SOSC 4300 / SOSC 5500: Computational Social Science
Spring 2023
Lecture Time: Monday 9:00 - 11:50
Lecture Room: LSK 1033

This version prepared on Jan 27, 2023

Instructor
ZHANG, Han
Office Room 2379, Academic Building, Lift 15
Email zhangh@ust.hk
Office Hour Monday 16:00 - 17:00

Teaching Assistant
CHEN, Pei
Room 3001, Academic Building, Lift 4
pchenam@connect.ust.hk

• Course homepage is: https://github.com/HKUST-SOSC4300-5500/
• Syllabus will be regularly updated at https://github.com/HKUST-SOSC4300-5500/Lecture-Material

Prerequisites
• Students are expected to be familiar with the materials covered in basic statistics (e.g., SOSC 2400 for UG students and SOSC 5090 for PG students). Students with statistics knowledge but do not meet prerequisite can seek instructor’s approval for enrollment.
• Students should also have basic literacy in at least one statistical programming language. We will use R and Python in tutorials. You can also use other programming languages such as Matlab, Julia, etc., as long as you can finish course assignment and projects with the codes.

Goals

Upon finishing the course, students should be able to:

1. Describe the opportunities and challenges of social research in the age of big data
2. Evaluate research on social phenomena from different fields, including social sciences and computer science/data science.
3. Practice the essential techniques to analyze social big data
4. Propose research questions that are suited to be examined by computational methods with big data
5. Write a research article that utilizes the techniques and methods of computational social science to address social science problems, or design a project that use computational social science to address some real-world problems.
Grading

Your score will be accessed based on the following five components (no mid-term and final exams):

<table>
<thead>
<tr>
<th>Component</th>
<th>%</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance and participation</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Homework assignments</td>
<td>30%</td>
<td>Two weeks</td>
</tr>
<tr>
<td>Literature review Report</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Literature review Presentation</td>
<td>5% (5 min)</td>
<td>mid April</td>
</tr>
<tr>
<td>Final Paper/Project Presentation</td>
<td>15% (15 min)</td>
<td>May 8</td>
</tr>
<tr>
<td>Write-up</td>
<td>35%</td>
<td>May 22</td>
</tr>
</tbody>
</table>

Homework, literature review, and final paper/project need to be submitted on Github. We will cover the steps in the first lecture.

Grouping

- You should finish all tasks in groups
  - If there is any MPhil or PhD student in a group: max group size is 2
  - Otherwise: 3 to 4 in a group (e.g., 4 UG in a group)
- Finish grouping by Week 3

Attendance and participation in class activities

- Based on class attendance and involvement in lecture and tutorial. you are expected to be either able to answer questions about the assigned readings or ask questions about the parts you did not understand. If you are uncomfortable speaking up in class, send the question in Zoom’s chat window, post them on Canvas’s discussion forums, come to my office hours, or send your questions via e-mail.

Homework assignments

- There will be 3 assignments to test your knowledge of applying and evaluating basic machine learning algorithms
- Each exercise is due in two weeks after the release of assignment.

Literature review

Select a research topic and summarize how past researchers have used computational methods and/or big data to study this particular research area.

- Some examples of research areas:
  - Sociology: internal migration, international migration, social inequality, race and ethnicity relations, happiness,
  - Political science: government performance, government policy (and its effectiveness), election, social movements
  - Economics: measuring economic growth with big data
  - History: historical development of an idea
  - Psychology: measuring personality with big data
  - Communication and information science: content and spread of fake news/hate speeches
• You are recommended to select a research areas that are similar to your final research paper. Students can discuss with instructors and TA for possible topics or feasibility.

• Your performances will be accessed by:
  – **Written Report**: limit your report to **8 pages, 12 points, double space**. Spell out clearly contributions of each group member in the first page of your report.
  – **Presentation of literature review (5 minutes)**: each student/group needs to present their literature reviews in class.

• Include the following items in written report and presentations:
  – What is the research area you have chosen, and why it’s important or interesting
  – How people studied it traditionally (e.g., what data they use, what methods they use), and what are limitations of traditional methods/data?
  – What are the advantages of using computational social science methods and data?
  – What are the shortcomings of using computational social science methods and data?

**Final paper/project**

You can choose to write a research final paper, or a project that analyze a “real-world” social science problem. The differences between two options lie in their intended audience: research final paper should talk to researchers, while project talk to lay audience. The paper/project needs to be performed in the same group for presentation. It’s recommended that you discuss your ideas with the instructor in early weeks of the course, during offices hours or through emails.

