

CS 291A: Deep Learning for NLP

Project advice & Brainstorming

William Wang
UCSB Computer Science
william@cs.ucsb.edu

Including slides from Jason Eisner.

Announcements

- Course website and syllabus:
 - <http://william.cs.ucsb.edu/courses>
 - Registered students should sign up on the presentation schedule (Ke Ni, ke00@ucsb.edu), or we will just assign you to a slot.
 - Students who joined us from the waitlist, let Ke Ni know.
- We will upload lecture slides to resource page on Piazza.
 - <http://www.piazza.com/ucsb/winter2018/cs291a>
- Feel free to use Piazza to discuss project ideas or seek feedback.
 - There are several 291A students who have already published at top conferences.

Differences between two DL classes at UCSB

- CS291K by Dr. Xifeng Yan: Advanced DL applications with focuses on IE, QA, chatbots. Prerequisite: 291A. This course has a qualification exam.
- CS291A (this course): fundamental neural network concepts and models, with wide-range of topics in DL4NLP and focuses on the research component.

Other related courses

- CS 190I/291A (by me): Introduction to NLP.
 - Next offering: Fall 2018.
- CS 165B (by me), Machine Learning
 - Next offer: Spring 2018 (next quarter).
- CS 165A (Yan/Singh/Turk): Artificial Intelligence
- Deep Learning Course at Math
 - Please contact Paul Atzberger.
- Computational Linguistic Courses
 - Please contact Fermín Moscoso del Prado Martín.

Can I get into this course?

- There will be ~10-15 students dropping the class...
- You are welcome to sit in the class.
- We will offer the same course again in next year, hopefully with more resources.

Re: Google Cloud

- Is this a requirement?
 - No. It's a free GPU computing resource that 291A provided, courtesy of Google Cloud.
- You are welcome to use your own computing infrastructure for this course.

Today's Agenda

- Projects:
 - How to propose a novel research project.
 - What to avoid.
 - Project ideas and open-research problems for DL4NLP.
 - Selected engineering advice.
- Brainstorming:
 - Sketching out some ideas.
 - Verify whether it is practical.
 - Structured Group discussions.
 - Instructor feedback.

William's guide to

WHAT'S A NOVEL DL4NLP RESEARCH PROJECT?

Novelty: Problem

- Is this problem a significant new problem? Is there any prior research?
 - Example: *Thumbs up?: sentiment classification using machine learning techniques*, Pang and Lee, ACL 2002
 - Background: Back in the early 2000s, Internet reviews are not pervasive. Even there are a lot of work on text classification, but not on sentiment.
 - This paper has been cited 6000+ and created a new area called Sentiment Analysis.
 - Nominated as a candidate for NAACL 2018 Test-of-Time Paper.

Novelty: Problem

- New problem is great, but...
 - If you want people to work on your problem, make it easy for them to get started and to measure their progress. Provide:
 - **Test data**, for evaluating the final systems
 - **Development data**, for measuring whether a change to the system helps, and for tuning parameters
 - An **evaluation metric** (formula for measuring how well a system does on the dev or test data)
 - **Labeled training data** and other data resources
 - **Open-source** your code.

Novelty: Task

- New tasks within existing problems are also valuable, so in addition to review sentiment classification:
 - Some recent work in sentiment analysis.
 - Aspect based sentiment analysis: *What is Prof. Wang thinking about Yoshinoya?*
 - Entity-based opinion mining: *How does Courtyard café compare to UCen?*
 - Sentiment summarization: *What is the specialty of Yoshinoya?*

Novelty: Technique

- Is this a new model that significantly addresses the limitation of previous work?
 - HMM→MEMM→CRF→Bi-LSTMs w.Attention
- This is almost a **must** in DL4NLP research.
- This class project is not about running some example code in TensorFlow.
- Good example: *a new attentive seq2seq model with hierarchical encoders that can incorporate structured prior knowledge (Freebase).*

Novelty: Insight

- Can you offer new theoretical or practical insights to existing models?
- For example, Word2Vec's SkipGram model is not well understood.
 - “It's implicitly factorizing a shifted PMI matrix” Levy and Goldberg (2014 NIPS)
 - “It's exponential family principle component analysis” Cotterell et al (2017 EACL)
- Another example: How to explain, visualize, and interpret deep learning models?

