CSE 587: Deep Learning for Natural Language Processing

How to Do a Research Project in NLP?

Rui Zhang

Spring 2023

PennState

Project Format

Project Topics. This project aims to conduct original and independent research over **any NLP-related topics**.

- Must deal with natural language data
- Must include some degree of model implementations and experiments

Group Policy. You can work on the course project in a group of **1 - 3** people. You are allowed to combine this project with your research projects or projects from other courses.

Project Deliverables

- **Project Proposal**. Write a 3-page proposal that outlines your plan including what problem or task you want to address, what dataset(s) you want to work on, what metrics you need to employ, what baselines you would like to compare with. You should also cite a few relevant prior papers.
- **Final Report**. Your final report should use our Latex template with at least 8-page plus references. Your report should begin with an abstract and introduction to clearly state the problem you want to solve and contributions you have made. It should also have a section on related work, a section on your methodology, a section on your experimental settings and results, and a section on conclusions.
- **Code and Data**. Please submit your data and runnable code with a detailed instruction.

Project Timeline

Group Formation: January 27

Group Registration: February 3 (Canvas: People -> Project)

Project Proposal: February 17 (You must submit after registration)

Final Report/Code and Data: May 5 (during the final exam week)

Possible Project Approaches

This project aims to conduct original and independent research over NLP-related topics. Students can choose from several possible approaches:

1. Invent a new and important task in NLP and create a dataset for it, e.g., MACSum: Controllable Summarization with Mixed Attributes
2. Create a new dataset for an existing NLP tasks, e.g., Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task
3. Pick an existing NLP task and dataset and try to get good results, e.g., CONTaiNER: Few-Shot Named Entity Recognition via Contrastive Learning and check this awesome website Papers With Code: The latest in Machine Learning
4. Try to pick a NLP problem and dive deep into it with comprehensive analysis, e.g., Selective Annotation Makes Language Models Better Few-Shot Learners

Basic Requirements

Project Topics. This project aims to conduct original and independent research over **any NLP-related topics**.

Must deal with natural language data. You can

- use existing datasets
- develop new one for good reason.

Must include some degree of deep learning model implementations and experiments. You can

- build novel NLP system pipeline
- improve existing models for improving performance
- apply existing models for some novel task you are interested.

Grading Criteria

- No matter which approach you take, we will hold your project to a high standard. Excellent projects will result in published papers in top-tier NLP/AI/ML conferences or journals (e.g., ACL, NAACL, EMNLP, AAAI, ICLR, NeurIPS, ICML, TACL, ...).
- A typical case is that you show your method provide **good results** on the task you choose.
- It is not necessary to beat state-of-the-art, but it should **work reasonably well**.
- No matter what performance you get, you should provide **analysis and explanation**.
- Your project is not (only) graded based on the "accuracy" number, but we consider **originality**, **novelty**, **complexity of your problem**, **soundness of your approach**, **thoroughness of your experiments**, **performance**, and **writing quality**.

Project Ideas

- Start with existing published research at ACL, NAACL, EMNLP, AAAI, ICLR, NeurIPS, ICML, TACL, ...
- Browse Papers with Code, NLP-progress, Hugging Face Datasets
- Welcome to check out my research and publications: https://ryanzhumich.github.io/ and https://scholar.google.com/citations?user=8GCLiUwAAAAJ&hl=en
- Project Reports from Stanford CS224n: Natural Language Processing with Deep Learning: http://web.stanford.edu/class/cs224n/project.html

Research Trends in 2023

LLMs

- Emergent Abilities of Large Language Models

Evaluating and improving NLP models beyond accuracy

- Robustness to domain shift
- Fairness, Bias, Trustworthiness
- Explainability, Interpretability

Efficient NLP models

- Small models with similar performance: model pruning, model quantization
- Small training data: few-shot learning, zero-shot learning

Multilingual NLP

- NLP beyond English
- Low-resource languages

NLP for Social Good

- Education, Healthcare, Security, Privacy

How to do a RESEARCH project in NLP?

What is the **Task**?

- Can you **name** the task? Text Classification? Question Answering? Summarization?
- What is the **input** and **output**?
- Can you give some **examples**?
- What is your **motivation**? Why is it an **important** task?

How to do a RESEARCH project in NLP?

What is the **Main Goals(s)** / **Research Question(s)**?

- What is **challenging** about the task?
- What is the **research question**? An excellent project starts with a good research question.
- What is your **hypothesis**?
- What does success look like for your project? What new knowledge will you create after you finish your projects?
- Be ambitious, and we encourage ambitious and creative ideas.
- Your project grade is not judged on what accuracy you get. Your project does not have to work or solve the problem. An excellent project is a step towards a correct direction. So, formulate your research question.

How to do a RESEARCH project in NLP?

What people have done regarding your research questions?

