

EECS 395/495: ADVANCED TOPICS IN MACHINE LEARNING

Winter 2017

Instructor:	Doug Downey	Email:	ddowney@eecs.northwestern.edu
TA:	Nishant Subramani	Email:	nishant.subramani23@gmail.com
Time:	Tu/Th 3:30 - 4:50	Place:	Tech F281

Course Pages:

1. http://www.cs.northwestern.edu/~ddowney/courses/395_Winter2017/

Course Description: This course will cover topics in machine learning for statistical language modeling. Statistical language models 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 language models. 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.

Office Hours: After 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: Leading a paper discussion (30%), Class participation (20%), Project (50%).

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.

Background Resources:

1. [Deep Learning Textbook](#)
2. [Deep learning in neural networks: An overview](#) - (Schmidhuber; Science 2015)
3. [A Neural Probabilistic Language Model](#) - (Bengio et al; JMLR 2003)
4. [Advances in NLP](#) - (Hirschberg and Manning; Science 2015)
5. [Practical Recommendations for Gradient-Based Training of Deep Architectures](#) - (Bengio 2012)

Tentative Calendar: note, starting 1/24 we will only discuss one paper; the supplementary papers are for you to look at but you do not need to read or summarize them. The supplements are sometimes but not always related to the discussion paper.

- 1/10
[Recurrent neural network based language model](#) - (Mikolov et al; Interspeech 2010)
[Extensions of recurrent neural network language model](#) - (Mikolov et al; ICASSP 2011)
Presenter: Professor
- 1/12
[Linguistic Regularities in Continuous-Space Recurrent Neural Network Language Models](#) - (Mikolov et. al NAACL-HLT 2013)
[Distributed Representations of Words and Phrases and their Compositionality](#) - (Mikolov et. al 2013)
- 1/17
[GRU Paper: Learning Phrase Representations using RNN EncoderDecoder for Statistical Machine Translation](#) - (Cho et al 2014)
[Gated Feedback Recurrent Neural Networks](#) - (Chung et. al; ICML 2015)
- 1/19
[Pointer Sentinel Mixture Models](#) - (Metamind; 2016)
[Quasi-Recurrent Neural Networks](#) - (Metamind; 2016)
- 1/24
Discussion Paper:
[Generating Sequences with RNNs](#) - (Graves 2013)
Supplement:
[Resource to Understand LSTMs better](#)
[Long Short Term Memory](#) - (Hochreiter & Schmidhuber 1997)
- 1/26
Discussion paper:
[Improved Semantic Representations From Tree-Structured LSTMs](#) - (Tai, Socher, Manning 2015)
Supplement:
[An Empirical Exploration of Recurrent Network Architectures](#) - (Jozefowicz ICML 2015)

- 1/31

Discussion paper:

[A Way out of the Odyssey: Analyzing and Combining Recent Insights for LSTMs](#) - (Metamind; 2016)

Supplement:

[Phased LSTM: Accelerating Recurrent Network Training for Long or Event-based Sequences](#) - (NIPS 2016)

- 2/2

Discussion paper:

[Show, Attend and Tell: Neural Image Caption Generation with Visual Attention](#) - (Xu et al; ICML 2015)

Supplement:

[Resource to understand attention](#)

[Describing Multimedia Content using Attention-based EncoderDecoder Networks](#) - (Cho, Courville, Bengio; IEEE 2015)

- 2/7 – no class

- 2/9

Discussion paper:

[Deep Visual-Semantic Alignments for Generating Image Descriptions](#) - (Karpathy, Fei Fei Li; CVPR 2015)

Supplement:

[Knowing When to Look: Adaptive Attention via A Visual Sentinel for Image Captioning](#) - (Metamind; 2016)

- 2/14

Discussion paper:

[Dynamic Coattention Networks for Question Answering](#) - (Metamind; 2016)

Supplement:

[Effective Approaches to Attention-based Neural Machine Translation](#) - (Luong, Pham, Manning 2015)

- 2/16

Discussion paper:

[Ask Me Anything: Dynamic Memory Networks for Natural Language Processing](#) - (Metamind; 2016)

Supplement:

