CIS 5930-02

Advanced Data Mining

Spring 2019

Syllabus

Course Website: http://www.ifmlab.org/courses.html

Instructor: Prof. Jiawei Zhang
Office: 169 James Love Building
Phone: (850) 645-2174
Email: jzhang@cs.fsu.edu
Office Hours: M/W 5:00-6:00PM

Course Teaching Assistant: Yuxiang Ren
Email: yren@cs.fsu.edu
No Office Hours for TA

Course Information:
- **Date**: January 7, 2019 – April 26, 2019
- **Time**: Monday/Wednesday 3:35-4:50PM
- **Venue**: Lov 0301

Course Description:
The course on Advanced Data Mining focuses on selective areas of importance in data mining and machine learning. Data mining and machine learning are both the recently emerged hot topic in AI studies, and solutions developed in the research hold substantial impacts in many important applications. Selective topics will be covered in the Advanced Data Mining course.

Topics Covered:
- Network Mining & Graph Mining
- Deep Learning
- Broad Learning
- Statistical Learning
- Text Mining
- Recommender System
- Misc.

Prerequisites:
You are expected to have background knowledge in Data Structure, Algorithm, Discrete Mathematics. You will also need to be familiar with basic Linear Algebra, basic Statistics, and can master at least one programming language and have programming experiences.

Textbook:

No required textbook, but some Reference Books are recommended.

[1] Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to Data Mining. 2nd Edition. Pearson.

[2] Reza Zafarani, Mohammad Ali Abbasi, Huan Liu. Social Media Mining: An Introduction. Cambridge University Press.

[3] Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press.

Course Format: No exams or homework in this course.

The objective of this course is to familiarize students with the latest research topics related to data mining and machine learning. Course activities include 1) paper reading and paper review; 2) paper presentation and discussion; and 3) research oriented course paper writing.

- **Paper Reading and Paper Review**: Each class will discuss one academic paper. Before class, the students should read the paper to be presented in class, and write a short review (no longer than 1 page) for the paper. The review should cover: (a) a summary of the paper; (b) 3 strong points of the paper; (c) 3 weakness of the paper; (d) potential ideas of future works; and (e) questions about the paper you would like to ask the presenter. Students need to submit the review **before** class starts. Some external reference papers/articles can be provided for students to understand the papers presented in class (no need to present or write reviews for external reference papers).
- **Paper Presentation and Discussion**: Each student will present 2 papers selected during the period of this course, and student will not need to write the reviews of papers you choose to present. Paper presentation is expected to take **45-50** minutes to cover (a) background knowledge, (b) problem description, (c) definitions, (d) ideas, (e) proposed methods and techniques, (f) experiments, (g) results, (h) related works, and (i) potential future works, etc. Additionally, we will have 15-20 minutes for in-depth discussion about the paper and Q&A with the audiences. Students can have Q&A during and after the presentation. Student needs to submit the slides **before** the end of presentation day.
- **Paper Proposal and Paper Writing**: Each student needs to finish an independent research-oriented academic paper in this course, based on the problems studied in this course about data mining and machine learning. The paper should be original work, not recycled published/submitted/on-going work with another faculty or classes. Student needs to submit a paper proposal (no longer than 2 pages) at the mid of the class to provide the planning for the paper. By the end of this semester of this course, student needs to submit the final paper (10 pages in ACM double-column format). Student needs to submit the proposal and final paper before the deadlines.

Paper List and Schedule:

- The course presentation paper list and schedule is available at
  https://docs.google.com/spreadsheets/d/1dvOHuQq6wDldZODruxdDYp505IhxbL5aerjlviiI_gU/edit?usp=sharing
- Each student may need to sign **2** papers you are interested in to present in class. Please type in your name in the slots available (M/W only during the semester).
- External reference papers/articles will not be presented in class, and students can read them after class if you are interested in them. The external reference list is available at
  https://docs.google.com/spreadsheets/d/1hwcAZa3rg3t6xyFuD9IMOhUVqOOht54-MfWW6U4SmhY/edit?usp=sharing

Course Grading Policy:

- **In-class presentation: 30%**. Powerpoint presentation needs to be submitted on the day of the presentation, before 11:59PM (midnight) of your presentation day. Copying existing presentation from the web is regarded as plagiarism.

- **Course paper: 50%**. Single authored original work on data mining/machine learning. Not recycled published/submitted/on-going work with another faculties or classes.
  - **Paper proposal: 10%**, which will be due on March 15, 2019, 11:59PM (midnight).
  - **Final paper: 40%,** which will be due on April 26, 2019, 11:59PM (midnight).
- **Course participation and Q&A: 20%**. A summary/review of each in-class discussion paper needs to be submitted **before** each class starts (i.e., **before** 3:35PM M/W). During the class, presenter and audiences can have Q&A with the pre-prepared questions in the review report.
- **Final Grade**
  - A: 100-90, A-: 90-85;
  - B+ 85-80, B: 80-75, B-: 75-70;
  - C: 70-60;
  - F: 60-0.

Late Submission Policy:
- Late paper review submission will not be accepted. During the semester, you can miss up to **2** paper reviews without penalty.
- Late presentation slides submission will not be accepted.
- Late proposal submission will not be accepted.
- Late final paper submission will get your grade $* \ 1(t <= 24)e^{(-\frac{ln(2)}{12}t)}$, if you are t hours late.

Resources:
- Online MOOC Courses:
  - Data Mining (https://www.coursera.org/specializations/data-mining)
  - Machine Learning (https://www.coursera.org/learn/machine-learning)
- Paper Review Template (https://docs.google.com/document/d/1jVkhizzDj9z8kUIrp-mgUFoj_ccqZ01hBOZs3TGp204/edit?usp=sharing)
- Proposal and Paper ACM Template (http://www.acm.org/publications/proceedings-template)
  - Latex Template (http://www.acm.org/binaries/content/assets/publications/consolidated-tex-template/acmart-master.zip)
  - Windows Word Template (http://www.acm.org/binaries/content/assets/publications/consolidated-tex-template/acm_windows_word_template.zip)
  - Mac Word Template (http://www.acm.org/binaries/content/assets/publications/consolidated-tex-template/acm_mac_2016_word_template.zip)
- Python Machine Learning Toolkit: Scikit-learn (http://scikit-learn.org/stable/)
- Deep Learning Toolkit: Tensorflow (https://www.tensorflow.org)