DL/NLP Resources

Dr. Wenye Li
wyli@cuhk.edu.cn
Shenzhen Research Institute of Big Data

Last Revision: September 19, 2016

Table of Contents

Deep Learning and NLP

Books

- *Deep Learning*, by I. Goodfellow, Y. Bengio, A. Courville, (in preparation for publication by MIT Press, availabe at: `http://www.deeplearningbook.org/`).
- *Neural Networks, Tricks of the Trade* (2nd edition), edited by G. Montavon, G. Orr, and K. Muller, Springer, 2012.

Tutorials

- `https://nips.cc/Conferences/2015/Schedule?event=4891`: *Deep Learning NIPS'2015 Tutorial*, by G. Hinton, Y. Bengio, and Y. LeCun. also: `https://www.iro.umontreal.ca/~bengioy/talks/DL-Tutorial-NIPS2015.pdf`.
- `http://deeplearning.net/tutorial/deeplearning.pdf`: *Deep Learning Tutorial*, by LISA lab, University of Montreal, 2015.

Papers on Deep Learning

1. Learning Deep Architectures for AI
   Y. Bengio (Montreal Univ)
   `https://www.iro.umontreal.ca/~lisa/pointeurs/TR1312.pdf`
   Note: Foundations and Trends in Machine Learning, 2009

2. Deep Learning
   Y. LeCun, Y. Bengio, and G. Hinton (New York Univ, Montreal Univ, Toroto Univ)
   `www.nature.com/articles/nature14539`
   Note: Nature'2015

3. Deep Learning in Neural Networks: An Overview
   J. Schmidhuber (IDSIA Switzerland)
   `https://arxiv.org/abs/1404.7828`
   Note: Neural Networks, 2015, with 888 references

4. A Fast Learning Algorithm for Deep Belief Nets
   G. Hinton, S. Osindero, and Y.W. Teh (Toronto Univ)
   `https://www.cs.toronto.edu/~hinton/absps/fastnc.pdf`
   Note: Neural Computation, 2006

5. Learning representations by back-propagating errors
   D. Rumelhart, G. Hinton, R. Williams (UCSD, CMU)
   `http://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf`.
   Note: Nature'1988, Neural Networks and backpropagation – for named entity recognition

6. Imagenet classification with deep convolutional neural networks
   A. Krizhevsky, I. Sutskever, G. Hinton (Toroto Univ)
   `https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-network.pdf`
   Note: NIPS'2012.

Papers on Language Understanding

The following is a collection of papers on language understanding and related techniques with deep learning techniques. A number of reviews are given first, followed by specific techniques in NLP and suggested readings from Dr. He Xiaodong. Traditional (non-DL) NLP techniques are not listed. If needed, the readings can be obtained easily for the topics of sentence segmentation, word tokenizing, word normalization, n-grams, named entity recognition, part of speech tagging, etc.

1. Advances in natural language processing.
   Julia Hirschberg, Christopher D. Manning (Columbia University, Stanford University)
   `http://science.sciencemag.org/content/349/6245/261`
   Note: Science' 2015, A recent review

2. Ask Me Anything: Dynamic Memory Networks for Natural Language Processing.
   Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, Richard Socher (MetaMind, Palo Alto, CA)
   `http://arxiv.org/abs/1506.07285`.
   Note: arXiv:1506.07285, 2015, A dynamic memory network which processes input sequences and questions, forms episodic memories, and generates relevant answers.

3. A Primer on Neural Network Models for Natural Language Processing.
   Yoav Goldberg (Bar Ilan University, Israel)
   `https://arxiv.org/abs/1510.00726`.
   Note: arXiv:1510.00726, 2015, 75 page summary of state-of-the-art.

