
(http://nlp.stanford.edu/)

CS224n: Natural Language Processing with Deep Learning

(index.html)

Course Project

Overview

The Course Project is worth a significant portion of your grade. It offers you the chance to apply your newly acquired skills towards an in-depth application.

Your project can be on any topic of your choice related to deep learning for NLP. To be precise: That means that the project should make substantive use of deep learning and substative use of human language data. You also have the option of doing a default final project (aka Assignment 4) that we will provide scaffolding code for. This will be similar to the other assignments but longer and more open-ended.

Here (https://docs.google.com/document/d/1i-3HwziSVPem-hgJ1BW7MueLIxeDjUyO9fb5_lr3WDQ/edit?usp=sharing) is a list of project ideas from Stanford researchers!

To inspire more ideas, you might look at recent deep learning publications from top-tier NLP conferences and labs:

- ACL (https://www.aclweb.org/): Association for Computational Linguistics
- EMNLP (http://emnlp2014.org): Empirical Methods in Natural Language Processing
- NIPS (http://nips.cc/): Neural Information Processing Systems
- ICML (http://icml.cc/): International Conference on Machine Learning
- ICLR (http://iclr.cc/): International Conference on Representations
- arXiv (https://arxiv.org/list/cs.CL/recent): e-prints archive
- Stanford NLP Group (http://nlp.stanford.edu/publications.shtml): New and relavent papers from local faculty
- Kaggle challenges (http://www.kaggle.com/): An online machine learning competition website
- Stanford's CoreNLP (https://github.com/stanfordnlp/CoreNLP): A Java suite of Core NLP tools
- NLP+Twitter (https://github.com/aritter/twitter_nlp): Twitter NLP tools
- Past cs224d Projects (http://cs224d.stanford.edu/reports_2016.html)
- Past cs224n Projects (http://nlp.stanford.edu/courses/cs224n/)

We have included below some pretty cool papers for further inspiration.

- **Natural Language Processing (Almost) From Scratch (http://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf)**
- **LSTM: A Search Space Odyssey (http://arxiv.org/abs/1503.04069)**
- **Named Entity Recognition (http://www.aclweb.org/anthology/W09-1119)**
- **Part of Speach Tagging (http://nlp.stanford.edu/pubs/CICLing2011-manning-tagging.pdf)**
- **Coreference Resolution (http://www.eecs.berkeley.edu/~gdurrett/papers/durrett-klein-emnlp2013.pdf)**
- **CBOW and Skip-Gram (http://arxiv.org/pdf/1301.3781.pdf)**
- **Toward AI Complete Question Answering: A Set of Prerequisite Toy Tasks (http://arxiv.org/abs/1502.05698)**
- **Memory Networks (http://arxiv.org/pdf/1410.3916v8.pdf)**

Important Dates

Course project proposal (only for students not doing the default final project): due February 9.
The poster session will be held 12:15-3:15pm on March 21.
Final course project: due **March 17** (11:59pm).

Finding a Mentor

Students proposing their own project are **required** to have a mentor who will provide high-level guidance for the project. You must find a mentor before submitting your project proposal. To find a mentor

1. Contact them in office hours or through email. Briefly go over your background, interests, and ideas you have for the project.
2. If they agree to mentor you, **email cs224n.win1617.mentors@gmail.com with your mentor cc'd**. In the email state who your mentor is as well as the names and SUNetIDs of the people in your team.

Here are the mentors affiliated with cs224n:

- Chris Manning (http://nlp.stanford.edu/manning/) (manning@cs.stanford.edu)

- Richard Socher (http://socher.org) (richard@socher.org)
- Danqi Chen (http://cs.stanford.edu/~danqi/) (danqi@cs.stanford.edu)
- Arun Chaganty (http://arun.chagantys.org) (chaganty@cs.stanford.edu)
- Kevin Clark (http://cs.stanford.edu/~kevclark/) (kevclark@cs.stanford.edu)
- Ignacio Cases (http://stanford.edu/~cases/) (cases@stanford.edu)

However, you are free to ask anyone else (such as other Stanford faculty and grad students) to mentor you as long as they have substantial experience with deep learning.

Mentors will take up to 25 teams each on a first-come-first-serve basis, so it is recommended you contact them early!

