Course Syllabus: COMPSCI 690V

Visual Analytics

Course Description: In this course, students will work on solving complex problems in data science using exploratory data visualization and analysis in combination. Students will learn to deal with the Five V's: Volume, Variety, Velocity, Veracity, and Variability, that is with large data, complex heterogeneous data, streaming data, uncertainty in data, and variations in data flow, density and complexity. Students will be able to select the appropriate tools and visualizations in support of problem solving in different application areas. The course is a practical continuation of CS590V - Data Visualization and Exploration and focuses on complex problems and applications. It does not require CS590V. The data sets and problems will be selected mainly from the IEEE VAST Challenges, but also from the KDD CUP, Amazon, Netflix, GroupLens, MovieLens, Wiki releases, Biology competitions and others. We will solve crime, cyber security, health, social, communication, marketing and similar large-scale problems. Data sources will be quite broad and include text, social media, audio, image, video, sensor, and communication collections representing very real problems. Hands-on projects will be based on Python or R, and various visualization libraries, both open source and commercial.

3 credits.

Required Textbooks: None. The instructor will provide Readings.

Required Software: The assignments will be split between both open source and commercial analysis systems such as Python, R, SAS, and other analytic/computational systems, as well as data visualization and exploration systems. All required software will be freely available, either as open source software or by the vendors for student use.

Course Website: The course website will be hosted on Moodle at https://moodle.umass.edu/. The course website will host lecture notes, assignments, and pointers to readings and videos. We will use Piazza (https://piazza.com/ Change this link to course specific link when course starts) for a course discussion forum.

Grading Plan: The coursework will include assignments, a midterm exam, and a final project.

Graded Homework Assignments	50%
Midterm Exam	20%
Final Project	30%

Grading Scale:

A	90-100
B	80-89
C	70-79
D	60-69
F	0-59

Accommodation Statement: The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements.

Academic Honesty Statement: Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (http://www.umass.edu/dean_students/codeofconduct/acadhonesty/).

Course Schedule:

Week 1: Introduction

- Course logistics
- How do analysis and visualization work in tandem in solving complex data science problems?
- Example tools
- VAST Challenge problem (single data type)

Week 2: Dealing with classical data

- Types of data

- Exploratory visualizations of classical databases
- Comparing visualizations systems (open source and commercial)

Week 3: Dealing with classical data

- Exploratory analytics
- Comparing analytic/computational systems (open source and commercial)
- Integrating visualization and analytics
- VAST Challenge problem (single data type, larger data)

Week 4: Selected IEEE VAST Challenges

- Classifying the VAST Challenges, the KDD, Amazon, Netflix and Biology Challenges
- Comparative evaluations: identify what data science approaches will be needed to attach the problems.
- Identify what tools would be most appropriate and what teams would look like to solve the problems.
- Groups select one of the VAST Challenge or one of the other data sets problems

Week 5: Dealing with text and time

- Identify semantic ontologies that support heterogeneous data
- Compare analysis vs visualization for one of the text/time VAST Challenges or other data set.

Week 6: Graphs and time

- Using one of the VAST Challenges that deals with graphs and time extend the class ontology
- Compare analysis and visualization results
- Identify where analytics is most useful and how visualizations support computational steering

Week 7: Text and Documents

- Using one of the VAST Challenges that deals with text, emails or web pages over time extend the class ontology
- Compare analysis and visualization results
- Identify where analytics is most useful and how visualizations support computational steering

Week 8: Audio and video

- Using one of the VAST Challenges that deals with imagery and video, or using one of the other data sets that deals with imagery and video extend the class ontology

- Compare analysis and visualization results
- Identify where analytics is most useful and how visualizations support computational steering
- Design an extension to these tools that would support dealing with imagery or video over time

Week 9: Dealing with data that does not fit into memory

- Data and dimensional reduction with guarantees (or constraints)
- Alternative algorithms with guarantees
- Sampling with guarantees

Week 10: Streaming data

- Real time databases
- Task driven and exploratory approaches
- Explore and solve one of the more complex heterogeneous data sets problems

Week 11-12: Project development

Week 13: Final Project Presentations

- Students present final project solutions in class
- Students generate debrief report using IEEE CG&A format