
Preamble

This repository contains the lecture slides and course description for the Deep Natural Language Processing course offered in Hilary Term 2017 at the University of Oxford.

This is an advanced course on natural language processing. Automatically processing natural language inputs and producing language outputs is a key component of Artificial General Intelligence. The ambiguities and noise inherent in human communication render traditional symbolic AI techniques ineffective for representing and analysing language data. Recently statistical techniques based on neural networks have achieved a number of remarkable successes in natural language processing leading to a great deal of commercial and academic interest in the field

This is an applied course focussing on recent advances in analysing and generating speech and text using recurrent neural networks. We introduce the mathematical definitions of the relevant machine learning models and derive their associated optimisation algorithms. The course covers a range of applications of neural networks in NLP including analysing latent dimensions in text, transcribing speech to text, translating between languages, and answering questions. These topics are organised into three high level themes forming a progression from understanding the use of neural networks for sequential language modelling, to understanding their use as conditional language models for transduction tasks, and finally to approaches employing these techniques in combination with other mechanisms for advanced applications. Throughout the course the practical implementation of such models on CPU and GPU hardware is also discussed.

This course is organised by Phil Blunsom and delivered in partnership with the DeepMind Natural Language Research Group.

Lecturers

- Phil Blunsom (Oxford University and DeepMind)
- Chris Dyer (Carnegie Mellon University and DeepMind)
- Edward Grefenstette (DeepMind)
- Karl Moritz Hermann (DeepMind)
- Andrew Senior (DeepMind)
- Wang Ling (DeepMind)
- Jeremy Appleyard (NVIDIA)

TAs

- Yannis Assael
- Yishu Miao
- Brendan Shillingford
- Jan Buys

Timetable

Practicals

- Group 1 - Monday, 9:00-11:00 (Weeks 2-8), 60.05 Thom Building
 - Group 2 - Friday, 16:00-18:00 (Weeks 2-8), Room 379
1. Practical 1: word2vec
 2. Practical 2: text classification
 3. Practical 3: recurrent neural networks for text classification and language modelling
 4. Practical 4: open practical

Lectures

Public Lectures are held in Lecture Theatre 1 of the Maths Institute, on Tuesdays and Thursdays (except week 8), 16:00-18:00 (Hilary Term Weeks 1,3-8).

Lecture Materials

1. Lecture 1a - Introduction [Phil Blunsom]

This lecture introduces the course and motivates why it is interesting to study language processing using Deep Learning techniques.

[slides] [video]

2. Lecture 1b - Deep Neural Networks Are Our Friends [Wang Ling]

This lecture revises basic machine learning concepts that students should know before embarking on this course.

[slides] [video]

3. Lecture 2a- Word Level Semantics [Ed Grefenstette]

Words are the core meaning bearing units in language. Representing and learning the meanings of words is a fundamental task in NLP and in this lecture the concept of a word embedding is introduced as a practical and scalable solution.

[slides] [video]

Reading

Embeddings Basics

- Firth, John R. “A synopsis of linguistic theory, 1930-1955.” (1957): 1-32.
- Curran, James Richard. “From distributional to semantic similarity.” (2004).
- Collobert, Ronan, et al. “Natural language processing (almost) from scratch.” *Journal of Machine Learning Research* 12. Aug (2011): 2493-2537.
- Mikolov, Tomas, et al. “Distributed representations of words and phrases and their compositionality.” *Advances in neural information processing systems*. 2013.

Datasets and Visualisation

- Finkelstein, Lev, et al. “Placing search in context: The concept revisited.” *Proceedings of the 10th international conference on World Wide Web*. ACM, 2001.
- Hill, Felix, Roi Reichart, and Anna Korhonen. “Simlex-999: Evaluating semantic models with (genuine) similarity estimation.” *Computational Linguistics* (2016).
- Maaten, Laurens van der, and Geoffrey Hinton. “Visualizing data using t-SNE.” *Journal of Machine Learning Research* 9.Nov (2008): 2579-2605.

Blog posts

- Deep Learning, NLP, and Representations, Christopher Olah.
- Visualizing Top Tweeps with t-SNE, in Javascript, Andrej Karpathy.

Further Reading

- Hermann, Karl Moritz, and Phil Blunsom. “Multilingual models for compositional distributed semantics.” *arXiv preprint arXiv:1404.4641* (2014).

