EECS-397 AND EECS-497: SEMINAR IN STATISTICAL LANGUAGE MODELING

Spring 2019

Instructors: David Demeter / Prof. Larry Birnbaum   **Email:** EECS397Northwestern@gmail.com

Time: Tuesdays 6:00-9:00PM   **Place:** MUDD #3514

Course Description: Statistical language models (SLMs) assign probabilities to sequences of words, and are used in systems that perform speech recognition, machine translation, and many other tasks. In recent years, deep neural networks have provided radically improved SLMs. This course will cover both the fundamental technologies that comprise statistical language models, from more classical n-gram models to recent memory-based deep neural networks, along with applications of the techniques to important tasks in artificial intelligence. Students will be required to read and present research papers, and to complete a substantial course project. Students registered for the 497 section of the class will be required to implement two additional projects in TensorFlow extending the work of papers covered in the class.

Office Hours: Before class, or by appointment.

Course Materials: No textbook. The course material will be comprised of research papers in the field.

Course Goals: The goal of this course is to familiarize graduate students and advanced undergraduates with the current state-of-the-art in statistical language modeling. Students will read recently published papers in the field.

Prerequisites: EECS 349 or permission of the instructor.

Grading Policy: Grades are assigned using the standard scale (given in the "introduction" lecture notes), so 93-100 points is an A, 90-93 points is an A-, etc. Points will be allocated as follows:

	397	497
Class Participation	20 pts	10 pts
Leading a Paper Discussion	30 pts	20 pts
Assigned Participation	10 pts	10 pts
Project #1	-	10 pts
Project #2	-	10 pts
Course Project	40 pts	40 pts

Students will be required to lead a paper discussion once individually, or twice as a group of two (preferred). Students not presenting in that week are required to prepare short (two paragraph) summaries of two of the papers being presented that week which should be submitted before the start of class. Students will have to opportunity sign-up for paper presentation slots. Projects should be completed in groups of two (preferred) or individually. Specific project requirements will be provided when assigned.

Course Objectives:

- Have a general understanding of the current state-of-the art in statistical language models.
- Understand how at least one statistical language model is implemented and can be applied (via the course project).
- Be able to understand, and think critically about, recent research papers in the field of statistical language modeling.

Calendar:

Week #1: Introduction

- Introductory Remarks
- Review of Background Materials

1. Bengio, Yoshua et al. "A Neural Probabilistic Language Model." NIPS (2000).

Week #2: Word Embeddings

2. Pennington, Jeffrey et al. "Glove: Global Vectors for Word Representation." EMNLP (2014).
3. Mikolov, Tomas et al. "Linguistic Regularities in Continuous Space Word Representations." HLT-NAACL (2013).
4. Linzen, Tal. "Issues in evaluating semantic spaces using word analogies." RepEval@ACL (2016).
5. Levy, Omer and Yoav Goldberg. "Linguistic Regularities in Sparse and Explicit Word Representations." CoNLL (2014).

Week #3: RNN Architectures

6. Mikolov, Tomas et al. "Recurrent neural network based language model." INTERSPEECH (2010).
7. Zaremba, Wojciech et al. "Recurrent Neural Network Regularization." CoRR abs/1409.2329 (2014).
8. Józefowicz, Rafal et al. "Exploring the Limits of Language Modeling." CoRR abs/1602.02410 (2016).
9. Bojanowski, Piotr et al. "Alternative structures for character-level RNNs." CoRR abs/1511.06303 (2015).

Week #4: Controlling Properties of NNLMs

10. Srivastava, Nitish et al. "Dropout: a simple way to prevent neural networks from overfitting." Journal of Machine Learning Research 15 (2014): 1929-1958.
11. Merity, Stephen et al. "Regularizing and Optimizing LSTM Language Models." CoRR abs/1708.02182 (2018).
12. Jang, Eric et al. "Categorical Reparameterization with Gumbel-Softmax." CoRR abs/1611.01144 (2016).
13. Grave, Edouard et al. "Improving Neural Language Models with a Continuous Cache." CoRR abs/1612.04426 (2017).

- Project #1 Due Friday at 11:59PM
- Course Project Proposals Due Friday at 11:59PM

Week #5: Applications of NNLMs

- Review of implementing language models in TensorFlow

14. Kushman, Nate et al. "Learning to Automatically Solve Algebra Word Problems." ACL (2014).
15. Andreas, Jacob et al. "Learning to Compose Neural Networks for Question Answering." HLT-NAACL (2016).
16. Johnson, Justin et al. "DenseCap: Fully Convolutional Localization Networks for Dense Captioning." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 4565-4574.

Week #6: Improving Efficiency of NNLMs

17. Mikolov, Tomas et al. "Efficient Estimation of Word Representations in Vector Space." CoRR abs/1301.3781 (2013).
18. Mnih, Andriy and Koray Kavukcuoglu. "Learning word embeddings efficiently with noise-contrastive estimation." NIPS (2013).
19. Grave, Edouard et al. "Efficient softmax approximation for GPUs." ICML (2017).
20. Vaswani, Ashish et al. "Attention Is All You Need." NIPS (2017).

Week #7: Text Generation Using NNLMs

21. Graves, Alex. "Generating Sequences With Recurrent Neural Networks." CoRR abs/1308.0850 (2013).
22. Sutskever, Ilya et al. "Sequence to Sequence Learning with Neural Networks." NIPS (2014).
23. Guu, Kelvin et al. "Generating Sentences by Editing Prototypes." Transactions of the Association of Computational Linguistics 6 (2018): 437-450.
24. Noraset, Thanapon et al. "Definition Modeling: Learning to define word embeddings in natural language." AAAI (2017).

- Project #2 Due Friday at 11:59PM
- Course Project Progress Reports Due Friday at 11:59PM

Week #8: Applications of NNLMs

25. Bahdanau, Dzmitry et al. "Neural Machine Translation by Jointly Learning to Align and Translate." CoRR abs/1409.0473 (2015).
26. See, Abigail et al. "Get To The Point: Summarization with Pointer-Generator Networks." ACL (2017).
27. Zhou, Ming et al. "Learning to Generate Product Reviews from Attributes." EACL (2017).
28. Luong, Minh-Thang and Christopher D. Manning. "Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models." CoRR abs/1604.00788 (2016).

Week #9: Current Topics

29. Zellers, Rowan et al. "SWAG: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference." EMNLP (2018).
30. Devlin, Jacob et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." CoRR abs/1810.04805 (2018).
31. Yang, Zhilin et al. "Breaking the Softmax Bottleneck: A High-Rank RNN Language Model." CoRR abs/1711.03953 (2018).
32. Peters, Matthew E. et al. "Deep contextualized word representations." NAACL-HLT (2018).

- Course Project Due Friday at 11:59PM

Weeks #10 and #11: Course Project Presentations