Course Information

        CSC529: Advanced Data Mining
        Winter 2015
        Wednesday 5:45PM-9:00PM
        Loop Campus, CS&TC 00222
        Course Management System: http://d2l.depaul.edu

Instructor Information

        Instructor: Daniela Stan Raicu
        Office: CDM Center, Room 718
        Office Hours: Monday, 4:15-5:00pm, 9:00-9:45pm
                        Wednesday, 4:15-5:00pm, 9:00-9:45pm
                        Other times by appointment
        Phone: (312) 362-5512
        Fax: (312) 362-6116
        Email: draicu@cti.depaul.edu
        Home page: http://facweb.cs.depaul.edu/~dstan

Course Description

The course is for students with prior background in data mining or machine learning techniques. The course will cover advanced modeling techniques, such as ensemble learning, extended linear models, probabilistic graphical models, mixture and latent variable models, and matrix factorization. First the theoretical foundations of these techniques will be presented and augmented with in-class examples and homework problems. Second, the state-of-the-art research related to these techniques will be presented and augmented with paper reviews that highlight the practical applications of these advanced data mining techniques. Applications of the models will be presented in popular domains, including social computing and health informatics.

Course Learning Goals

At the end of the course, students should be able to:
- Understand the basic principles and computational tasks of knowledge discovery from data
- Understand how information in real world applications can be formulated and represented as different genres of data, such as matrices, sequences, data streams, graphs/networks
- Identify the major data mining challenges specific to different genres of data
- Apply the state-of-the-art data mining techniques that solve these problems
- Understand the recent trends and open directions in the field of data mining.

Recommended Books

- *Data Mining: Practical Machine Learning Tools and Techniques* by Witten, Frank, and Hall, 3rd Edition, ISBN 978-0-12-374756-0
  - This book has a focus of practical applications and the use of the WEKA toolkit.
- *Probabilistic Graphical Models, Principles and Techniques* by Daphne Koller and Nir Friedman, ISBN 978-0-262-01319-2
  - This book has a focus on theoretical foundations of probabilistic graphical models http://mitpress.mit.edu/books/probabilistic-graphical-models
- *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* by Hastie, Tibshirani, Friedman
  - This book has a focus on theoretical foundations of data mining; PDF available at http://www-stat.stanford.edu/~tibs/ElemStatLearn

Prerequisites

CSC 424 and (IS 567 or ECT 584 or CSC 578)

1/6/2015

Grading
The homework/programming assignments will be worth 40% of the course grade, paper reviews will be 20%, and the final project will be worth 40%.

The summary of the weights of each assignment for contributing to the final grade is as follows:

Assignment	Weight in final grade
Homeworks & Programming Assignments	40%
Paper Review	20%
Final Project	40%

The final grade will be assigned according to the following scale:

Percentage Grade	Letter Grade	Manner of fulfillment
95-100	A	Excellent
90-94	A-	
85-89	B+	Very Good
80-84	B	
75-79	B-	
70-74	C+	Satisfactory
65-69	C	
60-64	C-	Poor
55-59	D+	
50-54	D	
0 – 50	F	

Homework/Programming Assignments, Papers' Reviews, and Final Project Policies

Homework/programming assignments
There will be 4 homework/programming assignments, which are due at the beginning of class one or two weeks after they are assigned. Late assignments will be accepted up to one lecture later than the assigned due date with a 25% penalty – this penalty will be assessed in full to assignments turned in from the end of class on the day that the assignment is due up until the beginning of next lecture. No assignments will be accepted beyond the beginning of class one lecture beyond the due date. Any submitted documents (homeworks, reports, etc) must be typed and submitted through D2L website: http://d2l.depaul.edu It is your responsibility to check that your homework/project files are uploaded correctly on D2L and you should always keep a copy of your submission.

Extra credit points will be given for additional problems in assignments and active participation in the lectures and Discussion Forum.

