CSEP 517-Wi21 Project Instructions

Course Project (15% of grade)

In the course project, we would like you to consider what you are excited about or what you want to try out related to NLP, and explore in this direction. Perhaps, you want to dig deeper into existing methods in a specific domain, extend and improve existing algorithms on certain tasks, or come up with a creative application to another dataset of your interest. In the end, you will explore topics of your interest and will be expected to turn in a final report describing their project. There will be no presentation or poster session.

To help you pick a project topic, we have listed three options for the project. You are welcome to try out different options. We encourage you to submit **an optional 0.5-1 page project proposal by 3/3/21 (Wed) on canvas**, so TAs could provide feedback and suggestions on your project plan. If you are uncertain how to best get started, again just be sure to ask us for help! We have also listed some project ideas in the later section.

Project Options

1) **Replicating an existing paper [suggested length: 4-6 pages]**

Choose an existing paper and replicate its result. Try to first understand the main contribution of the paper and summarize it in the report in your own words. Then, you can either implement the method/algorithm yourself, or use an existing/published code base the authors provide. If you choose the latter option, we expect you to do more explorations, e.g. experiments with perhaps different hyperparameters and ablations, minor modifications to the algorithms, or provide thorough analysis and case study.

2) **Extending existing algorithms [suggested length: 4-6 pages]**

Choose existing algorithms or models in NLP and think of how you can extend them to improve the performance. This can be done on the same task, or perhaps you wish to modify the algorithms to perform well in a new task. Then, code up your new algorithm and compare with the baseline models. This option requires more novelty and might be more difficult than other options, especially if you'd like to see the performance gain. But it is more relevant to what people typically do in NLP research.

3) **Application to new task/dataset [suggested length: 4-6 pages]**

Before you took this class, you might have had some tasks you wanted to solve, or a dataset you wanted to play with. Then, this option is for you, where you can use what you have learned in class in a real world application, and be creative with your task. You should clearly define which problem and dataset you are interested in. We encourage you to think why you are excited about this task or what makes the task interesting. Then, design a program that achieves your goal and report the results.

4) **Literature review [suggested length: 6-8 pages]**

Choose a topic you are interested in and write a literature review about it. First, pick a list of prior works and decide which ones to include and compare (~3-5 papers). Explain why each work was proposed based on previous limitations, and consider the goal of each work and if it has been achieved in the paper. One way is to make subsections for each paper or method. Then, provide the reported results and make a critical comparison among them (perhaps on a standard benchmark dataset). For this option, you do not have to implement the papers, but you are welcome doing so.

Programming Expectation: You will be expected to write code for Options #1-3. In your implementation, you are **allowed to use external libraries**, such as nltk, spacy, sklearn, and etc. We do not expect you to implement models such as CRF from scratch as in homework. As stated in Option #1, you can use published code as part of your implementation, but we encourage you to make some modifications to the code. This includes trying different hyperparameters, slight modification to the algorithm, ablation studies, or using a different dataset of your interest. In the report, please make it clear what you have tried and how the experiments were run. Overall, we expect the programming workload to be roughly comparable to the one in homework.

Groups: The ideal group size would be 2, but group of 3 is also allowed. We will have group sign ups in Ed to help find your group members. But it is fine to work alone if you are comfortable doing so. You only need to submit one report for each group, but please **include the names of your group members** in the report. Also, please indicate your personal contribution to your group project.

Deadline: The final report deadline will be a week after the exams week, which is on **3/21/21 Sun 11:59pm**. **No late days can be used for the final project due to a tight deadline for the official grade submission!** We recommend you to plan and start early on your project.

Report Format

Recommended length for the report is: **4-6** pages for **Options 1-3** and **6-8** pages for **Option 4**. You do not have to fill out the max page limit and will not be penalized if you don't. In fact, it is less preferable to read paper that is long, repetitive content than the one concise and straight to the point!

