CS 6890 Syllabus

Social Media Mining
Utah State University, Fall 2014
Tuesday and Thursday 10:30am - 11:45am, Old Main (MAIN) 406

Instructor:
Kyumin Lee, MAIN 401D, (435) 797-8420, kyumin.lee@usu.edu
Office Hours: 11:45am - 12:45pm Tu/Th, or by appointment

Course Summary:
This course introduces research topics in social media mining and addresses relevant theoretical foundations, methods, and tools. The course topics include, but are not limited to: link analysis, social media, geography and the social web, crowdsourcing, big data management (e.g., MapReduce), privacy and security issues related to social systems (e.g., Facebook, Twitter and Google+), etc. Students will learn how to collect, parse and analyze social media data with data mining techniques (e.g., classification and clustering).

By the end of the semester you will be able to:
• Define and explain the key concepts, methods, and tools relevant to social media mining
• Design, implement and evaluate core algorithms and approaches
• Identify the salient features and apply recent research results in social media mining
• Have hands-on experience by performing programming assignments and a project that will reinforce the theoretical aspects covered in lectures.

Mapped objectives in IDEA:
• Learning fundamental principles, generalizations, or theories
• Learning to apply course material (to improve thinking, problem solving, and decisions)
• Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course

Communication:
We are going to use Google Groups for all course communication, so you should check our Google Group often. If you have a question to discuss with everyone, please post it to the group! If you have found a cool link to share with classmates, share it to the group. We will monitor the group and provide feedback. But everyone is encouraged to contribute.

Prerequisites:
Graduate classification or approval of instructor.

Required Text:
No primary textbooks, but we will refer to the following textbooks:
In addition, we will read some research papers from top computer science conferences and journals.

Class Format:
During the half of this semester, I will present an overview of social media mining and related algorithms and techniques. The remaining weeks will be devoted to intense paper discussion.

Grading:
The course grading policy is as follows:
- 5% In-class discussion
- 10% Assignments
- 15% Paper presentation
- 20% Midterm
- 15% Project proposal
- 15% Project presentation
- 20% Final project deliverable


Assignments: There are two programming assignments with/without written assignments. Each assignment is proportion to 5% of your grade. You will have total 2 late days during the semester. You can use up to 2 late days for each assignment without penalty. After you consume the total 2 late days for an assignment, then you will get penalty proportion to extra late days (e.g., 10% off for the next late day, 20% for the next two late days and so on).

For example, you submitted your first assignment 2 days late. You will not get any penalty, but use 2 out of 2 late days. Or if you submit your first assignment 3 days later than due date, you will use 2 late days (again up to 2 late days for an assignment), and get 10% off penalty because of the third late day.

For each assignment, we will NOT accept your solution more than 5 days late.

You may discuss an assignment with your colleague, but you should write a program by yourself and should NOT copy and paste your colleague's program. If you discussed an assignment with your colleague, explicitly report the colleague's name and what you discussed in your submission.

Midterm: The midterm exam is closed book and will be held in class. You may bring one standard 8.5" by 11" piece of paper with any notes you think appropriate or significant (front and back). No electronic devices allowed.

The most significant portion (50%) of your grade is based on the project, for which you may work in teams of up to two persons. The general goal of research is to contribute new knowledge. Thus, it is important to ask yourself what research question(s) you aim at answering and what challenges you aim at solving.

Project: In general, a project should include both: (i) an empirical evaluation of an algorithm or model on an interesting dataset, in order to better understand these methods, and possibly further
improve them; and (ii) a proposal for a new algorithm and model, including a comparison with a baseline.

When picking a topic, try to ask yourself the following questions:

- Is the topic addressing an important problem? Would anyone care about it if you solve the problem?
- To what extent has previous research work addressed this problem? And what remains unknown?
- Do you have any idea at all about how to solve the problem? If not, can you reformulate the problem to make it easier?
- Would you be able to evaluate your solution? That is, how can you demonstrate your solution is good and solves the problem well?

Proposal: The emphasis of the course is on recent and current scholarly material. We will read a number of papers from the past few years and these papers should stimulate you to think about unexplored avenues of research. For the project proposal, you should pick two or three related papers as the basis for your proposal. These related papers will serve as the research grounding for your proposal. So first in your proposal, you must address these questions (in about 2 or 3 pages, single column, normal fonts and margins):

- What is the key technical content and interesting ideas behind the papers?
- How do the papers relate to the topics presented in the course?
- What are the strengths of the paper? What are its weaknesses? And how might you go about improving on the weaknesses?

In the remainder of the proposal (about 2 pages), I expect you to dig deeply into your proposed work (which should naturally build on the background materials you’ve already covered).

- What is your research question? Clearly define the research problem/question.
- Why is this an interesting question to ask and why would we care about the answer to this question or a solution to the problem?
- Has any existing research work tried to answer the same or a similar question, and if so, what is still unknown?
- How do you plan to work out the answer to the question. (At the proposal stage, you are only expected to have a sketch of your methods.)
- How would you evaluate your solution. That is, how do you plan to demonstrate that your solution/answer is good or is reasonable.
- A rough timeline to show when you expect to finish what. List a couple of milestones.

You should plan to write a proposal of 3 ~ 5 pages. The project proposal is due on October 9 (Thursday) by 11:59pm. Email me your project proposal in PDF with a pithy filename. One per team.

Final Project Deliverable: At the end of the semester, you will deliver a final paper and participate in a presentation (maybe with a demo) session.

- The presentation session will be held on December 2 and 4. 15% of your project grade is based on the project presentation. Your team’s grade will be based on the clarity of the presentation itself, audience feedback, and your answers to my questions.
- The final report is due on December 12, 2013 by 11:59pm. You should email me your final report in PDF with a pithy filename. The paper should be in the standard ACM conference
template http://www.acm.org/sigs/publications/proceedings-templates (double-column) and up to 10 pages maximum. The paper counts for 20% of your project grade. You should write your report as if you were writing a conference paper. You should address the same questions as those you have addressed in the proposal, only with more details, especially regarding some of the challenges that you need to solve and your experimental results. You should also include your conclusions from the study and point out how your work can be further extended (i.e., future work).

Cheating Statement
Academic Dishonesty: This course adheres to the cheating policy for courses in the Department of Computer Science posted on the bulletin board outside the CS office on the 4th floor of Old Main and posted online at http://cs.usu.edu/htm/cheating-policy/.

ADA Compliance
Students with ADA-documented physical, sensory, emotional or medical impairments may be eligible for reasonable accommodations. Veterans may also be eligible for services. All accommodations are coordinated through the Disability Resource Center (DRC) in Room 101 of the University Inn, (435)797-2444 voice, (435)797-0740 TTY, or toll free at 1-800-259-2966. Please contact the DRC as early in the semester as possible. Alternate format materials (Braille, large print or digital) are available with advance notice.