Paper Review:
Throughout the quarter, the students will be also provided with a list of research papers related to the theory discussed in class. Each student will have to review these papers and participate in their discussion. Two paper reviews will have to be submitted and they will represent 20% of the final grade.

The paper readings will be selected from the recent literature in major journals and conference proceedings in the field of data mining. They include but are not limited to the ACM KDD Conference on Knowledge Discovery and Data mining (KDD), the IEEE International Conference on Data Mining (ICDM), the

SIAM International Conference on Data Mining, the IEEE Transactions on Knowledge and Data Engineering!(TKDE), and the ACM Transactions on Knowledge Discovery from Data (TKDD).

Final Project
The purpose of the final project is to demonstrate students' ability to apply the knowledge and the techniques learned during this course. The final project for this class is more extensive analysis task, chosen by students from among the topics we discuss. Final projects will include a presentation to the rest of the class at the end of the quarter, in place of a final exam. As part of the final project, students will also be asked to critique their classmates' projects. These critiques will be collected by me, collated, and passed on anonymously to the presenter.

Whenever it is possible, it is recommended that the DL students attend the final presentations to participate in the live discussions of the final projects and to complete critiques of the other projects. However, appropriate accommodations will be made for the DL students not being able to give the presentations in class; the DL students will still have to submit their critiques on the other projects.

Deliverables for the final project:

Proposal (February 11th): One page proposal describing the problem, the proposed approach, and at least three references other than text book or class notes.

Presentation (March 11th & March 18th): Each project is to be presented using PowerPoint, and the PPT file will have to be submitted to be published on course web site.

Report (March 18th): The report will be written in the format of a paper (abstract, introduction, literature review, methodology, results, discussion, conclusions and future work). The literature review for the final report consists of reading and summarizing about 5 to 6 published papers on the review topic. While the internet can serve as a good source of information, the DePaul Library also has extensive holdings, most of them available electronically.

Software
The use of Matlab will be taught in class. There will be also a lab session scheduled during the second week of school to familiarize the students with Matlab and the Statistics Toolbox. Students may use any data mining tool of their choice when completing class assignments.

Attendance
It is expected that you will attend every class; it is the single most important action you can take in mastering the course objectives. You are responsible for all material covered, assignments delivered or received, and announcements made in class sessions that you miss. For distance learning students, this means viewing the classes in a timely manner, participate in the discussion forum, and being sure to email or call in any questions that you have.

For online students:
Recordings of each lecture will be available a few hours after the "live" class, and can be found at the course website https://col.cdm.depaul.edu. Online students are expected to watch the lectures every week and to keep up with the course information posted on the course website.

Email
Email is the primary means of communication between faculty and students enrolled in this course outside of class time. Students should be sure their email listed under "demographic information" at http://campusconnect.depaul.edu is correct.

Changes to Syllabus
This syllabus is subject to change as necessary to better meet the needs of the students. Significant changes are unlikely, and will be thoroughly addressed in class. Minor changes, especially to the weekly agenda, are

1/6/2015

possible at any time. If a change occurs, it will be thoroughly addressed during class and posted under Announcements in COL.

Class Cancellation
Unless DePaul University closes because of weather, we will have class.

Cell Phones/On Call
 If you bring a cell phone to class, it must be off or set to a silent mode. Should you need to answer a call during class, students must leave the room in an undisruptive manner. Out of respect to fellow students and the professor, texting is never allowable in class. If you are required to be on call as part of your job, please advise me at the start of the course.

School policies:

Online Course Evaluations
Instructor and course evaluations provide valuable feedback that can improve teaching and learning. The greater the level of participation, the more useful the results. As students, you are in the unique position to view the instructor over time. Your comments about what works and what doesn't can help faculty build on the elements of the course that are strong and improve those that are weak. Isolated comments from students and instructors' peers may also be helpful, but evaluation results based on high response rates may be statistically reliable (believable). As you experience this course and material, think about how your learning is impacted. Your honest opinions about your experience in and commitment to the course and your learning may help improve some components of the course for the next group of students. Positive comments also show the department chairs and college deans the commitment of instructors to the university and teaching evaluation results are one component used in annual performance reviews (including salary raises and promotion/tenure). The evaluation of the instructor and course provides you an opportunity to make your voice heard on an important issue – the quality of teaching at DePaul. Don't miss this opportunity to provide feedback!

