OR/STAT 719 – CSI 775
Graphical Probability Models for Inference and Decision Making

Prof. Paulo C. G. Costa, PhD
Department of Systems Engineering and Operations Research
George Mason University
http://mason.gmu.edu/~pcosta

Course Description

Spring 2014

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: STAT 652, SYST/STAT 664, or permission of instructor
Co-requisites: None

Classes
** Wednesdays, from 7:20 p.m. to 10:00 p.m.*
** Room 1001 of the West building.*

Office hours
** Room 2227 of the Engineering Building.*
** Wednesdays, from 2:00 p.m. to 3:30 p.m., or by appointment.*

** Dr. Costa contact data: (703) 993-9989 / pcosta@gmu.edu*

Administrative

** Registration and drop without tuition penalty deadline: Jan 28th.*
** Drop with 33% tuition penalty: Feb 11th.*
** Final Drop deadline (66% tuition penalty): Feb 21st.*

Course Logistics

1. All course communication will be done via the Blackboard system. Students are expected to have access and be able to use the system before classes start. Blackboard is accessible via the MyMason portal at https://mymasonportal.gmu.edu/. Instructions for using the Blackboard system are provided in the "resources" link at the bottom of the portal page.
2. Volgenau School Computing Resources has answers to many questions about school systems on their web site: http://labs.vse.gmu.edu and will try to help you if have problems connecting to school computing systems. However, they will not provide assistance with general computing questions or course assignments. Please contact me if you have any questions about how to use software to complete your assignments.
3. Accommodations for disability: If you have a documented learning disability or other condition that may affect your academic performance you should: a) make sure this documentation is on file with Office for Disability Services (SUB I, Rm. 4205; 993-2474; http://ods.gmu.edu) to determine the accommodations you need; and b) let me know about your accommodation needs as soon as possible. If you have contacted the Center for Disability Services and are waiting to hear from a counselor, please keep me updated during the whole process.
4. Inclement weather: Class sessions may be cancelled due to inclement weather or other University emergencies. Check the Announcements area of the course website for updates.

Expected Behavior

1. Attendance in class is essential. Information will be presented that will not necessarily be in the book, and is almost certain to be in both the midterm and final exams.
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, tablet, pager, etc., in silent mode. Should you find yourself in *extreme* need to answer an incoming call, just leave the room to do so.
4. With a few exceptions, almost all of the course deliverables are submitted electronically (e.g. class-work and homework), scheduled in advance, and with some flexibility for students to change. Should any scheduled event impact a student's participation in class activities and assignments, it is the student's responsibility to coordinate with me in advance.

5. Students are permitted to interact on homework 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.
6. Late assignments, when properly justified, will receive reduced credit in accordance with the late assignment policy (below in this document). No points will be awarded if homework is turned in after solutions have been posted.
7. The exam dates and scheduling provided below are tentative, and it is the students' responsibility to keep abreast of changes.
8. Make-up exams will *only* be given for extreme situations, and *only* if I am contacted before the exam is given and full arrangements are established. Full adherence to this policy is the responsibility of the student.
9. Religious observances are one common example of events that might impact students' activities. Students are responsible for planning ahead. Please, refer to the GMU's calendar of religious holidays at http://ulife.gmu.edu/religious_calendar.php.
10. Academic Policy: All academic policies as given in the Honor System and code will be strictly followed. These are available at http://catalog.gmu.edu/content.php?catoid=19&navoid=4113.
11. General Policies: All general policies defined in the University Catalog are in place for this course. You can access those at http://catalog.gmu.edu/content.php?catoid=19&navoid=4114.
12. George Mason University is an Honor Code university. Please see the Office of Academic Integrity website (http://academicintegrity.gmu.edu/honorcode/) for a full description of the honor code and the honor committee process.

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

Course Outline:

Unit 1: Course Overview and Introduction

Unit 2: Graphical Probability Models

Unit 3: Representing Knowledge in an Uncertain World

Unit 4: Belief Propagation in BNs: Propagation in Junction Trees

Unit 5: Learning Bayesian Networks from Data

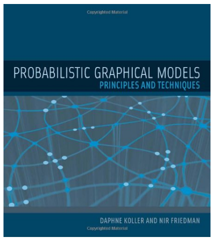
Unit 6: Belief Propagation in BNs: Other Inference Algorithms

Unit 7: Knowledge Engineering and Modeling

Unit 8: Planning and Decision Making

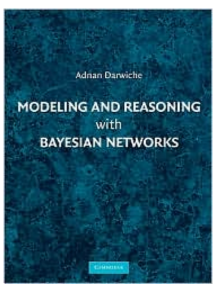
Textbook

The course will be based on the class notes. However, the following books are recommended:



Probabilistic Graphical Models: Principles and Techniques
Daphne Koller and Nir Friedman
The MIT Press, 1st Edition (July 31, 2009). 1280 pp.
ISBN-10: 0262013193.
ISBN-13: 978-0262013192.

Bayesian Networks and Decision Graphs
Thomas Nielsen and Finn Jensen
Springer; 2nd edition (July 4, 2007). 447 pp.
ISBN-10: 0387682813.
ISBN-13: 978-0387682815.



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

Other recommended books include:

- Learning Bayesian Networks (Paperback), Richard E. Neapolitan. Prentice Hall; illustrated edition (April 6, 2003). ISBN-10: 0130125342.
- Bayesian Artificial Intelligence, 2nd edition. Kevin Korb and Ann Nicholson. Chapman and Hall, 2010.
- Causality: Models, Reasoning and Inference, Judea Pearl. Cambridge University Press; 2nd edition (September 14, 2009). ISBN-10: 052189560X.

Please, note that all of the above books are available from the GMU Library. Also, the second recommended book (Nielsen-Jensen, 2007) is available as an online resource, so students can download it directly from the GMU Library website.

Grading

The grading structure of this course is as follows:

- Assignments (20% of grade)
- Midterm (25% of grade)
- Final Exam (25% of grade)
- Team 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 on how well students are understanding the subject.

You are not prevented from working with your peers on the class work and homework 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 had a hard time with the exams.

Assignments must be submitted via Blackboard and can be of two types:

***Homework Assignment**:* Each homework assignment is out of 100 points. Unless stated otherwise, I will present the solutions at the beginning of the next class after the assignment was handed. If you handle your assignment after it is due but before I present the solutions you can earn a max of 70 points. An assignment handled after the solutions are posted will yield 0 points.

***Tests, Quizzes, or Challenges**:* These are conducted in class and each will be out for an amount of points to be disclosed prior to the class. The details of each test, quiz, or challenge will be explained during its respective announcement.

Files should be named with the following convention:

OR719_AssigmentTypeAndWeek_LastnameFirstname.
Examples: OR719_Hwk2_DoeJohn, OR719_ClassWork2_PoppinsMary, etc.

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

Both the Midterm and the Final exams will be take-home, individual work (Honor code applies). Details will be announced later.

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. UnBBayes-MEBN is a free, java-based, open source package that has many features (only so far to implement MEBN), but it is in beta state at this point. 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 5. 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 5/7 7:20 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 others have done? 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!!!

Sunday, January 3, 2014.