EECS 595: Natural Language Processing
Fall 2023
Final Project Description

October 21, 2023

1 Introduction

This project provides you an opportunity to explore a topic that you are interested in. You can choose any topics related to natural language processing or the application of NLP techniques to your own research areas. You can choose to work solo by yourself, or by a team of ***up to three members***. You can choose to use any publicly available software or resources as long as a proper acknowledgment is provided.

2 Topics

2.1 Option 1: Choose Your Topic

Feel free to pick any topic of your choice related to NLP. You are encouraged to discuss potential topics with the instruction team. When considering different topics, you are encouraged to think outside the box. For example, you can use the existing data to define new interesting problems which may be different from the original tasks set out by these data.

ACL Anthology As in most areas in Computer Science, for the NLP and Computational Linguistic community, the latest research advances are often first reported at conferences (although some of them may have appeared ahead of time in pre-print arXiv). The conferences particularly relevant to this class include (but not limited to): the Annual Meeting of the Association for Computational Linguistics (ACL), the conference on Empirical Methods in Natural Language Processing (EMNLP), the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), the European Chapter of ACL (EACL), and the Asia-Pacific Chapter of ACL (AACL). The ACL Anthology makes available a large collection of research papers on various topics in NLP. You are encouraged to explore these resources to identify the topic you would like to work on.

SemEval For example, a collection of tasks and datasets have been developed to examine machine's ability in semantic processing. This ranges from lexical semantics such as word sense disambiguation to sentence-level semantics such as semantic parsing. You can start with the SemEval competition each year (up to 2018). A number of tasks were defined for each competition. Data for these tasks are often provided, so are the relevant papers. You will be able to compare your results with the published results.

Other Potential Topics You can also find some example final projects from Stanford NLP class and EECS595.

2.2 Option 2: Default Topics

If you prefer an assigned topic, the following is a list of default topics for you to consider. You can choose any one of these default topics.

2.2.1 Default Topic 1: Conversation Entailment

Example 1:

Conversation Segment:

B: My mother also was very very independent. She had her own, still had her own little house and still driving her own car,
A: Yeah.
B: at age eighty-three.

Hypothesis:

(1) B's mother is eighty-three.
(2) B is eighty-three.

Example 2:

Conversation Segment:

B: Have you seen *Sleeping with the Enemy*?
A: No. I've heard that's really great, though.
B: You have to go see that one.

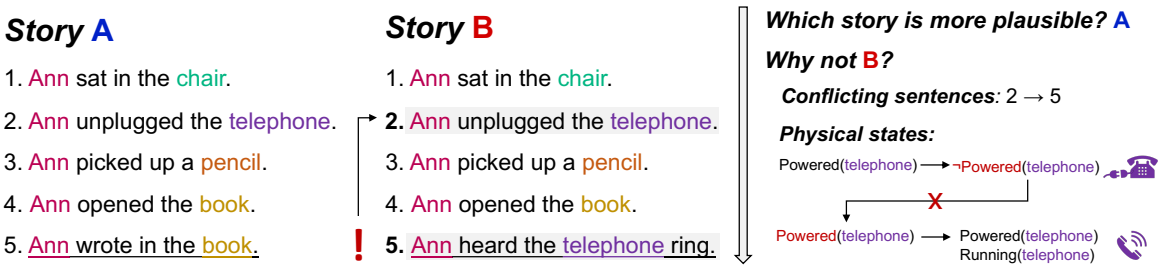
Hypothesis:

B suggests A to watch *Sleeping with the Enemy*.

Overview Conversation Entailment is an NLP dataset on textual entailment for conversations. In this task, a multi-turn natural language dialogue is provided as a premise. Given a hypothesis sentence based on the dialog, we must predict whether or not the hypothesis can be inferred from the dialog as a binary text classification. When this paper was released, the baseline system performance was low (less than 60% entailment accuracy). A recent paper from the SLED Lab revisits this task and finds that while state-of-the-art transformer-based language models can greatly improve the accuracy, their understanding is incoherent and may be based on spurious intermediate evidence. For this project, you may consider approaches to improve the accuracy and/or coherence on this task.

Resources You can find the original paper here. The dataset is available on GitHub and HuggingFace.

2.2.2 Default Topic 2: Tiered Reasoning for Intuitive Physics



Overview Tiered Reasoning for Intuitive Physics (TRIP) is a new NLP dataset from the SLED Lab at UMich. It poses a multi-layered physical commonsense reasoning task where AI systems

must determine which of two stories (each describing a series of physical actions applied to household objects) is plausible. Further, they must justify their decision by identifying which sentences in the implausible story are conflicting with each other, and the specific physical state changes that cause the conflict. When we evaluate systems on this tiered prediction, baseline results are very low (only up to 10% of stories are classified correctly and verified with coherent supporting evidence). For this project, you may consider approaches to improve the baseline performance on TRIP.

Resources You can find the published paper here. The code and TRIP dataset can be found on GitHub and HuggingFace.