4. Statistical Language Models based on Neural Networks.
   Tomas Mikolov (Brno University of Technology)
   `http://www.fit.vutbr.cz/~imikolov/rnnlm/thesis.pdf`
   Note: A Ph.D thesis

5. A Neural Probabilistic Language Model.
   Yoshua Bengio, Rejean Ducharme, Pascal Vincent, Christian Jauvin (Montreal University)
   `http://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf`.
   Note: Journal of Machine Learning Research, 2003, Seminal paper on word vectors.

6. A Neural Conversational Model.
   Oriol Vinyals, Quoc Le (Google)
   `http://arxiv.org/abs/1506.05869`.
   Note: arXiv:1506.05869, 2015, Conversational modelling

7. From Frequency to Meaning: Vector Space Models of Semantics.
   Peter D. Turney, Patrick Pantel (National Research Council Canada)
   `http://www.jair.org/media/2934/live-2934-4846-jair.pdf`.
   Note: Journal of artificial intelligence research, 2010, A must-read for semantics representation.

8. Distributed Representations of Words and Phrases and their Compositionality.
   Tomas Miklov et al. (Google)
   `http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-com`
   `pdf`.
   Note: Advances in neural information processing systems, 2013, Simple Word Vector representations: word2vec, GloVe

9. Efficient Estimation of Word Representations in Vector Space.
   Tomas Miklov et al. (Google)
   `http://arxiv.org/pdf/1301.3781.pdf`.
   Note: arXiv:1301.3781, 2013, Simple Word Vector representations: word2vec, GloVe

10. GloVe: Global Vectors for Word Representation.
    Jeffrey Pennington, Richard Socher, Christopher D. Manning (Stanford)
    `http://www-nlp.stanford.edu/pubs/glove.pdf`.
    Note: EMNLP'2014, Advanced word vector representations: language models, softmax, single layer networks

11. Improving Word Representations via Global Context and Multiple Word Prototypes.
    Eric H. Huang, Richard Socher, Christopher D. Manning, Andrew Y. Ng (Stanford)
    `http://www.aclweb.org/anthology/P12-1092`.
    Note: ACL'2012, Advanced word vector representations: language models, softmax, single layer networks

12. Natural Language Processing (almost) from Scratch.
    Ronan Collobert, Jason Weston, Leon Bottou, Michael Karlen, Koray Kavukcuoglu, Pavel Kuksa (NEC Lab)
    `http://arxiv.org/pdf/1103.0398v1.pdf`.
    Note: Journal of Machine Learning Research, 2011, Project Advice, Neural Networks and Back-Prop

13. A Neural Network for Factoid Question Answering over Paragraphs.
    Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, Hal Daume III (Maryland University)
    `https://cs.umd.edu/~miyyer/pubs/2014_qb_rnn.pdf`.
    Note: EMNLP'2014, Project Advice, Neural Networks and Back-Prop

14. Grounded Compositional Semantics for Finding and Describing Images with Sentences.
    Richard Socher, Andrej Karpathy, Quoc V. Le*, Christopher D. Manning, Andrew Y. Ng (Stanford)
    `http://nlp.stanford.edu/~socherr/SocherKarpathyLeManningNg_TACL2013.pdf`.
    Note: Transactions of the Association for Computational Linguistics, 2014, Project Advice, Neural Networks and Back-Prop

15. Deep Visual-Semantic Alignments for Generating Image Descriptions.
    Andrej Karpathy, Li Fei-Fei (Stanford)
    `http://cs.stanford.edu/people/karpathy/deepimagesent/devisagen.pdf`.
    Note: CVPR'2015, Project Advice, Neural Networks and Back-Prop

16. Practical recommendations for gradient-based training of deep architectures.
    Yoshua Bengio (Montreal Univ)
    `http://arxiv.org/abs/1206.5533`.
    Note: in Neural Networks: Tricks of the Trade, 2012, Practical tips: gradient checks, overfitting, regularization, activation functions, details

