CS 590.03: Intro to NLP Project Guidelines

For the final project you either need to: (1) *reproduce* the experiments of an existing NLP paper; or (2) *build* a new NLP system for a task of your choosing. We have included a list of papers below which can help you generate ideas for what to do in the project.

## Timeline

Oct 13: Group information due

Oct 20: Project Proposal due

November 17, 19: In-class presentations of proposed work/preliminary results.

Dec 13: Final report due

## Group Information

The final project will be conducted in groups of 2-4 students. Please form your own groups and let us know what they are via this [google form](https://forms.gle/xqJt4bNhXmRQmrLs6) by Oct 13. Even if you do not have a group please fill out the form and let us know which topics you are interested in and we will match you with someone else. Note that bigger groups will have more cloud credits for running experiments but will also be expected to pursue more ambitious projects.

## Project Proposal

Please submit a 2-page report summarizing what you plan to do for the project. You should consider the following questions when formulating a proposal (from [here](https://people.cs.umass.edu/~miyyer/cs585/lectures/03-loglinlms.pdf)):

* What is the research question? / Which paper are you trying to reproduce?
* What’s been done before? / What are the main findings and why are they interesting?
* What experiments will you do?
* How will you know whether it worked?
  + For reproducibility: results to compare
  + If data exists: held-out accuracy
  + If data doesn’t exist: manual evaluation of system output. Or, annotate new data.

Accordingly, proposals should include the following sections:

1. Introduction + problem statement + motivation (Why is this problem interesting?)
   1. For replication projects include a description of the model/setup you’re going to replicate here.
2. A brief literature review (What has previously been done on this problem?)
3. The datasets you will use and how you will evaluate your project (e.g., perplexity, F1-score, human evaluation).
4. The tools and resources you plan to use (e.g., pytorch, previous implementations)
5. Project milestones: what you plan to do and by when.
6. Team breakdown: broadly, which team members will be responsible for what.

## Grading / Deliverables

There are three deliverables for the project which will contribute to your final grade as follows:

* Initial proposal (2 pages): 20%
* Presentation: 30%
* Final report (8 pages): 50%

Both the project proposal and final report should include a section on who will do / did what in the project. Additional guidance on the expected structure of the final report will be forthcoming.

Each of the items above will be judged based on the clarity of the writing / presentation, correctness of the technical content, and interestingness of the ideas pursued.

## Google Cloud Credits

To help with experimentation every student will receive $50 in Google Cloud credits. Here is the URL you will need to access in order to request a Google Cloud coupon. You will be asked to provide your school email address and name. An email will be sent to you to confirm these details before a coupon is sent to you.

[Student Coupon Retrieval Link](https://urldefense.com/v3/__https://gcp.secure.force.com/GCPEDU?cid=6S55J2ZcL9cLH0qQA2k6RENdoJVMMZ7iy1OSnKNSkQJ4L*2FJ1j4HHsNurdAA*2FjNjL*__;JSUv!!OToaGQ!-HwTE9o6wCYN15UrGmWX_eFCB11_7wz3UQO9q6QKNjq4HMYjbTTpEO1ZoL5VM2uojfM$)

* You will be asked for a name and email address, which needs to match your school domain. A confirmation email will be sent to you with a coupon code.
* You can request a coupon from the URL and redeem it until: 12/23/2021
* Coupon valid through: 8/23/2022
* You can only request ONE code per unique email address. **If you have both a cs.duke.edu address and a duke.edu address, please use only one of them!**

Please contact the instructors if you have any questions or issues.

**Note that $50 will buy you somewhere between 10-100 GPU hours depending on the** [**type of GPU**](https://cloud.google.com/compute/gpus-pricing) **you select. Please plan your projects accordingly!**

## Additional Guidelines for Reproducibility Studies

* You are expected to re-implement *all the code* necessary to generate the results. This includes data preprocessing, model implementation, and evaluation (though it is fine to use existing libraries like hugging face).
* Your final report should include a description of the challenges you faced in re-implementing.
* It is fine if you are unable to reproduce the results, but you should be able to clearly describe why it was difficult and which details were missing from the original paper.
* The best projects in this category will go beyond the experiments in the paper and check things like: how sensitive the results are to hyperparameters, how robust is the model to domain shift or adversarial examples, etc.
* You can read more about reproducibility at the [ML Reproducibility Challenge](https://paperswithcode.com/rc2021) website. We also encourage teams to consider submitting their work to this challenge (deadline Feb 4, 2022).

## Project Ideas

Below is a list of topics and some recent papers on those topics. This is just a starting point -- we encourage you to explore references in these papers as well as other papers which cite these (you can find these on Google Scholar) to discover more work on a topic.

### Recent papers

You can look at recent NLP conferences to get ideas for projects (or for papers to replicate).