Novelty: Dataset/Software

- In computer vision, if there's no **ImageNet** dataset, the progress in CV could be slower.
- In NLP, if there's no **Penn Treebank**, the progress on parsing would be slower.
- If Mikolov did not release his **Word2Vec code**, it will not create such impacts.
- So always try to open-source datasets and code if possible.

Always ask yourself

- What are your contributions?
 - Problem?
 - Task?
 - Theoretical Insight?
 - Technical approach / model?
 - Empirical results?
 - Dataset / Code?

William's guide on

WHAT TO AVOID

Spend the entire quarter on data collection or annotation

- When you have an idea, the first question is: where do I get the data?
- “Neural Relation Extraction from Doctor’s Notes”
 - It’s challenging to obtain a dataset (privacy).
 - Unless you have connections / agreements with doctors and hospitals.
- But I am working on weakly-supervised algorithms
 - You still need ground truth labels for evaluation.

Not enough idea / no insights

- Add one more layer to an existing neural network model without good justification.
- Run an existing model on a new dataset of the same task / problem.
- A 1-2% improvement on Stanford SQuAD dataset, and you don't know why.
- Minor modification of the objective function's loss term and regularization.

Too ambitious

- We only have 9 weeks.
 - It's enough time for picking up low-hanging fruits.
 - So we need reasonable plans that can deliver.
- Some proposals that might be too ambitious:
 - Too many tasks all at once.
 - Too many approaches all at once.
 - Too many promises.
 - Problem too difficult.

Out of scope

- “A novel variational auto-encoder for testing hardware failure”
- “Zero-shot generative adversarial networks for particle simulation”
- “A new kernel for support vector machines”
- These are great ideas, but 291A proposals should be related to representation learning, neural networks, NLP, and related modalities.

Not suitable for DL

- Any problem with training instances $< 10K$ might not be suitable for supervised DL.
- You can often try logistic regression, SVM, or CRF to compare the performance.

No literature survey

- Same idea has been done (even better) in the past.
- No reasonable baselines in mind.
- No plan for comparison.

Not well thought-out

- Vague idea, and no concrete plans.
 - “I sort of want to work on abstractive summarization of tweets...”
- Ethical concerns.
 - “use profile photo recognition and facebook / twitter posts to identify user attributes...”
- Practical concerns.
 - “Large-scale hyperparameter tuning for ResNet”

William's guide on

FINDING PROJECT IDEAS

Pop Quiz: Are these good topics?

- “Extremely Deep CNN-LSTMs for speech recognition”
- “A VAE-GAN hybrid model for dependency parsing”
- It’s hard to compete with the industry on these well-studied topics:
 - Google, MSFT, Apple, Baidu, IBM have tons of engineers working on ASR engineering, and WERs are very low.
 - Google’s Parsey McParseface gives 90+% performances on WSJ/PTB dependency parsing, and 97% on PoS tagging.
- But dependency parsing on other genres is still not well studied (e.g., social media, chat / dialog).

The NLP Research Community

- **Papers**

- [ACL Anthology](#) has nearly everything, free!
 - Over 36,000 papers!
 - Free-text searchable
 - Great way to learn about current research on a topic
 - New search interfaces currently available in beta
 - » Find recent or highly cited work; follow citations
 - Used as a dataset by various projects
 - Analyzing the text of the papers (e.g., parsing it)
 - Extracting a graph of papers, authors, and institutions (Who wrote what? Who works where? What cites what?)

The NLP Research Community

- **Conferences**

- Most work in NLP is published as 8-page conference papers with 3 double-blind reviewers.
- Main annual conferences: ACL, EMNLP, NAACL
 - Also EACL, IJCNLP, COLING
 - + various specialized conferences and workshops
- Big events, and growing fast! [ACL 2015](#):
 - About 1500 attendees
 - 692 full-length papers [submitted](#) (173 accepted)
 - 648 short papers submitted (145 accepted)
 - 14 [workshops](#) on various topics

The NLP Research Community

- **Institutions**

- **Universities:** Many have NLP faculty

- Several “big players” with many faculty
 - Some of them also have good linguistics, cognitive science, machine learning, AI

- **Companies:**

- Old days: AT&T Bell Labs, IBM
 - Now: FAIR, DeepMind, Google, Microsoft, IBM, many startups ...
 - Speech: Nuance, ...
 - Machine translation: Language Weaver, Systran, ...
 - Many niche markets – online reviews, medical transcription, news summarization, legal search and discovery ...