17. Recurrent neural network based language model.
    Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan Honza Cernock, Sanjeev Khudanpur (Brno Univ of Technology, John Hopkins Univ)
    `http://www.fit.vutbr.cz/research/groups/speech/publi/2010/mikolov_interspeech2010_IS100722.pdf`.
    Note: Recurrent neural networks – for language modeling and other tasks

18. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank.
    Richard Socher et al. (Stanford)

`http://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf`.
Note: EMNLP'2013, Recursive neural networks – for different tasks (e.g. sentiment analysis)

19. Extensions of recurrent neural network language model.
Tomas Mikolov et al. (Brno Univ of Technology)
`http://www.fit.vutbr.cz/research/groups/speech/publi/2011/mikolov_icassp2011_5528.pdf`.
Note: ICASSP'2011, Recurrent neural networks – for language modeling and other tasks

20. Opinion Mining with Deep Recurrent Neural Networks.
Ozan Irsoy, Claire Cardie (Cornel Univ)
`http://www.cs.cornell.edu/~oirsoy/drnt.htm`.
Note: EMNLP'2014, Recurrent neural networks – for language modeling and other tasks

21. Long Short-Term Memory.
Sepp Hochreiter, Jurgen Schmidhuber (TU Munich, IDSIA Switzerland)
`http://web.eecs.utk.edu/~itamar/courses/ECE-692/Bobby_paper1.pdf`.
Note: Neural computation, 1997, GRUs and LSTMs in language processing

22. Gated Feedback Recurrent Neural Networks.
Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, Yoshua Bengio (Montreal Univ)
`http://arxiv.org/pdf/1502.02367v3.pdf`.
Note: CoRR, abs/1502.02367, 2015, GRUs and LSTMs in language processing

23. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.
Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, Yoshua Bengio (Montreal Univ)
`http://arxiv.org/pdf/1412.3555v1.pdf`
Note: arXiv:1412.3555, 2014, GRUs and LSTMs in language processing

24. Parsing with Compositional Vector Grammars.
Richard Socher, Cliff Chiung-Yu Lin, Andrew Y. Ng, Christopher D. Manning (Stanford)
`http://nlp.stanford.edu/pubs/SocherBauerManningNg_ACL2013.pdf`.
Note: ACL'2013, Recursive neural networks – for parsing

25. Parsing Natural Scenes and Natural Language with Recursive Neural Networks.
Richard Socher, John Bauer, Christopher D. Manning, Andrew Y. Ng (Stanford)
`http://www-nlp.stanford.edu/pubs/SocherLinNgManning_ICML2011.pdf`.
Note: ICML'2011, Recursive neural networks – for parsing

26. Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection.
Richard Socher, Eric H. Huang, Jeffrey Pennington, Andrew Y. Ng, Christopher D. Manning (Stanford)
`http://papers.nips.cc/paper/4204-dynamic-pooling-and-unfolding-recursive-autoencoders-for-parap`
`pdf`.
Note: NIPS'2011, Recursive neural networks – for different tasks (e.g. sentiment analysis)

27. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks.
Kai Sheng Tai, Richard Socher, Christopher D. Manning (Stanford)
`http://arxiv.org/pdf/1503.00075v2.pdf`.
Note: arXiv:1503.00075, 2015, Recursive neural networks – for different tasks (e.g. sentiment analysis)

28. A Convolutional Neural Network for Modelling Sentences.
Nal Kalchbrenner, Edward Grefenstette, Phil Blunsom (Oxford Univ)

http://www.aclweb.org/anthology/P14-1062.
Note: ACL'2014, Convolutional neural networks – for sentence classification

29. Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models.
Minh-Thang Luong, Christopher D. Manning (Stanford)
https://arxiv.org/abs/1604.00788.
Note: arXiv:1604.00788, 2016, Deep Learning in machine translation

30. Addressing the Rare Word Problem in Neural Machine Translation.
Minh-Thang Luong, Ilya Sutskever, Quoc V. Le, Oriol Vinyals, Wojciech Zaremba (Stanford, Google)
https://arxiv.org/abs/1410.8206.
Note: arXiv:1410.8206, 2014, Deep Learning in machine translation