- Do literature reviews. You should read at least **20 papers** on your research question.
- You don't need to read every line of them, but you need to know their motivations and what they have done. **Just read the title and abstract.**
- Sometimes, an interesting idea may arise by **combining ideas of 2 papers**, or **borrowing ideas from another field**.
- Use Google Scholar, ACL Anthology, ArXiv, Papers with Code, NLP-progress

How to do a RESEARCH project in NLP?

What **datasets** are you going to use?

- An existing dataset? Where can you download it?
- A new one? If so, why? Is it because it is a new task and there is no dataset for this, or existing datasets have some flaws and do not suit your research question? How do you create your own dataset?
- Why do you think the dataset you choose is appropriate for your research goals of this project?

How to do a RESEARCH project in NLP?

What **models** are you going to use?

- A set of existing baseline models?
- To adapt a state-of-the-art model?
- To create your own model?

How to do a RESEARCH project in NLP?

What **evaluation metrics** are you going to use?

- Automatic evaluation metrics?
- Human evaluation?
- Combination of the two?

How to do a RESEARCH project in NLP?

What **experiments** you need to answer your research question?

- How do you design experiments to answer your research question and your hypothesis?
- How to interpret your **overall results**? How does your approach compare with other baselines and previous best model?
- Can you do some **ablation study** by removing some components of your model and see how it works?
- Can you perform interesting **analysis** to show how model works? It can be **quantitative** or **qualitative**. e.g., Give some examples of when it succeeds and when it fails. Error analysis. Break down the performance by some criteria. Visualization of Attention. Visualization of Embeddings.

An Example Project: Fake News Detection

1. Define Task: Fake News Detection
   - Input: A news article
   - Output: fake or real

2. Datasets
   - Use existing datasets, e.g., LAIR, FakeNewsNet
   - Collect your own data: harder but a new opportunity.

3. Dataset Hygiene
   - Define training, development, test split

4. Metrics
   - Automatic Evaluation: Accuracy, F1
   - Human Evaluation: how well do people can solve this task?

An Example Project: Fake News Detection

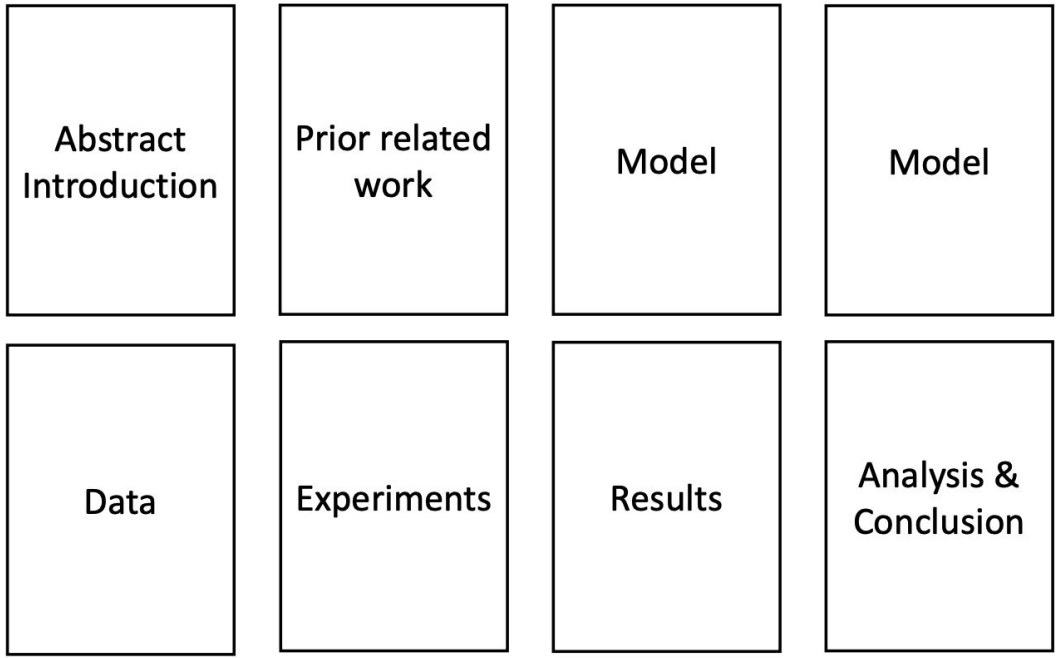
5. Baselines
- Implement a simple baseline first: e.g., logistic regression on unigrams or bigrams, or deep average networks, or LSTMs
- Literature review on other papers on this task. What approaches do they use?

6. Implement your own models
- Try different models and approaches: RNNs, CNNs, attention-based models, Transformers, pretrained language models, BERT, RoBERTa
- Improve performance
- You goal is to do better than state-of-the-art

7. Collect Results and Perform Analysis

Project Report

8 pages. A typical publication in ACL will look like this.



Data Hygiene

Training Data: Use this and only this to train your neural networks.