Beginning Winter quarter 2014-15 Online Course and Teaching Evaluations (OTEs) will be available to students via mobile devices (phone, tablet, laptop). This new modality offers the possibility to administer the evaluations in class, for instructors teaching face-to-face courses.

Academic Integrity and Plagiarism
This course will be subject to the academic integrity policy passed by faculty. More information can be found at http://academicintegrity.depaul.edu/.
The university and school policy on plagiarism can be summarized as follows: Students in this course should be aware of the strong sanctions that can be imposed against someone guilty of plagiarism. If proven, a charge of plagiarism could result in an automatic F in the course and possible expulsion. The strongest of sanctions will be imposed on anyone who submits as his/her own work any assignment which has been prepared by someone else. If you have any questions or doubts about what plagiarism entails or how to properly acknowledge source materials be sure to consult the instructor.

Withdrawal
Students who withdraw from the course do so by using the Campus Connection system (http://campusconnect.depaul.edu. Withdrawals processed via this system are effective the day on which they are made. Simply ceasing to attend, or notifying the instructor, or nonpayment of tuition, does not constitute an official withdrawal from class and will result in academic as well as financial penalty.

Retroactive Withdrawal
This policy exists to assist students for whom extenuating circumstances prevented them from meeting the withdrawal deadline. During their college career students may be allowed one medical/personal administrative withdrawal and one college office administrative withdrawal, each for one or more courses in a single term. Repeated requests will not be considered. Submitting an appeal for retroactive withdrawal does not guarantee approval.

College office appeals for CDM students must be submitted online via MyCDM.

The deadlines for submitting appeals are as follows:

Autumn Quarter: Last day of the last final exam of the subsequent winter quarter
Winter Quarter: Last day of the last final exam of the subsequent spring quarter
Spring Quarter: Last day of the last final exam of the subsequent autumn quarter
Summer Terms: Last day of the last final exam of the subsequent autumn quarter

Excused Absence
In order to petition for an excused absence, students who miss class due to illness or significant personal circumstances should complete the Absence Notification process through the Dean of Students office. The form can be accessed at http://studentaffairs.depaul.edu/dos/forms.html. Students must submit supporting documentation alongside the form. The professor reserves the sole right whether to offer an excused absence and/or academic accommodations for an excused absence.

Incomplete
An incomplete grade is a special, temporary grade that may be assigned by an instructor when unforeseeable circumstances prevent a student from completing course requirements by the end of the term and when otherwise the student had a record of satisfactory progress in the course. CDM policy requires the student to initiate the request for incomplete grade before the end of the term in which the course is taken. Prior to submitting the incomplete request, the student must discuss the circumstances with the instructor. Students may initiate the incomplete request process in MyCDM.
•   All incomplete requests must be approved by the instructor of the course and a CDM Associate Dean. Only exceptions cases will receive such approval.
•   If approved, students are required to complete all remaining course requirement independently in consultation with the instructor by the deadline indicated on the incomplete request form.
•   By default, an incomplete grade will automatically change to a grade of F after two quarters have elapsed (excluding summer) unless another grade is recorded by the instructor.
•   An incomplete grade does NOT grant the student permission to attend the same course in a future quarter.

Students with Disabilities
Students who feel they may need an accommodation based on the impact of a disability should contact the instructor privately to discuss their specific needs. All discussions will remain confidential.
To ensure that you receive the most appropriate accommodation based on your needs, contact the instructor as early as possible in the quarter (preferably within the first week of class), and make sure that you have contacted the Center for Students with Disabilities (CSD) at:
Student Center, LPC, Suite #370
Phone number: (773)325.1677
Fax: (773)325.3720
TTY: (773)325.7296