31. Neural Machine Translation by Jointly Learning to Align and Translate.
Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio (Montreal Univ)
https://arxiv.org/abs/1409.0473.
Note: arXiv:1409.0473, 2014, Deep Learning in machine translation

32. Sequence to Sequence Learning with Neural Networks.
Ilya Sutskever, Oriol Vinyals, Quoc V. Le (Google)
https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf.
Note: NIPS'2014, sequence to sequence learning

33. Neural Programmer: Inducing Latent Programs with Gradient Descent.
Arvind Neelakantan, Quoc V. Le, Ilya Sutskever (University of Massachusetts Amherst, Google)
http://arxiv.org/abs/1511.04834.
Note: arXiv:1511.04834, 2015, end-to-end differentiable neural network with basic arithmetic and logic operations

34. Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing.
Antoine Bordes, Xavier Glorot, Jason Weston, Yoshua Bengio (UTC France, Montreal Univ)
http://www.jmlr.org/proceedings/papers/v22/bordes12/bordes12.pdf.
Note: AISTATS'2012, learning meaning representation

35. Semi-supervised recursive autoencoders for predicting sentiment distributions.
Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, Christopher D. Manning (Stanford)
http://nlp.stanford.edu/pubs/SocherPenningtonHuangNgManning_EMNLP2011.pdf.
Note: EMNLP'2011, sentence-level prediction of sentiment label distributions

36. Framewise phoneme classification with bidirectional LSTM and other neural network architectures.
Alex Graves, Jurgen Schmidhuber (IDSIA Switzerland, TU Munich)
ftp://ftp.idsia.ch/pub/juergen/nn_2005.pdf.
Note: Neural Networks, 2005, framewise phoneme classification

37. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.
Kyunghyun Cho et al. (Montreal Univ)
https://arxiv.org/abs/1406.1078.
Note: arXiv:1406.1078, 2014, learning a fixed-length vector representation of phrases

38. Character-level Convolutional Networks for Text Classification.
Xiang Zhang, Junbo Zhao, Yann LeCun (New York Univ)

http://arxiv.org/abs/1509.01626
Note: NIPS'2015, ConvNets for text classification.

39. Learning Deep Structured Semantic Models for Web Search using Clickthrough Data
Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, Larry Heck (Microsoft Research)
https://www.microsoft.com/en-us/research/publication/learning-deep-structured-semantic-models-f
Note: CIKM'2013, suggested by Dr. He Xiaodong

40. Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Spoken Language Understanding
Gregoire Mesnil, Xiaodong He, Li Deng, Yoshua Bengio (Microsoft Research, Montreal Univ)
https://www.iro.umontreal.ca/~lisa/pointeurs/RNNSpokenLanguage2013.pdf
Note: Interspeech 2013, suggested by Dr. He Xiaodong

41. A Latent Semantic Model with Convolutional-Pooling Structure for Information Retrieval
Xiaodong He, Jianfeng Gao, Li Deng, Gregoire Mesnil (Microsoft Research)
https://www.microsoft.com/en-us/research/publication/a-latent-semantic-model-with-convolutional-
Note: CIKM'2014, suggested by Dr. He Xiaodong

42. Semantic Parsing for Single-Relation Question Answering
Wen-tau Yih, Xiaodong He, Christopher Meek, Scott Wen-tau Yih, Chris Meek (Microsoft Research)
https://www.microsoft.com/en-us/research/publication/semantic-parsing-for-single-relation-questi
Note: ACL'2014, suggested by Dr. He Xiaodong

43. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base
Wen-tau Yih, Ming-Wei Chang, Xiaodong He, Jianfeng Gao, Scott Wen-tau Yih (Microsoft Research)
https://www.microsoft.com/en-us/research/publication/semantic-parsing-via-staged-query-graph-ge
Note: AFNLP'2015, suggested by Dr. He Xiaodong