Grading Policy

The final project is worth 30% for your grade (with your poster presentation comprising additional 2%). You are allowed to work in groups of up to 3 people and may use up to 3 late days. See the grading page (grading.html) for more details.

Project Proposal

The project proposal should be a few short paragraphs (200-400 words overall). If you are doing the default final project, you do not need to submit a proposal. **If you do not submit a proposal on time, you have to do the default final project**. Your proposal should contain the following headings:

- **Mentor:** Who is your mentor for the project? You should have already emailed cs224n.win1617.mentors@gmail.com with this information.
- **Problem Description:** What is the problem that you will be investigating? Why is it interesting?
- **Data:** What data will you use? If you are collecting new datasets, how do you plan to collect them?
- **Methodology/Algorithm:** What method or algorithm are you proposing? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations?
- **Related Work:** What reading will you examine to provide context and background?
- **Evaluation Plan:** How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g. plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)?

Submission: Please upload one proposal per team on Gradescope.

Final Submission

Your final write-up should be between **6 - 8** pages using the provided template (project_template/template.zip). After the class, we will post all the final reports online so that you can read about each others' work. If you do not want your writeup to be posted online, then please let us know when you submit your writeup.
You should include a brief statement on the contributions of different members of the team. Team members will normally get the same grade, but we reserve the right to differentiate in egregious cases.

Submit your final submission as instructed below:

1. A PDF file of your final report submitted through Gradescope.
2. A zip file with supplementary materials through our submission script. You are required to include all the code for your project in the supplementary materials.

Report. The following is a suggested structure for the report:

- Title, Author(s)
- **Abstract**: It should not be more than 300 words
- **Introduction**: this section introduces your problem, and the overall plan for approaching your problem
- **Background/Related Work**: This section discusses relevant literature for your project
- **Approach**: This section details the framework of your project. Be specific, which means you might want to include equations, figures, plots, etc
- **Experiments**: This section begins with what kind of experiments you're doing, what kind of dataset(s) you're using, and what is the way you measure or evaluate your results. It then shows in details the results of your experiments. By details, we mean both quantitative evaluations (show numbers, figures, tables, etc) as well as qualitative results (show images, example results, etc).
- **Conclusion**: What have you learned? Suggest future ideas.
- **References**: This is absolutely necessary.

Supplementary Material is not counted toward your 6-8 page limit.
Examples of things to put in your supplementary material:

- Source code (required).
- Cool videos, interactive visualizations, demos, etc. (optional)

Examples of things to not put in your supplementary material:

- All of a submodules (Theano, Caffe, CoreNLP) source code.
- Any code that is larger than 1MB.
- Model checkpoints.
- A computer virus.

Poster Session

We will hold a poster session in which you will present the results of your projects im the form of a poster. The poster session will happen on Match 21st, 12:15-3:15pm (location: Lathrop second floor). Poster boards and easels will be provided.

Example Project Reports

Your project reports should structure like a NLP conference paper (NIPS, ICML, EMNLP, ACL, etc.). You can find publications from Stanford NLP Group from here (http://nlp.stanford.edu/publications.shtml). In addition, you may also take a look at some previous projects from other Stanford CS classes, such as CS221 (http://web.stanford.edu/class/cs221/sample-projects/), CS229 (http://cs229.stanford.edu/projects2013.html), CS224W (http://web.stanford.edu/class/cs224w/projects.html) and CS231n (http://web.stanford.edu/class/cs224w/projects.html) as well as previous cs224n (http://nlp.stanford.edu/courses/cs224n/) and cs224d (http://cs224d.stanford.edu/reports_2016.html) projects.

Collaboration Policy

You can work in teams of up to **3** people. Larger teams are expected to do larger projects. See the grading page (grading.html) for more details.

Honor Code

You may use any existing code, libraries, etc. and consult and any papers, books, online references, etc. for your project. However, you must cite your sources in your writeup and clearly indicate which parts of the project are your contribution and which parts were implemented by others. Under no circumstances may you look at another cs224n group's code or incorporate their code into your project.

If you are doing a similar project for another class, you must make this clear and write down the exact portion of the project that is being counted for CS224n.