cambridge University Press; 2nd edition (September 14, 2009). ISBN-10: 052189560X.

Lecture Notes

Lecture notes for each chapter will be made available from the Blackboard course page either before class or just after it. You will need to download Adobe Acrobat Reader to read these lecture notes.

Grading

The grading structure of this course is as follows:

- Assignments (20% of grade)
- Midterm (25% of grade)
- Final Exam (25% of grade)
- Project (30% of grade)

Assignments

There will be assignments posted via Blackboard during the course. Each assignment will have its respective due date defined during the announcement. I might sometimes not grade the assignments in detail, but will always use it to gain insight about the understanding of the students on the subject.

You are not prevented to work with your peers on the exercises, and are even encouraged to do so. However, each student must provide his/her own answers, and I might verify whether he/she actually worked in his/her respective exercise and understood the solution provided. In any case, past experience consistently shows that students who didn't keep up with the assignments have had a hard time with the exams and the project.

Each assignment is out of 10 points. Late assignment policy: Late by 3 days: you can earn a max of 7 points, provided you have all correct answers. If late beyond 3 days then it will not be graded. Always check for grades on Blackboard. If you don't see the grade, report to me by the next class after assignments have been returned. I will not entertain missing grade requests that come later in the semester.

Exams (Midterm and Final)

Take-home, individual work, Honor coded. Details will be provided via Blackboard during the course.

Course Project

Overview. Students are required to do a semester long research project applying what they have learned in this course to a problem of their own choosing. The most straightforward type of research project is a modeling exercise. Here is a step-by-step description of an average project:

1. Consider a problem related to your work or dissertation topic in which uncertainty plays an important role.
2. Construct a model for the problem using a Bayesian network package such as Netica, UnBBayes, JavaBayes, Hugin, Genie or another belief network software package (see links on the Blackboard website).
3. Apply the network engineering process described in class to do the knowledge engineering for your model.
4. Even if you are your own expert, make sure you document:
    - the elements of your model,
    - the rationale for your design decisions,
    - the process by which you evaluated and refined your model, and
    - your overall evaluation of the results of modeling.

If you are more ambitious, you might try a project in learning, in assembling network fragments to build a complex model, or in embedding belief network software into some other application.

For embedding belief net software into applications, Java Bayes is a free (for research use) belief network package with API's. Netica also has API's, but not in the free student version. IBAL is a language for specifying probabilistic models. Primula, which interoperates with SamIam, is a relational modeling system that allows you to build complex models with repeated structure. Be forewarned: the learning curve for that software package is steep! Finally, students with a software bent may request to do an implementation -- an inference or learning algorithm (e.g., junction tree, K-2, particle filter), or a GUI for dragging and dropping nodes and entering probabilities.

A "think piece" that discusses issues and posits an architecture with no implementation is discouraged. I believe ideas need to be tested in the real world to determine whether they have merit. The existence of many good (and free!) publicly available Bayesian network software packages means that you do have an opportunity to test out your ideas. You are challenged to make your ideas real.

Deliverables for the project include a project proposal, a presentation and a written final report.

Project proposal. Due March 2. The proposal should be no more than 3 single-spaced pages (it can be less). The purpose is to get feedback on your idea to help you produce a better project. Here is a sample proposal outline :

- Introduction: Describe the problem. Why is the problem important? What role does uncertainty play? Why is it important to incorporate uncertainty? Give a brief synopsis (few sentences) of your project and how it addresses these issues.
- Background: You should do a survey of the literature relevant to your problem and to the role uncertainty plays. What others have done? What are the open issues? How does your work relate to what has been done previously?
- Research plan: Describe what you plan to do.

Oral presentations. Scheduled during the last 3 weeks of class. You will give a brief in-class presentation of your project. Demonstrations with commentary are encouraged. Slides must be submitted via Blackboard no later than 2 p.m., Eastern Time, of **the day before the presentation!** It is tolerable to make changes to your presentation after submitting it, although you are expected to handle a reasonably "close-to-final" version of the actual presentation.

Final report. A hard copy of your written report is due on Monday 5/3, 4:30 p.m. Eastern Time. A soft copy must be submitted via the Blackboard system, no later than the hard copy's deadline, and must contain the following structure:

1) Title
2) Abstract: A brief description of the project.
3) Introduction: Describe the problem. Why is the problem important? What role does uncertainty play? Why is it important to incorporate uncertainty? Give a brief synopsis (few sentences) of your project and how it addresses these issues. Data collection method for values
4) Background: You should do a survey of the literature relevant to your problem and to the role uncertainty plays. What has been done by others? What are the open issues? How does your work relate to what has been done previously?
5) Description of your project. If you built a belief network model, you should include a graphic showing your network structure, justify the conditional dependence and independence assumptions you are making, and describe qualitatively the probability assessments you made and the reasons for them.

Describe the test cases you ran, and discuss the results. Are they what you expected? Why or why not?

6) Evaluation: What did you learn from your project? Was it a success? Were your results satisfactory or unsatisfactory? Why? What would you change if you had it to do over?
7) Summary and Conclusions.
8) Appendices: You should include full conditional probability tables for your networks, as well as full documentation of test cases you ran (values of input variables and probability distributions for output variables). Include printout of source and documentation for any code you developed for your project.
9) Softcopy: Please provide softcopy of code, models, and reports you produced for this project.

BEST WISHES FOR A GREAT SEMESTER!!!

Fairfax, January 10, 2010.