44. From Captions to Visual Concepts and Back
Hao Fang et al. (Microsoft Research)
https://arxiv.org/abs/1411.4952
Note: CVPR'2015, suggested by Dr. He Xiaodong

45. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems
Ali Mamdouh Elkahky, Yang Song, Xiaodong He
https://www.microsoft.com/en-us/research/publication/a-multi-view-deep-learning-approach-for-cr
Note: WWW'2015, suggested by Dr. He Xiaodong

46. Deep Reinforcement Learning with a Natural Language Action Space (Microsoft Research)
Ji He, Jianshu Chen, Xiaodong He, Jianfeng Gao, Lihong Li, Li Deng, Mari Ostendorf
https://arxiv.org/abs/1511.04636
Note: ACL'2016, suggested by Dr. He Xiaodong

47. Rich Image Captioning in the Wild
Kenneth Tran et al. (Microsoft Research)
http://arxiv.org/abs/1603.09016
Note: CVPR'2016, suggested by Dr. He Xiaodong

48. Deep Sentence Embedding Using Long Short-Term Memory Networks: Analysis and Application to Information Retrieval

Hamid Palangi et al. (Microsoft Research)
`https://arxiv.org/abs/1502.06922`
Note: IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, suggested by Dr. He Xiaodong

49. Character-Level Question Answering with Attention
David Golub, Xiaodong He (Microsoft Research)
`https://arxiv.org/abs/1604.00727`
Note: EMNLP'2016, suggested by Dr. He Xiaodong

50. Stacked Attention Networks for Image Question Answering
Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Smola (Microsoft Research, CMU)
`https://arxiv.org/abs/1511.02274`
Note: CVPR'2016, suggested by Dr. He Xiaodong

Groups

- G. Hinton (U. Toronto, Google) and his group.
- Y. LeCunn (New York Univ., Facebookresearch) and his group.
- Y. Bengio (U. Montreal) and his group.
- A. Ng (Stanford, Baidu) and his group.
- T. Poggio (MIT) and his group.
- C. Manning (Stanford) and his group.
- Li Deng (Microsoft Research) and his group.
- ...

Conferences and Journals

- NIPS: Annual Conference on Neural Inforatmion Processing, http://www.nips.cc/
- ICML: International Conference on Machine Learning, http://www.icml.cc/
- ICPR: International Conference on Learning Representations, http://www.iclr.cc/
- ACL: Annual meeting of the Association for Computational Linguistics
- EMNLP: Conference on Empirical Methods on Natural Language Processing
- JMLR: Journal of Machine Learning Research
- TPAMI: IEEE Transactions on Pattern Analysis and Machine Intelligence
- TNNLS: IEEE Transactions on Neural Networks and Learning System
- Neural Computation

Software

- Theano: CPU/GPU symbolic expression compiler in python. `http://deeplearning.net/software/theano/`
- Torch: a Matlab-like environment for state-of-the-art machine learning algorithms. `http://torch.ch/`
- TensorFlow: an open source software library for numerical computation using data flow graphs. `http://www.tensorflow.org/`
- Caffe: a deep learning framework made with expression, speed, and modularity in mind. `http://caffe.berkeleyvision.org/`
- CNTK: a unified deep-learning toolkit by Microsoft Research. `https://github.com/Microsoft/CNTK/wiki`
- ...

MISC

- `https://en.wikipedia.org/wiki/Deep_learning`
- `http://deeplearning.stanford.edu/`
- `http://deeplearning.net/`
- `https://www.technologyreview.com/s/513696/deep-learning/`

Revision History

- 2016-09-09, initial version with deep learning resources.
- 2016-09-17, added language understanding readings.
- 2016-09-18, added authors&affiliations and suggested papers from Dr. He Xiaodong.