You can write the report in Word doc or latex, but please **submit the final report in PDF**. The page limit should generally be single spaced and single column, and includes figures and equations. References section will not be counted towards the page limit. If you used one of the programming options, please **include the link to your code (e.g. github)** in the report. For your reference, here is a latex template from a recent NLP conference: emnlp2020-templates.zip.

In general, your report will be in the following format (but not mandatory):

1) Introduction
    a) What is it you are trying to solve or achieve? Why do you think it is important or exciting?
    b) Main summary and contribution of your project.
2) Related Work
    a) Cite relevant works and provide a short summary for each of them.
    b) For literature review, this section might be longer than the other options.
3) Methods & Algorithms
    a) Description of algorithms and model implementation.
    b) Model figures, equations, algorithms.
    c) Explanation on why you chose such algorithms.
    d) You do not need to re-derive equations and repeat content in the lecture.
4) Dataset
    a) Type of data collected and how they were annotated and preprocessed
    b) Dataset statistics (e.g. # of question-answers)
5) Experiments & Results
    a) List of models, hyperparameters, evaluation scheme, etc. used in your experiments.
    b) Training curves showing evaluations on the dev set.
    c) Table summarizing quantitative results for each model.
    d) Write down your findings and provide possible explanations and analysis.
    e) Figure of qualitative examples if available.
6) Conclusion
    a) Wrap up and summarize your findings.
    b) Point to future work and extensions.
7) References (not counted towards the page limit)

Some considerations

- Plan ahead to choose your project topic. Decision making can be curiosity-driven fun activities to wrap up the class with your own independent fun exploration. We have listed some project ideas below. It's always helpful to decide and start early. Feel free to talk to TAs about project ideas and ask for help.
- Feel free to access data of your interest and run experiments.
- Try to avoid direct copying from the paper you choose; summarize the approach in your own words and be specific about what exploration/improvement you do.
- We are not looking for getting the state of the art performance for option 2 & 3, but please provide detailed analysis and explanations on results you see.
- We recommend you to check the computational requirements for reproducing the paper and take into account the resources available to you for the project. Some authors will have had access to infrastructure (like TPUs) that is way out of your budget.
  - In the class, each individual will be given Google/Azure credits for computation which you can use for the project and assignments. **Details TBA**.

Project Ideas

Here is a list of project ideas and datasets suggested by the TAs. You are welcome to use any of the ideas for your project, but you are not required to use them and you are absolutely welcome to come up with different ones!

Project Ideas

- RNNs or Transformers for Classification Tasks: (w/ dataset of your interest)
  - *Sequence Classification*: POS Tagging/Named Entity Recognition
    - Use the same corpus in HW
    - Compare the performance with models used in HW, e.g. HMM, CRF.
  - *Text Classification*: Sentiment analysis, news categorization, question answering, natural language inference.
    Comprehensive review for reference: https://arxiv.org/pdf/2004.03705.pdf
  - See other interesting tasks in the dataset list section.
- Literature Review
  - Survey on Text Summarization, Neural Conversational Systems, Probabilistic Graphical Models, Question Answering Systems
  - Deep Word Embeddings
    Skip-Gram, Word2vec, GloVe, ELMO, BERT (not all papers are required)