* The [ACL Anthology](https://aclanthology.org/) contains the full proceedings (i.e., all published papers) of nearly all NLP conferences.
* Focus your search on recent years, and on those papers appearing at one of ACL, EMNLP, NAACL, TACL, or Findings (which are listed under the “Venue” column in the anthology).
* It’s also fine to replicate or be inspired by an NLP paper from a recent top-tier machine learning conference, such as NeurIPS, ICML, AISTATS, or ICLR.

### Common Tasks

* Classification Tasks
  + Sentiment (and related) detection  
    [SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020)](https://arxiv.org/pdf/2006.07235.pdf)  
    [SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval)](https://aclanthology.org/S19-2010/)
  + Natural Language Inference  
    [What Can We Learn from Collective Human Opinions on Natural Language Inference Data?](https://arxiv.org/abs/2010.03532)  
    [Adversarial NLI: A New Benchmark for Natural Language Understanding](https://aclanthology.org/2020.acl-main.441/)
  + Fact verification / Propaganda  
    [Fool Me Twice: Entailment from Wikipedia Gamification](https://arxiv.org/abs/2104.04725)  
    [SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles](https://aclanthology.org/2020.semeval-1.186.pdf)  
    [FEVEROUS: Fact Extraction and VERification Over Unstructured and Structured information](https://arxiv.org/abs/2106.05707)  
    [Fact or Fiction: Verifying Scientific Claims](https://aclanthology.org/2020.emnlp-main.609/)
* Structured prediction
  + Parsing
  + Relation / Event extraction  
    [DocRED: A Large-Scale Document-Level Relation Extraction Dataset](https://aclanthology.org/P19-1074/)
  + Coreference resolution
* Question Answering
  + Retrieval  
    [Dense Passage Retrieval for Open-Domain Question Answering](https://arxiv.org/abs/2004.04906)  
    [PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them](https://arxiv.org/abs/2102.07033)  
    [Sparse, Dense, and Attentional Representations for Text Retrieval](https://arxiv.org/abs/2005.00181)
  + Reading Comprehension  
    [Natural Questions: A Benchmark for Question Answering Research](https://research.google/pubs/pub47761/)  
    [Text-based NP Enrichment](https://arxiv.org/abs/2109.12085) (\*\*NEW\*\*)
  + KB-QA  
    [Case-based Reasoning for Natural Language Queries over Knowledge Bases](https://arxiv.org/abs/2104.08762)  
    [Measuring Compositional Generalization: A Comprehensive Method on Realistic Data](https://arxiv.org/abs/1912.09713)  
    [The Web as a Knowledge-base for Answering Complex Questions](https://arxiv.org/abs/1803.06643)
  + Closed-book QA  
    [How Much Knowledge Can You PackInto the Parameters of a Language Model?](https://arxiv.org/pdf/2002.08910.pdf)  
    [Language Models as Knowledge Bases?](https://arxiv.org/abs/1909.01066)
* Text generation  
  [The GEM Benchmark: Natural Language Generation, its Evaluation and Metrics](https://arxiv.org/pdf/2102.01672.pdf)
  + Table-to-text  
    [ToTTo: A Controlled Table-To-Text Generation Dataset](https://arxiv.org/abs/2004.14373)
  + Story generation  
    [STORIUM: A Dataset and Evaluation Platform for Machine-in-the-Loop Story Generation](https://arxiv.org/abs/2010.01717)
  + Summarization
  + Machine translation
* Language modeling

### Techniques

* Prompt engineering for language models  
  [Prefix-Tuning: Optimizing Continuous Prompts for Generation](https://arxiv.org/abs/2101.00190)  
  [Cutting Down on Prompts and Parameters: Simple Few-Shot Learning with Language Models](https://arxiv.org/abs/2106.13353)
* Domain adaptation  
  [DEMix Layers: Disentangling Domains for Modular Language Modeling](https://arxiv.org/abs/2108.05036)  
  [Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks](https://aclanthology.org/2020.acl-main.740/)
* Faster pretraining / finetuning
* Efficient transformers for long inputs  
  [Big Bird: Transformers for Longer Sequences](https://arxiv.org/abs/2007.14062)
* Few-shot / zero-shot learning  
  [Neural Data Augmentation via Example Extrapolation](https://arxiv.org/abs/2102.01335)
* Counterfactually augmented data  
  [Learning the Difference that Makes a Difference with Counterfactually-Augmented Data](https://arxiv.org/abs/1909.12434)
* Incorporating linguistic structure into transformers
* Compositional Reasoning  
  [ReaSCAN: Compositional Reasoning in Language Grounding](https://arxiv.org/abs/2109.08994)
* Updating knowledge in language models  
  [Editing Factual Knowledge in Language Models](https://arxiv.org/abs/2104.08164)