Development Data: Use this to decide when to stop training your neural networks, and do hyperparameter tuning.

Test Data: Use test data after picking your model. You should use test data only once.

Computing Resources

- Physical Lab: 135 Lab, Westgate Building. Nvidia A4500 (20GB) GPUs
- Remote Servers: A6000 (48 GB) GPUs. Provided by Department. Instruction TBD.

Use Free Resources

- Google Colab
- OpenAI API

How to train and debug neural networks

Most common neural net mistakes by Andrej Karpathy
A Recipe for Training Neural Networks by Andrej Karpathy
Deep Learning for NLP Best Practices by Sebastian Ruder
Practical Advice for Building Deep Neural Networks by Matt H and Daniel R

Resources

Start with existing code. Many papers publish with a github link.

Open-source NLP libraries

- Stanford CoreNLP: https://stanfordnlp.github.io/CoreNLP/
- spaCy: https://spacy.io/
- PyTorch: https://pytorch.org/
- TensorFlow: https://www.tensorflow.org/

Project Management

- Use **Anaconda** to setup your python environment.
- Use an IDE to write code. I personally use Visual Studio. or Google Colab with GPUs!
- Use a **Github** to maintain your code.
- Use **wandb** to monitor training and record experiments. Highly recommended.
- Use **Overleaf** with **provided latex templates** to write your proposal and report.

Overleaf Template

Proposal

- https://www.overleaf.com/read/nqhjfcfjvxpq

Progress Report / Final Report

- https://www.overleaf.com/read/grpvkxgcdqdt

Projects in Spring 2022

14 teams

Final Reports are grades in three levels:

- 15 out of 15
- 10 out of 15
- 5 out of 15

Four sample reports are provided in Canvas.

Note that the total score is different this year.

Sample Project

Eavesdropping Continuous Speech on Smartphones via Motion Sensors

15 out of 15

+ interesting and important problem on speech eavesdropping on smartphones; the research challenges span speech recognition, IoT, machine learning, and security.
+ clear descriptions of the overall architecture
+ Synthetic Training Data Generation is most interesting and novel part

- it is not clear how the ASR model is different from Fayek 2016.
- Is it possible to try Transformers-based ASR approaches?
- What are the baselines for comparisons? Would you plan any user study?
- How would different smartphone hardware affect your solution and results?

Sample Project

A Human-labeled dataset and benchmark for real time social media data and news summarization

10 out of 15

+ summarization with real-time social media data is an interesting new direction
+ signification potential social impact on public policy
+ extensive discussion on related work including news summarization and tweet summarization

- Given there is no dataset for this task, the project proposed to start with NewsRoom and MultiNews, however, it is not clear how to build a benchmark including tweet summarization.
- This is a new direction and a new dataset would be really valuable. Or more research in existing related datasets would be helpful.
- In the proposed framework, the most challenging part is "Combination". However, it is not clear what methods exactly are proposed to do this.
- What is the best way to evaluate the summarization quality? Would you plan any human evaluation to test your "hypothesis that there is a value in combining social media data with news summarization tasks to obtain more information regarding real-time feedback and public opinions"?

Sample Project

Improving negative sampling for contrastive learning

10 out of 15

+ interesting problem! generating/selecting hard negative examples for contrastive learning is important for effective representation, yet this remained underexplored
+ the proposed approach sounds interesting and viable
+ clear experimental plan including the datasets and baselines to compare

- is it possible to use your technique for unsupervised contrastive learning?
- while GLUE is certainly a reasonable choice, yet the dataset performance has been saturated.
- need a more clear comparisons with previous work on data augmentation and negative sampling for contrastive learning to show why your proposed method is better
- suggest organizing your related work under several topics: contrastive learning; data augmentation;
- more description of your methods are needed
- it is not clear what is random cropping and how it is used exactly
- it is not clear where synonyms and antonyms come from. Are you using WordNet?
- it is not clear why Quick-Thought Vectors is chosen, especially considering that there have been other sentence embedding approaches such as SimCSE and Sentence-BERT etc.

Sample Project

Sentiment Analysis for IMDB Reviews

5 out of 15

+ clear problem
+ include different approaches including transformers/bert/cnn
+ description of network hyperparameters progress of experiments have been documented

- this topic is not very research-oriented, because sentiment analysis has already been studied extensively.
- the goal of the project is not clear: is it to provide a better sentiment analysis model? or is it compare existing ones and provide new analysis and insights?
- more citations will be better

More Advice

Start early.

Choose a project that is **research-oriented**.

Position your project in the context of other related work.

Write your report like an research paper, not just an experimental log.

Expectation management.

- If you want to secure a high score, a good strategy is to combine this project with your current research agenda.
- If you want to explore high-risk ideas using this project, you can also do so.
- The ultimate goal of this project is not to score high, it is the chance for you to practice your research capabilities and advance your research agenda.