Course Projects

Final Project

- Groups of 3-4
- **Scope:** on the order of one of the programming assignments
  - But, need to define the problem, come up with right feature representation, write up results in a formal report.

Selecting a Topic

- Part of your thesis? Great! (discuss with advisor)
- Find a problem you are interested in where you think NLP can help.
- Experiment with one of the algorithms we discussed in class.
- OK to build on existing code / datasets.
- **First question:** what is the dataset?

Datasets

- Papers with code (https://paperswithcode.com/area/natural-language-processing)
- NLP Progress (http://nlpprogress.com/)
- Various Semeval Tasks:
  - http://alt.qcri.org/semeval2018/index.php?id=tasks
- Many more…

Requirements

- 4 Page Report
  - Due Friday, December 9
  - **Late reports will not be accepted**
- What is your contribution with respect to previous work?
- Include empirical analysis of your approach
  - Report performance on dev / test set
  - Compare against appropriate baseline methods (example: majority class + LSTM baselines)

Grading Rubric

For the reasonably well-prepared reader, is it clear what was done and why? Is the report well-written and well-structured?
Clarity (1-5):

How original is the approach? Does this project break new ground in topic, methodology, or content? How exciting and innovative is the work that it describes?
Originality / Innovativeness (1-5):

First, is the technical approach sound and well-chosen? Second, can one trust the claims of the report -- are they supported by proper experiments, proofs, or other argumentation?
Soundness / Correctness (1-5):

Does the author make clear where the problems and methods sit with respect to existing literature? Are any experimental results meaningfully compared with the best prior approaches?
Meaningful Comparison (1-5):

Overall (1-5):

General Advice

- **First question:** is the data you need easily available?
- Try to get a simple baseline working as early as possible to determine whether your project idea is feasible.
- Start with a manageable-sized dataset
  - Then scale up…

Group Formation Time (10 minutes)

https://app.gather.town/invite?token=ZhaoC1IpRbSyAQHyG2-X