2.2.3 Default Topic 3: Physical Causality Reasoning

	Top Action Predictions	Top Effect Descriptions		Top Action Predictions	Top Effect Descriptions
	**bake potato** peel potato boil potato fry potato	potato crispy potato is crushed eggs get beaten potato browned		wrap book **tear book** fold paper shave hair	book is ripped paper become creased book into smaller pieces meat is being prepped
	peel carrot cut wood **chop carrot** grate carrot	carrot into little sections tree into pieces carrot into tiny pieces wood is being chopped		stain paper close drawer **squeeze bottle** crack bottle	bottle is pressed together meat is exposed paper around itself drawer is pushed back
	chop onion cut onion slice onion **background**	onion is being cut onion in banana is made banana is removed		chop onion cook onion grate potato **background**	onion is heated onion into small pieces onion into multiple pieces onion is chopped

Overview The Physical Action-Effect Prediction is a vision-language task which is designed for physical causality reasoning by connecting actions (as causes) to their effects. The dataset contains action-effect information for 140 verb-noun pairs. It has two parts: effects described by natural language, and effects depicted in images. Specifically, given an image which depicts a state of the world, the task is to predict what concrete actions could cause the state of the world. Unlike traditional action recognition where the underlying actions are visually captured, the images in this task only capture the effects of actions. The model described in the paper is very simple and the performance is yet to be improved. For this project, you may consider modern approaches to improve the baseline performance on the Physical Action-Effect Prediction, *e.g.*, involving pre-trained vision-language models.

Resources You can find the published paper here. The dataset is available on GitHub and HuggingFace.

2.3 Tools

HuggingFace HuggingFace maintains a wide range of NLP datasets on its Hub. You may explore the Datasets available there.

Papers With Code Papers With Code maintains a wide range of NLP benchmarks and their state-of-the-art solutions. You may find the NLP page helpful.

StanfordNLP The Stanford NLP group has made many resources/tools available to the community, *e.g.* Stanza and CoreNLP you tried in HW3.

3 Scope

The scope of the project should be proportional to the size of the team (i.e., solo vs. three members). For example, if your team has three members, one option is to look at a shared problem from three different angles or approaches. It will allow you to compare and contrast as a whole to reach a better understanding of the problem/solution space. You need to show that the overall scope of the project justifies the size of the team and there is an equal division of the effort. ***If it's a team project, you need to include a brief statement in your proposal and in the submission page of your final report about the work division and responsibility of each team member.*** Note that it is possible that team members receive different points for the final project depending on their contributions.

4 Grading of the Project

The final project is counted for 40 points out of 100. It consists of the following components. It is your responsibility to make sure each component is delivered on time to receive a full point. If it's a team project, you only need to make one submission from one of the team members (no duplicate submissions please).

4.1 Project Proposal (10pts)

You need to submit a two page project proposal for approval. In your proposal, you should specify the following:

1. The problem you are trying to address.
2. The proposed approaches.
3. The data set that will be used.
4. Any previous work on this topic? Provide references.
5. The plan of implementing your approach and milestones for the rest of the semester.
6. The composition of the team and work division between team members.

The proposal is due on **November 6, 11:59pm**.

4.2 Project Presentation (10pts)

You will present your project to the class during one the following class sessions **November 27**, **November 29**, or **December 4**. The instructions for presentations will be made available on November 10. After the presentation, you will need to submit the final slides which address questions raised during the presentation by **December 8, 11:59pm**

4.3 Project Final Report (20 pts)

The final report of the project is due on **December 13, 11:59pm**.

Your report should include the following section:

1. Introduction: The problem statement
2. Related Work
3. Detailed description of your approaches (note if you have multiple people on your team, multiple approaches are expected)
4. Evaluation: comparison between different approaches, different configurations of your approaches. If baseline results are reported in previously published work, you should include that comparison too.
5. Discussion of your results
6. Conclusion
7. Division of Work (only needed if it's a team project)
8. An URL link to your code base
9. References

Depending on the size of the team, your report needs to meet the following page constraint (using ACL format, excluding references).

- A solo team: at least 5 content pages
- A two-member team: at least 7 content pages
- A three-member team: at least 9 content pages

The goal of the final project is for you to gain some hands-on experience in solving NLP problems. If the problem you choose to work on has been studied by previous work, do not worry if your results do not outperform the published results. You should elaborate on what you have learned from this experience (e.g., through error analysis).

5 Final Report Template

For both your proposal and final report, you should use the ACL paper format. The latest templates are available in this repo, and also on Overleaf (need to make your own copy to edit it).

6 A Summary of Timeline

Please keep the following timeline in mind:

- Project proposal: due **November 6, 11:59pm**
- Project presentation: in class **November 27**, **November 29**, or **December 4**. (The schedule will be made available on Nov. 10).
- Presentation slides: due **December 8 11:59pm**
- Project report: due **December 13 11:59pm**

7 Computing Resources

Google Colab could be an option for you to run large jobs. You may look into that at https://colab.research.google.com/notebooks/welcome.ipynb. You can also use our course Great Lakes account for your project.