[Dynamic Memory Networks for Visual and Textual Question Answering](#) - (Metamind; 2016)

- 2/21

Discussion paper:

[Sequence Level Training with RNNs](#) (Facebook; ICLR 2016)

Supplement:

[Sequence to Sequence Learning with Neural Networks](#) - (Sutskever et al NIPS 2014)

- 2/23

Discussion paper:

[A Neural Network for Factoid Question Answering over Paragraphs](#) (Boyd-Graber and others; EMNLP 2014)

Supplement:

[Improving Neural Language Models with a Continuous Cache](#) - (FAIR; ICLR 2017 Under Review)

- 2/28

Discussion paper:

[A Neural Conversational Model](#) - (Vinyals, Le; ICML 2015)

Supplement:

[A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks](#) - (Metamind; 2016)

- 3/2

Papers that we missed/new papers that were not included originally

- 3/7 – project presentations

- 3/9 – project presentations

Tentative Paper Outline:

Background

1. [Deep learning in neural networks: An overview](#) - (Schmidhuber; Science 2015)
2. [A Neural Probabilistic Language Model](#) - (Bengio et al; JMLR 2003)
3. [Advances in NLP](#) - (Hirschberg and Manning; Science 2015)
4. [Distributed Representations of Words and Phrases and their Compositionality](#) - (Mikolov et. al 2013)
5. [Practical Recommendations for Gradient-Based Training of Deep Architectures](#) - (Bengio 2012)

Sequence-to-Sequence

1. [Sequence to Sequence Learning with Neural Networks](#) - (Sutskever et al NIPS 2014)
2. [Sequence Level Training with RNNs](#) (Facebook; ICLR 2016)

Recurrent Neural Networks

- 1.
2. [Gated Feedback Recurrent Neural Networks](#) - (Chung et. al; ICML 2015)
3. [Pointer Sentinel Mixture Models](#) - (Metamind; 2016)
4. [Quasi-Recurrent Neural Networks](#) - (Metamind; 2016)

Long Short-Term Memory Networks

1. [Resource to Understand LSTMs better](#)
2. [Long Short Term Memory](#) - (Hochreiter & Schmidhuber 1997)
3. [Improved Semantic Representations From Tree-Structured LSTMs](#) - (Tai, Socher, Manning 2015)
- 4.
- 5.
- 6.
7. [Phased LSTM: Accelerating Recurrent Network Training for Long or Event-based Sequences](#) - (NIPS 2016)

Dynamic Memory Networks

1. [Ask Me Anything: Dynamic Memory Networks for Natural Language Processing](#) - (Metamind; 2016)
2. [Dynamic Memory Networks for Visual and Textual Question Answering](#) - (Metamind; 2016)

Attention

1. [Resource to understand attention](#)
2. [Describing Multimedia Content using Attention-based EncoderDecoder Networks](#) - (Cho, Courville, Bengio; IEEE 2015)
3. [Show, Attend and Tell: Neural Image Caption Generation with Visual Attention](#) - (Xu et al; ICML 2015)
4. [Effective Approaches to Attention-based Neural Machine Translation](#) - (Luong, Pham, Manning 2015)
5. [Dynamic Coattention Networks for Question Answering](#) - (Metamind; 2016)
6. [Knowing When to Look: Adaptive Attention via A Visual Sentinel for Image Captioning](#) - (Metamind; 2016)

Word Embeddings

1. [GloVe: Global Vectors for Word Representation](#) - (Pennington, Socher, Manning 2014)

Misc

1. [Matching Networks for One Shot Learning](#) - (DeepMind; 2016)
2. [neural machine translation by jointly learning to align and translate](#) - (Bandana Cho Bengio 2015)
3. [A Neural Network for Factoid Question Answering over Paragraphs](#) (Boyd-Graber and others; EMNLP 2014)
4. [Deep Visual-Semantic Alignments for Generating Image Descriptions](#) - (Karpathy, Fei Fei Li; CVPR 2015)
5. [A Neural Conversational Model](#) - (Vinyals, Le; ICML 2015)
6. [A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks](#) - (Metamind; 2016)
7. [Improving Neural Language Models with a Continuous Cache](#) - (FAIR; ICLR 2017 Under Review)