CS6501: Text Mining Course Policy

Hongning Wang

CS@UVa

Goal of this course

- Discuss fundamental problems in text mining research
  - Building blocks of text mining algorithms
  - Wide coverage of many applications
    - Document classification/clustering
    - Topic modeling
    - Sentiment analysis/recommendation
- Get hands-on experience by developing practical systems/components
- Prepare students for doing cutting-edge research in text mining and related fields
  - Open the door to the amazing job opportunities in data science industry

Structure of this course

- Lecture based
  - Six major topics will be covered
    - E.g., NLP pipelines, classification/clustering models, and social network analysis
  - Introduce state-of-the-art large-scale text analytics techniques
    - E.g., MapReduce framework, Apache Spark and GraphLab

Prerequisites

- Programming skills – Important!
  - Basic data structures: CS 2150 or equivalent
  - Java is required for machine problems
    - Most open source packages are written in Java
  - Any language you choose for the rest of this course
- Math background
  - Probability
    - Discrete/continuous distributions, expectation, moments
  - Linear algebra
    - Vector, matrix, dot product
  - Optimization
    - Gradient-based methods

Grading policy

- Homework (30%)
  - Machine problems (~4)
- In-class quizzes (15%)
  - To review the learned concepts (~5)
- Paper presentation (20%)
  - Graded by peer-review
- Course project (35%)
  - Research/development-oriented
- **No** midterm/final exams!
- **No** curve will be applied in final grading!

Quizzes

- Format
  - True/False questions
  - Multiple choice questions
  - Short answer questions
- Schedule
  - After each major lecture topic
  - Will be informed one week before the quiz
- <u>Closed</u> book and <u>closed</u> notes
  - No electronic aids or cheat sheets

Paper presentation

- Let students present the state-of-the-art research related to text mining
  - Choosing from recommended readings, or your favorite paper outside the list
  - 15-mins presentation including 2-mins Q&A
  - One paper one student
  - Register your choice early, first come first serve
  - Will be graded by the instructor and other students

Course project

- Appreciate research-oriented problems or "deliverables"
  - Work in groups (*not required*)
    - Up to 3 students
  - Project proposal (20%)
    - Discuss your topic with peers or the instructor first
    - Written report
  - Project report (40%)
    - Due **before** the final presentation
  - Project presentation (40%)
    - 15-mins in-class presentation
    - 5-mins Q&A

Deadlines

- Machine problems
  - Due in 7-days after posting

- Paper presentation
  - Sign up is due in **the end of 6th week**
  - Presentation starts on the **7th week**

- Project
  - Proposal due in **the end of 5th week**
  - Presentation in the last week of the semester

Late policy

- Homework
  - Everyone will have one chance to ask for extension (extra three days after deadline)
  - Request must be made **before** the deadline!
- Quizzes
  - **No** make-up quizzes unless under emergency situation
- Paper presentation
  - Must be presented on your selected date
- Course project
  - Proposal due early in the semester (~5th week, no extension)
  - Final report due before presentation (no extension)

Late policy

- If submit after the deadline without granted extension
  - 15% late penalty will be applied

Fairness among all the students will be guaranteed!

Contact information

- Lecture
  - Instructor: Hongning Wang
  - Time: Tuesday/Thursday 9:30am to 10:45am
  - Location: Rice Hall 340
- Office hour
  - Instructor's
    - Time: Thursday 11am to 12pm
    - Location: Rice Hall 408
- Course website
  - Website: http://www.cs.virginia.edu/~hw5x/Course/Text-Mining-2015-Spring/_site
  - Piazza: https://piazza.com/class/i3zaobi7ivhlx

Thank you!

QUESTIONS?