-
- Levy, Omer, and Yoav Goldberg. “Neural word embedding as implicit matrix factorization.” Advances in neural information processing systems. 2014.
 - Levy, Omer, Yoav Goldberg, and Ido Dagan. “Improving distributional similarity with lessons learned from word embeddings.” Transactions of the Association for Computational Linguistics 3 (2015): 211-225.
 - Ling, Wang, et al. “Two/Too Simple Adaptations of Word2Vec for Syntax Problems.” HLT-NAACL. 2015.

4. Lecture 2b - Overview of the Practicals [Chris Dyer]

This lecture motivates the practical segment of the course.

[slides] [video]

5. Lecture 3 - Language Modelling and RNNs Part 1 [Phil Blunsom]

Language modelling is important task of great practical use in many NLP applications. This lecture introduces language modelling, including traditional n-gram based approaches and more contemporary neural approaches. In particular the popular Recurrent Neural Network (RNN) language model is introduced and its basic training and evaluation algorithms described.

[slides] [video]

Reading

Textbook

- Deep Learning, Chapter 10.

Blogs

- The Unreasonable Effectiveness of Recurrent Neural Networks, Andrej Karpathy.
- The unreasonable effectiveness of Character-level Language Models, Yoav Goldberg.
- Explaining and illustrating orthogonal initialization for recurrent neural networks, Stephen Merity.

6. Lecture 4 - Language Modelling and RNNs Part 2 [Phil Blunsom]

This lecture continues on from the previous one and considers some of the issues involved in producing an effective implementation of an RNN language model. The vanishing and exploding gradient problem is described and architectural solutions, such as Long Short Term Memory (LSTM), are introduced.

[slides] [video]

Reading

Textbook

- Deep Learning, Chapter 10.

Vanishing gradients, LSTMs etc.

- On the difficulty of training recurrent neural networks. Pascanu et al., ICML 2013.
- Long Short-Term Memory. Hochreiter and Schmidhuber, Neural Computation 1997.
- Learning Phrase Representations using RNN EncoderDecoder for Statistical Machine Translation. Cho et al, EMNLP 2014.
- Blog: Understanding LSTM Networks, Christopher Olah.

Dealing with large vocabularies

- A scalable hierarchical distributed language model. Mnih and Hinton, NIPS 2009.
- A fast and simple algorithm for training neural probabilistic language models. Mnih and Teh, ICML 2012.
- On Using Very Large Target Vocabulary for Neural Machine Translation. Jean et al., ACL 2015.
- Exploring the Limits of Language Modeling. Jozefowicz et al., arXiv 2016.
- Efficient softmax approximation for GPUs. Grave et al., arXiv 2016.
- Notes on Noise Contrastive Estimation and Negative Sampling. Dyer, arXiv 2014.
- Pragmatic Neural Language Modelling in Machine Translation. Baltescu and Blunsom, NAACL 2015

Regularisation and dropout

- A Theoretically Grounded Application of Dropout in Recurrent Neural Networks. Gal and Ghahramani, NIPS 2016.
- Blog: Uncertainty in Deep Learning, Yarin Gal.

Other stuff

- Recurrent Highway Networks. Zilly et al., arXiv 2016.
- Capacity and Trainability in Recurrent Neural Networks. Collins et al., arXiv 2016.

7. Lecture 5 - Text Classification [Karl Moritz Hermann]

This lecture discusses text classification, beginning with basic classifiers, such as Naive Bayes, and progressing through to RNNs and Convolution Networks.

[slides] [video]

Reading

- Recurrent Convolutional Neural Networks for Text Classification. Lai et al. AAAI 2015.
- A Convolutional Neural Network for Modelling Sentences, Kalchbrenner et al. ACL 2014.
- Semantic compositionality through recursive matrix-vector, Socher et al. EMNLP 2012.
- Blog: Understanding Convolution Neural Networks For NLP, Denny Britz.
- Thesis: Distributional Representations for Compositional Semantics, Hermann (2014).

8. Lecture 6 - Deep NLP on Nvidia GPUs [Jeremy Appleyard]

This lecture introduces Graphical Processing Units (GPUs) as an alternative to CPUs for executing Deep Learning algorithms. The strengths and weaknesses of GPUs are discussed as well as the importance of understanding how memory bandwidth and computation impact throughput for RNNs.

