

COMP7607: Natural Language Processing

Final Project

Yiheng Xu

Email: yihengxu@connect.hku.hk

Credits: Many materials from CSE517@UW, COS484@Princeton, CS388@UTAustin, and CS-4650@Gatech with special thanks!

Final Project

- Goal: reproduce a top-tier NLP paper from the past two years.
- 3-5 student team
- Deliverables:
 - Project proposal (5%): a 1-2 page proposal
 - Presentation (10%): a 5 minute midway presentation with up to 5 slides
 - Final report (35%): a 4-8 page report

Project Timeline

- Project release: Oct 17
- Deadline for team information: Oct 24
- Deadline for research proposal: Oct 31
- Presentation: TBD (around Nov 21)
- Deadline for final report: Nov 28

Find your teammates

- Build your 3-5 student team.
 - Find classmates you already know.
 - Post on Moodle course forum and introduce yourself.
 - If you still can't find a teammate, fill in the form and we can help you.
- **Find your teammates early!**
 - Prepare proposal and report early.

Choose a paper

- Choose a main or finding paper from
 - ACL {2022, 2021}
 - EMNLP {2021}
 - NAACL {2022, 2021}
- ACL Anthology: <https://aclanthology.org/>

Choose a paper

- **Research topic:** Find the problem tackled in the paper interesting.
- **Public dataset:** Check that the data you will need is publicly accessible.
- **Code:** In many cases, the authors may have made code available; this may be a blessing or a curse.
- **Resource:** Estimate the computational requirements for reproducing the paper.

Project proposal (5%)

- Some study types you may perform:
 - **Method Study:** reproduce the baselines in the paper and come up with your own different hyper-parameters, initialization, objective, and components.
 - Change hidden size/lr; BERT-> ELECTRA; LSTM->Transformer
 - **Model/Dataset Analysis study:** reproduce results of various model ablations from the paper to analyse importance; propose different analyses on the model (e.g., test on new datasets, error analysis).
 - Compare different objectives/data scale; SQuAD -> SQuAD 2.0

Presentation(10%)

- 5 minute midway presentation with up to 5 slides.
- Basically your proposal + (initial) results
- Content:
 - Brief background
 - Study plan & settings
 - Results
 - Timeline
 - Individual Contribution

Final Report(35%)

- A 4-8 page project report.
- ACL templates: <https://github.com/acl-org/acl-style-files>
- Collaboration:
 - Coding: GitHub
 - Writing: Overleaf

This Stage

- Read the detailed instructions
- Find a Team
- Brain storming



Project Instructions

https://github.com/ranpox/comp7607-fall2022/blob/main/project/final_project.pdf

Any questions?