Final Project or HW5

Wei Xu
Your Two Choices

Choice 1 — complete HW5 (a programming homework) by yourself
  - Reimplement a classic Question Answering model (BiDAF; Seo et al. 2017) on the popular SQuaD dataset.
  - If you have limited experience with NLP/ML research, you should do this by default. We expect most students will choose this option.

Choice 2 — open-ended final project on research topics as a group of 2/3
  - If you know what you are doing, and have clear ideas what to do.
  - We expect more substantial work than HW5 x number of person.
  - Write a one-page research proposal —> get feedback —> complete the project —> write a 4-page final report, submit with code/data
HW5 — QA System

- Programming homework
  - BiDAF model: https://arxiv.org/abs/1611.01603
- We will discuss QA later in the course (tentatively, Week 12)

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: through contact with Persian traders
Why choose HW5?

- If you:
  - Have limited time, limited experience with research, don’t have any clear idea of what you want to do, need help from TAs...

- Then:
  - Do the HW5!
  - Most student should do it!

- Considerations:
  - HW5 is a clearly-defined problem.
  - It has starter code and support from TA...
Why choose a Course Project?

If you:

‣ Have some research project that you are exited about (and are possibly already working on), which substantively involves text data and machine learning models

‣ Have a lot of extra time and energy (> 10 hours per week)

‣ Want to see more of the process of defining a research goal, finding data and tools, working out new ideas, and evaluate it ...

‣ Want to join my research lab

Then:

‣ Do the course project!

‣ Team up! (group size of 2 or 3; if go solo, confirm w/ instructor first)
Project Proposal (one-page by 4/1)

- Find a relevant (key) research paper for your topic
- Write a summary of that research paper and what your takeaways
- Write what you plan to work on and how you can innovate
  - Relevant existing literature
  - What kind(s) of models you will use/explore; what is your baseline
  - The data you will use and how it is obtained
  - How you will evaluate success
  - An estimated timeline w/ milestones
- Use LaTeX template from ACL 2021.
- List all team members.

https://www.overleaf.com/latex/templates/instructions-for-acl-ijcnlp-2021-proceedings/mhxffkdwdymb
Think critically about a research paper
- What were the novel contributions or points?
- Is what makes it work something general and reusable?
- Are there flaws or neat details in what they did?
- How does it fit with other papers on similar topics?

How to do a good job on your project proposal
- You need to have an overall sensible idea.
- Do you have good data or a realistic plan to collect it
- Do you have a realistic way to evaluate your work
- Do you have appropriate baselines for comparisons
Final Project Writeup/Presentation

- **4-page writeup** due the day before final exam date (no late submission!)
- Use **LaTeX template** from ACL 2021
- Include references; statement of each group members’ contribution
- Writeup quality is very important to your grade!
- **X-minute oral presentation at the final exam time** ($X \in [5, 10]$)

Credit: Stanford CS224n
Some example research ideas ...

- Building a robust QA system which works on different QA datasets
- State-of-the-art NLP applied/adapted to non-English languages
- Better way to identify/collection African America English users/posts in Twitter
  "Demographic dialectal variation in social media: A case study of African-American English" by Su Lin Blodget, Lisa Green, and Brendan O’Connor (EMNLP 2016)
- Identifying Asian Hate Crime related information from social media
- LGBTQ — current gender prediction is usually limited to male/female
- and many more ...
Questions?