The ML Research Community

- **Papers**

- Most of recent ICML papers:
 - www.jmlr.org/proceedings
- Most of recent NIPS papers:
 - <https://papers.nips.cc/>
- Most of recent ICLR papers:
 - <http://www.iclr.cc/doku.php?id=iclr2017:previousyears>
- Other resources:
 - CVPR/ICCV/AAAI/IJCAI/MLJ/JMLR

Some example project ideas

- Understanding and detecting fake news
- Deep learning for abstractive humor generation
- Automatic detection of toxicity, aggression, and personal attacks in Web dialogues
- Generative adversarial networks for QA
- Learning to parse textual processes
- Learning to reason about requirements
- Efficient methods for learning to learn
- Adversarial variational auto-encoder

My own class projects

- CS6998 (Spring 2010 Columbia):
 - NLP for the Web
 - Project: *automatic vandalism detection in Wikipedia*
 - Deliverable: **COLING 2010.**
- ML10-710 (Fall 2011 CMU)
 - Structured Prediction.
 - Project: *Historical Analysis of Legal Opinions with a Sparse Mixed-Effects Latent Variable Model*
 - Deliverable: **ACL 2012.**

My own class projects (cont'd)

- ML 10-701 (Spring 2013 CMU):
 - CS PhD Machine Learning
 - Project: *A Semiparametric Gaussian Copula Regression Model for Predicting Financial Risks from Earnings Calls*
 - Deliverable: **ACL 2014.**
- LTI 11-712 (Spring 2014 CMU):
 - Lab in NLP.
 - Project: *Dependency Parsing on Chinese Social Media via Probabilistic Programming.*
 - Deliverable: **EMNLP 2014.**

Google's guide on

ML ENGINEERING RULES

Rules of Machine Learning: Best Practices for ML Engineering

(Martin Zinkevich, Google)

- http://martin.zinkevich.org/rules_of_ml/rules_of_ml.pdf
- Rule #2: Make metrics design and implementation a priority.
- Rule #4: Keep the first model simple and get the infrastructure right
- William's note: debug with small subsets first, and then increase the volume.

Rules of Machine Learning: Best Practices for ML Engineering

(Martin Zinkevich, Google)

- Rule #14: Starting with an interpretable model makes debugging easier.
- Rule #22: Clean up features you are no longer using.
- Rule #24: Measure the delta between models.
- Rule #27: Try to quantify observed undesirable behavior

William's NLP/ML engineering advice

- Preprocessing matters!
 - Depends on your applications.
 - Depends on your models.
- Hyperparameter tuning in DL is an art.
 - There are probably 10-20 of them.
 - Refer to literature for default settings.
 - Depends on your application.
 - Try warm-startup the embeddings.
 - Look out for overfitting.

BRAINSTORMING SESSION

Brainstorming Session (Part I)

- Write down two project ideas on your paper, and for each one:
 - Has this problem / task been done before?
 - Can I formulate this problem in Math / ML / DL?
 - Is my technical approach new?
 - Do I have a dataset?
 - Write down the baselines and metrics.

Brainstorming Session (Part II)

- Introduce yourself to your neighbors
 - Your name.
 - Your program and year.
- Pitch your top idea to your neighbor.
- Neighbors: evaluate feasibility and provide feedback.
- Now, switch roles.

Brainstorming Session (Part III)

- Group brainstorming (5 groups):
 - New or improve deep learning models.
 - Group 1: CNNs, RNNs, LSTMs, GRUs, Seq2seq.
 - Group 2: Adversarial and reinforcement learning.
 - Group 3: Theoretical and practical insights of existing DL models.
 - Novel DL4NLP applications.
 - Group 4: text based.
 - Group 5: text and other modalities.