CS 533: Natural Language Processing

Logistics

Karl Stratos

Rutgers University
Course Information

• **CS 533: Natural Language Processing (NLP)**
  - Wednesday 12:00–3:00pm at Beck Hall 252
  - All materials at course website: [http://karlstratos.com/teaching/cs533spring20/cs533spring20.html](http://karlstratos.com/teaching/cs533spring20/cs533spring20.html)

• **Instructor: Karl Stratos** (legal name: Jang Sun Lee, or Jangsun Lee)
  - Office Hours: Wednesday 3:20–4:30pm at Tillett 111H

• **TA: TBD**

• **Use Canvas:**
  - [https://rutgers.instructure.com/courses/44246](https://rutgers.instructure.com/courses/44246)
    1. To ask questions regarding lectures/homeworks/projects (and answer yourselves), discuss ideas, find collaborators, etc.
    2. To submit assignments
    3. To find announcements
About

This course is *not* about

- Philosophy of language
- Linguistic phenomena
- Social impact of language

This course *is* about

- **Models, statistical techniques, and algorithms** for computationally processing language as **data**
Grading

- Project: 40%
  - Written report: 30%
  - Presentation: 10%

- Exam (in-class and open book): 30%

- Assignments: 20%

- Participation: 10%
Project

Submit a proposal by later in the semester, do the work, and submit a final report and give an in-class presentation.

Has to be

1. **Substantial**: cannot be done trivially in a few hours (e.g., nontrivial implementation and experiments)
2. **Original**: has not been done already (e.g., new problem formulation, techniques, applications)

Ideally the quality of conference papers in NLP.

More information to come

But start thinking about projects early on.
Assignments

- The main way to learn

  - Tentative plan: \(\approx 4\) assignments, each a mix of written problems and coding in Python

  - A1 is already out (due 2/4 11:59pm)
    - If you cannot do A1 comfortably, you probably do not have the background needed for this course.

  - Work individually (okay to discuss). Do not copy:
    1. Honor
    2. Meaningless (in-class exam)
Exam

- Topics covered in lectures, readings, and assignments

- Tentative date: April 1

- You will be fine if you understand lectures and readings, and do well in assignments
  - I will tell you what will be on the exam
  - Definitely prioritize your research project over the exam
## Tentative Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Topics</th>
<th>Readings</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1 (January 22)</td>
<td>Introduction to NLP, Language Modeling</td>
<td>Michael Collins notes on <em>n</em>-gram models and log-linear models</td>
<td>A1 [code] (Due 1/31)</td>
</tr>
<tr>
<td>Week 2 (January 29)</td>
<td>Deep Learning for NLP: Neural Language Modeling</td>
<td>Colah's blogs on deep learning and LSTMs, NLM papers using feedforward (Bengio et al., 2003), recurrent (Mikolov et al., 2010; Melis et al., 2018), and attention-based (GPT-2) architectures</td>
<td></td>
</tr>
<tr>
<td>Week 3 (February 5)</td>
<td>Deep Learning for NLP: Conditional Neural Language Modeling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 4 (February 12)</td>
<td>Deep Learning for NLP: An Overview of Other Techniques and Applications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 5 (February 19)</td>
<td>Structured Prediction in NLP: Tagging</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 6 (February 26)</td>
<td>Structured Prediction in NLP: Parsing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 7 (March 4)</td>
<td>Unsupervised Learning in NLP: Latent-Variable Models and the EM Algorithm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 8 (March 11)</td>
<td>Unsupervised Learning in NLP: Variational Autoencoders</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Spring Recess</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9 (March 25)</td>
<td>Unsupervised Learning in NLP: Pretrained Neural Text Representations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 10 (April 1)</td>
<td>Exam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 11 (April 8)</td>
<td>Special Topics: TBD (Information Extraction, Question Answering)</td>
<td></td>
<td>Project proposal due</td>
</tr>
<tr>
<td>Week 12 (April 15)</td>
<td>Special Topics: TBD (Dialogue, Grounding)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 13 (April 22)</td>
<td>Special Topics: TBD (Maximal Mutual Information Representation Learning)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 14 (April 29)</td>
<td>Project Presentations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Other Requirements

- Use LaTeX to write up assignment and project reports
  - Template for assignment: http://karlstratos.com/teaching/cs533spring20/assignment_template.tar.gz
  - Template for project: http://karlstratos.com/teaching/cs533spring20/latex-example.tar.gz

- Use Python and its libraries like NumPy, PyTorch for coding

QUESTIONS?