- Nonliteral & Metaphorical Language Generation & Detection
    - Analogy Tasks & Dataset: https://aclweb.org/aclwiki/Analogy_(State_of_the_art)
    - Neural Metaphor Detection in Context: https://arxiv.org/abs/1808.09653
    - Sarcasm Generation with Commonsense: https://arxiv.org/pdf/2004.13248.pdf
    - Humor Recognition: https://www.aclweb.org/anthology/H05-1067.pdf
        - BERT implementation: https://arxiv.org/pdf/2004.12765.pdf
- Natural Language Generation (NLG)
    - Train a language model using RNNs on a book!
        - Compare perplexity of ngrams, character vs word level RNNs
        - Blank out some parts of the sentence and use decoding methods learned in NLG lecture to fill them in. Report accuracy & language metric in HW4. Decoding methods: greedy max, top-k sampling, and nucleus sampling
        - Free e-book dataset: https://www.gutenberg.org/
    - Generate natural language description from a tuple of facts (for example, inform(movie="Titanic", start_time="8:15PM", date="Monday, Feb 2"))
        - Dataset: MultiWoz (https://github.com/andy194673/nlg-sclstm-multiwoz)
        - An example, Semantically Conditioned LSTM-based NLG (SC-LSTM, EMNLP 2015 Best paper), https://arxiv.org/pdf/1508.01745.pdf , Pytorch Github: https://github.com/andy194673/nlg-sclstm-multiwoz
    - You can implement any model, and compare with the baseline models. (No need to be state-of-the-art)
- Table-to-Text generation
    - Generate the natural language for Tables.
    - Dataset: Long text, https://github.com/harvardnlp/boxscore-data, paper, https://arxiv.org/abs/1707.08052
    - Dataset: Wiki2Bio: https://github.com/tyliupku/wiki2bio
    - You can choose any of these datasets, and implement any model, and compare with the baseline models. (No need to be state-of-the-art)
    - Reading list: https://github.com/DrLiLiang/Data-to-Text-Generation
    - Table-based dialogue generation: Multi-domain Neural Network Language Generation for Spoken Dialogue Systems
- Story ending generation
    - ROCstories Dataset: https://cs.rochester.edu/nlp/rocstories/
    - Plan-And-Write: Towards Better Automatic Storytelling https://arxiv.org/abs/1811.05701

    - PlotMachines: Outline-Conditioned Generation with Dynamic Plot State Tracking https://arxiv.org/abs/2004.14967
- Cooking recipe generation
    - Reference-Aware Language Models https://www.aclweb.org/anthology/D17-1197/
- AI ethics and social bias
    - Quantifying Social Biases in Contextual Word Representations http://www.cs.cmu.edu/~ytsvetko/papers/bias_in_bert.pdf
    - Black is to Criminal as Caucasian is to Police: Detecting and Removing Multiclass Bias in Word Embeddings https://arxiv.org/abs/1904.04047
    - Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them https://arxiv.org/abs/1903.03862
    - Gender Bias in Coreference Resolution https://arxiv.org/pdf/1804.09301.pdf
    - Evaluating Gender Bias in Machine Translation https://www.aclweb.org/anthology/P19-1164/
    - Social Bias Frames: Reasoning about Social and Power Implications of Language https://arxiv.org/abs/1911.03891
    - Social Chemistry 101: Learning to Reason about Social and Moral Norms https://arxiv.org/abs/2011.00620
    - RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models https://arxiv.org/abs/2009.11462
- Winograd Schema Challenge
    - 2016 challenge link: http://commonsensereasoning.org/winograd.html
    - Paper reviewing the challenge: https://arxiv.org/pdf/2004.13831.pdf
    - More challenging version: WinoGrande
- Real-World Semi-Supervised Learning of POS-Taggers for Low-Resource Languages
    - https://www.aclweb.org/anthology/P13-1057.pdf
- Crawl text data using API to design a new dataset and task
    - Data crawling is NOT REQUIRED for this project and might take a lot of time to do than using an existing dataset.
    - This option is here if you want to work with your own dataset in the future.
    - Some examples (note we have not verified the codebase below):
        - Collect tweets and classify if the tweet is from a celebrity or not. https://github.com/jonbakerfish/TweetScraper
        - Collect data from reddit and train model to answer questions
            - AITA subreddit: https://github.com/iterative/aita_dataset
            - RedditAdvice: https://github.com/rowanz/turingadvice

        - Collect asr from youtube to classify the category of the video https://github.com/algolia/youtube-captions-scraper

Interesting Dataset

- Kaggle includes a great list of Machine Learning dataset and competitions: https://www.kaggle.com/tags/nlp
    - Toxic Comment Classification
    - Quora Question Pairs
    - Google Sentiment Analysis
    - News Summary
    - Sarcasm Detection
- WikiQA corpus
- Yelp Review Dataset
- ROCStory Dataset
- WordNet