Final Project

Final Project

- **Groups Size:** 2-3 people; 1 is possible (email me for permission).
- **Submission (slightly changed grading schema):**
  - 2-page project proposal (5%, due 10/10)
  - 4-page midway report (5%)
  - 8-page final report (17%) + final oral presentation (3%)

  (detailed instructions on the submission will be released later)

- **Example project reports — see Stanford CS224's past projects**
  https://web.stanford.edu/class/cs224n/project.html

- **Prize:** We will give out 1-3 best project awards. 🏆

Final Project

- **Shared project** with other classes is allowed
  - project is expected to be accordingly bigger/better
  - clearly declare at the beginning of your report that you are sharing project (with which class)
- **External collaborators** (e.g. non CS7650 students, phd advisor) are also allowed
  - clearly describe in the report which parts of the projects are your work

Project Proposal (5%)

- Two pages total
- 1-page summary of a relevant (key) research paper for your topic
  - Bibliographical information,
  - Background (motivation, related work, why this work is important),
  - Contributions (what's new this paper added to the ongoing research conversation — new algorithms, new experimental results and analysis, new meta-analysis of old papers, new datasets, or otherwise?)
  - Limitations and discussion (every paper has limitations and flaws)
  - Why this paper? What is the wider research context?

Project Proposal (5%)

- 1-page summary of what you plan to and how you can innovate?
  - Main goal and motivation of your project — why it is cool? why it is useful?
  - What NLP tasks(s)?
  - What data?
  - What methods?
  - What baseline?
  - How will you evaluate your results?

Why Project Proposal?

- From Chris Manning —

Skill: How to think critically about a research paper

- What were the main novel contributions or points?
- Is what makes it work something general and reusable or a special case?
- Are there flaws or neat details in what they did?
- How does it fit with other papers on similar topics?
- Does it provoke good questions on further or different things to try?
  - Grading of research paper review is primarily summative

How to do a good job on your project plan

- You need to have an overall sensible idea (!)
- But most project plans that are lacking are lacking in nuts-and-bolts ways:
  - Do you have appropriate data or a realistic plant to be able to collect it in a short period of time
  - Do you have a realistic way to evaluate your work
  - Do you have appropriate baselines or proposed ablation studies for comparisons
  - Grading of project proposal is primarily formative

Why Project Proposal?

- From Jason Eisner —

https://www.cs.jhu.edu/~jason/advice/how-to-read-a-paper.html

https://www.cs.jhu.edu/~jason/advice/write-the-paper-first.html

Finding Research Topics

- Two basic starting points, for all of science:
  - **Nails** — start with a (domain) problem of interest and try to find good/better ways to address it than are currently known/used
  - **Hammers** — start with a technical method/approach of interest, and work out good ways to extend or improve it or new ways to apply it

Credit: Stanford CS224n, Chris Manning

Typical Project Types

- This is not an exhaustive list —
- 1) Find an application/task of interest and explore how to approach/solve it effectively, often with an existing model
  - Could be task in the wild or some existing dataset or shared task (e.g.. WNUT or SemEval, etc.)
  - Or dialogue system, QA system, …
- 2) Analyze the behavior of models or existing datasets
  - how the model represents linguistic knowledge or what kinds of phenomena it can handle or errors that it makes.
  - what linguistic phenomena/errors exist in the dataset, how they affect model performance.

Credit: Stanford CS224n, Chris Manning

Typical Project Types

- This is not an exhaustive list —
- 3) Create a new dataset, conduct some analysis, train a prediction model
  - for a new topic/task, or for an existing task but better way to create higher quality dataset
  - may involve some manual annotation
  - conduct some quantitive and linguistic analyses
- 4) Implement a complex neural architecture and demonstrate its performance on some data, especially for non-English data
- 5) Come up with a new or variant neural network model and explore its empirical success (but this has become harder since 2020 —)

Credit: Stanford CS224n, Chris Manning

Place to start?

- Look at ACL Anthology for NLP papers:
  - https://aclanthology.org/
- Also look at the online proceedings of major ML/Web conferences
  - ICLR, NeurIPS , ICML
  - SIGIR, Web Conference, ICWSM (https://www.icwsm.org/2021/)
- Look at online preprint servers, especially:
  - https://arxiv.org/
- Look for an interesting problem in the world!
  - Psycholinguistics, computational social science, journalism, …

Finding a Topic

- Turing award winner and Stanford CS emeritus professor Ed Feigenbaum says to follow the advice of his advisor, AI pioneer, and Turing and Nobel prize winner Herb Simon:

  “If you see a research area where many people are working, go somewhere else.”

- But where to go? Wayne Gretzky:

  “I skate to where the puck is going, not where it has been.”

(Slides 51-55: https://web.stanford.edu/class/cs224n/slides/cs224n-2022-lecture08-final-project.pdf)

Credit: Stanford CS224n, Chris Manning

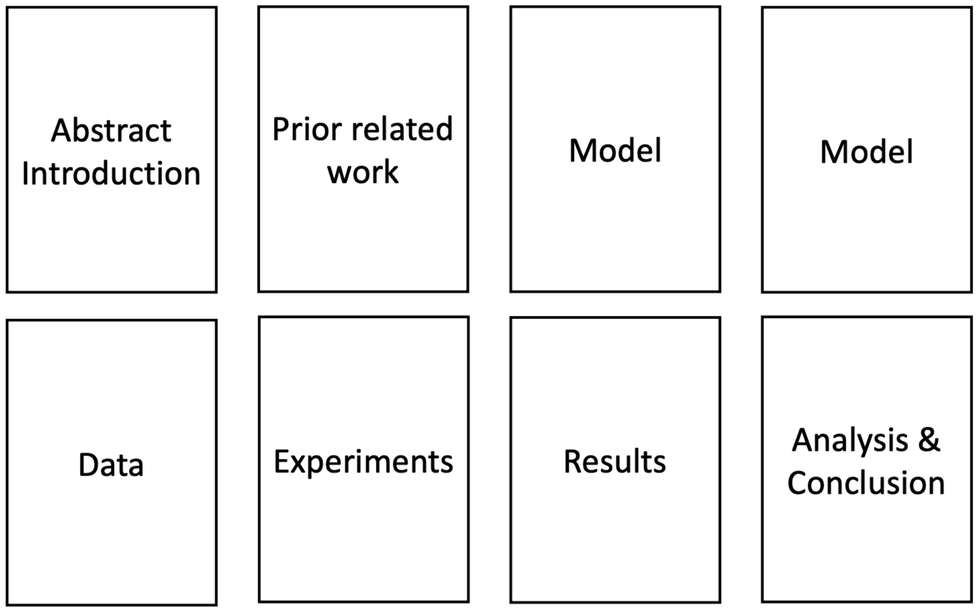
Finding Data

- Some people collect their own data for a project — **we like that!**
  - You may have a project that uses “unsupervised” data
  - You can annotate a small amount of data
  - You can find a website that effectively provides annotations, such as likes, starts, rating, responses, etc.
  - Look at research papers to see what data they use, how they get it
- Many others make use of existing datasets built by other researchers
  - Shared task at WNUT, WMT, SemEval, etc.
  - Datasets used in other papers (e.g. https://aclanthology.org/)

Credit: Stanford CS224n, Chris Manning

Final Project Writeup/Presentation

- Up to 8-**page writeup** due the day before final exam date (no late submission!)
- Use **LaTeX template** from ACL
- Include references; statement of each group members' contribution
- Writeup quality is important to your grade!
- X-minute **oral presentation at the final exam time** (X ∈ [3, 8])



Credit: Stanford CS224n

Have fun with your project!