**Research final paper**  choose a research topic and write a research paper **using computational social science methods or digital data**. This research article should follow the format of a standard research article, with the following components: introduction; review of past studies; research methods and data; results; conclusions. Consider the articles you read in class and for literature review as good examples of research articles. You can also browse the GitHub repo and see the projects of previous year’s students.

**Project**  Focus on a real-world case. Develop a website or mobile app or software. Consider that you want to sell some social science ideas to layman using cool data analysis and visualization. Some ideas of cool demo/projects can be found here:

- [https://projects.fivethirtyeight.com/](https://projects.fivethirtyeight.com/)
- [https://github.com/matiasmascioto/awesome-soccer-analytics](https://github.com/matiasmascioto/awesome-soccer-analytics)
- [https://github.com/academic/awesome-datascience](https://github.com/academic/awesome-datascience)

**Evaluation**

• Every group need to do a presentation (**15 minutes**): follow a standard presentation style for academic talks.

• If you are writing a final paper: limit your report to **20 pages, 12 points, double space, including Tables, Figures and References**. You can write fewer pages if you feel necessary.

• If you are doing a project: based on how your classmates like your website/app/software.
Grading policies

- **Late delivery** of due items will be marked down 75% if received within 1 day of the due date, and 50% if received within 3 days of the due date; you will receive zero credit if the due item is not delivered within 3 days of the due date. Contact the instructor if there are rare unforeseen circumstances.

- If you want to dispute a grade, please submit your argument in writing along with your assignment. We will evaluate the merit of your argument as well as perform a full reassessment of your entire assignment. This means that your grade may end up lower than it was originally.

- Final papers are checked by anti-plagiarism software. Students should take steps to avoid plagiarism and copying. For confirmed cases of plagiarism, severe sanctions including but not limited to a failure grade may be imposed.

Schedule (Tentative)

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2023-02-06 Mon</td>
<td>Introduction; big data</td>
</tr>
<tr>
<td>2</td>
<td>2023-02-13 Mon</td>
<td>Prediction;</td>
</tr>
<tr>
<td>3</td>
<td>2023-02-20 Mon</td>
<td>Prediction; Evaluation</td>
</tr>
<tr>
<td>4</td>
<td>2023-02-27 Mon</td>
<td>Text (I)</td>
</tr>
<tr>
<td>5</td>
<td>2023-03-06 Mon</td>
<td>Text (II); supervised</td>
</tr>
<tr>
<td>6</td>
<td>2023-03-13 Mon</td>
<td>Text (III); embedding</td>
</tr>
<tr>
<td>7</td>
<td>2023-03-20 Mon</td>
<td>Text (IV); unsupervised</td>
</tr>
<tr>
<td>8</td>
<td>2023-03-27 Mon</td>
<td>Network; basics</td>
</tr>
<tr>
<td>9</td>
<td>2023-04-03 Mon</td>
<td>Network; small worlds</td>
</tr>
<tr>
<td>10</td>
<td>2023-04-10 Mon</td>
<td>NO CLASS (mid-term break)</td>
</tr>
<tr>
<td>11</td>
<td>2023-04-17 Mon</td>
<td>Causal Inference and Big Data: network as example</td>
</tr>
<tr>
<td>12</td>
<td>2023-04-24 Mon</td>
<td>Image data (or other elective topics)</td>
</tr>
<tr>
<td>13</td>
<td>2023-05-01 Mon</td>
<td>NO CLASS (holiday)</td>
</tr>
<tr>
<td>14</td>
<td>2023-05-08 Mon</td>
<td>Presentation</td>
</tr>
</tbody>
</table>

Weekly reading material

The course materials will be drawn from lecture slides and assigned readings. Readings are available at Canvas. You are required to read the readings before the start of each class. Optional readings are for students who are interested to read more on the topic.

Lecture 1: Digital Traces


Optional readings


Lecture 2: Prediction; algorithms


Optional Readings


Lecture 3: Prediction; evaluation


Optional readings:


Lecture 4: Text (I); basics


Optional readings:


Lecture 5: Text (II); Dictionary and Supervised

Lecture 6: Text (III) Word Embeddings


Optional readings


Lecture 7: Text (IV); Unsupervised Methods and Topic Models


Optional readings


Lecture 8: Network, small world, and agent-based modeling


Optional readings


Lecture 9: Network, strength of ties, and diffusion


Lecture 10: Causal inference with big data


Lecture 11: Image data