[slides] [video]

Reading

- Optimizing Performance of Recurrent Neural Networks on GPUs. Appleyard et al., arXiv 2016.
- Persistent RNNs: Stashing Recurrent Weights On-Chip, Diamos et al., ICML 2016
- Efficient softmax approximation for GPUs. Grave et al., arXiv 2016.

9. Lecture 7 - Conditional Language Models [Chris Dyer]

In this lecture we extend the concept of language modelling to incorporate prior information. By conditioning an RNN language model on an input representation we can generate contextually relevant

language. This very general idea can be applied to transduce sequences into new sequences for tasks such as translation and summarisation, or images into captions describing their content.

[slides] [video]

Reading

- Recurrent Continuous Translation Models. Kalchbrenner and Blunsom, EMNLP 2013
- Sequence to Sequence Learning with Neural Networks. Sutskever et al., NIPS 2014
- Multimodal Neural Language Models. Kiros et al., ICML 2014
- Show and Tell: A Neural Image Caption Generator. Vinyals et al., CVPR 2015

10. Lecture 8 - Generating Language with Attention [Chris Dyer]

This lecture introduces one of the most important and influential mechanisms employed in Deep Neural Networks: Attention. Attention augments recurrent networks with the ability to condition on specific parts of the input and is key to achieving high performance in tasks such as Machine Translation and Image Captioning.

[slides] [video]

Reading

- Neural Machine Translation by Jointly Learning to Align and Translate. Bahdanau et al., ICLR 2015
- Show, Attend, and Tell: Neural Image Caption Generation with Visual Attention. Xu et al., ICML 2015
- Incorporating structural alignment biases into an attentional neural translation model. Cohn et al., NAACL 2016
- BLEU: a Method for Automatic Evaluation of Machine Translation. Papineni et al, ACL 2002

11. Lecture 9 - Speech Recognition (ASR) [Andrew Senior]

Automatic Speech Recognition (ASR) is the task of transducing raw audio signals of spoken language into text transcriptions. This talk covers the history of ASR models, from Gaussian Mixtures to attention augmented RNNs, the basic linguistics of speech, and the various input and output representations frequently employed.

[slides] [video]

12. Lecture 10 - Text to Speech (TTS) [Andrew Senior]

This lecture introduces algorithms for converting written language into spoken language (Text to Speech). TTS is the inverse process to ASR, but there are some important differences in the models applied. Here we review traditional TTS models, and then cover more recent neural approaches such as DeepMind's WaveNet model.

[slides] [video]

13. Lecture 11 - Question Answering [Karl Moritz Hermann]

[slides] [video]

Reading

- Teaching machines to read and comprehend. Hermann et al., NIPS 2015
- Deep Learning for Answer Sentence Selection. Yu et al., NIPS Deep Learning Workshop 2014

14. Lecture 12 - Memory [Ed Grefenstette]

[slides] [video]

Reading

- Hybrid computing using a neural network with dynamic external memory. Graves et al., Nature 2016
- Reasoning about Entailment with Neural Attention. Rocktäschel et al., ICLR 2016
- Learning to transduce with unbounded memory. Grefenstette et al., NIPS 2015
- End-to-End Memory Networks. Sukhbaatar et al., NIPS 2015

15. Lecture 13 - Linguistic Knowledge in Neural Networks

[slides] [video]

Piazza

We will be using Piazza to facilitate class discussion during the course. Rather than emailing questions directly, I encourage you to post your questions on Piazza to be answered by your fellow students, instructors, and lecturers. However do please do note that all the lecturers for this course are volunteering their time and may not always be available to give a response.

Find our class page at: <https://piazza.com/ox.ac.uk/winter2017/dnlpht2017/home>

Assessment

The primary assessment for this course will be a take-home assignment issued at the end of the term. This assignment will ask questions drawing on the concepts and models discussed in the course, as well as from selected research publications. The nature of the questions will include analysing mathematical descriptions of models and proposing extensions, improvements, or evaluations to such models. The assignment may also ask students to read specific research publications and discuss their proposed algorithms in the context of the course. In answering questions students will be expected to both present coherent written arguments and use appropriate mathematical formulae, and possibly pseudo-code, to illustrate answers.

The practical component of the course will be assessed in the usual way.

Acknowledgements

This course would not have been possible without the support of DeepMind, The University of Oxford Department of Computer Science, Nvidia, and the generous donation of GPU resources from Microsoft Azure.