Proposal for Tutorial on

Deep Learning and Distributed Word Representations

Proposed by

Sudha Bhingardive
Rudramurthy V
Kevin Patel
Prerana Singhal

Under direction of
Prof. Pushpak Bhattacharyya

Department of Computer Science & Engineering
IIT Bombay

Abstract

Deep learning techniques have demonstrated tremendous success in the natural language processing community in recent times. The goal is to move machine learning closer to one of its original goals: *Artificial Intelligence*. Deep learning involves learning multiple levels of representation and abstraction that help to make sense of data such as images, sound, and text. Various natural language processing applications have shown state-of-the-art results by using deep learning architectures. Much of these successes can be attributed to deep learning's ability to lend itself to representation learning for words. The focus of this tutorial is to establish a strong foundation of deep learning, give an extensive overview of existing deep learning approaches, and introduce different word representation approaches. This tutorial will help the audience in solving problems in language or text processing.

This tutorial will span over three parts. In the first part, we will discuss the basics of deep learning, it's architectures, distributed representations and various approaches to build such representations. We will direct the audience towards ready-to-use pre-trained models, open source tools to train your own models, and methods to evaluate these models.

In the second part of the tutorial, we place particular emphasis on several important applications, including (1) Named Entity Recognition, (2) Word Sense Disambiguation, (3) WordNet linking and (4) Sentiment Analysis. For each application, we will discuss what particular architectures of deep learning models are suitable given the nature of the application, and how learning can be performed efficiently and effectively using end-to-end optimization strategies.

In the third part, we are planning to have hands-on-sessions on various popular deep learning tools, and how deep learning techniques can be employed in various applications. Depending on availability of time, we plan to guide the audience to build either one of the following sample applications:
- 'Odd-one-out' application (find the outlier word in a set of words)
- 'Word origin annotation' application (detecting transliterated words in running text)

Proposed Duration: Full Day

Outline Plan

Part I: Basics
- Deep Learning (30 slides, 60 minutes)
  - Introduction [7][8][17][18][19]
    - What is deep learning?
    - Why deep learning?
  - Background [16][17][18][23][25]
    - From Perceptron to Neural Networks
      - Feedforward Neural Networks
      - Recurrent Neural Networks
      - Convolutional Networks
    - Challenges in training Neural networks
    - Unsupervised PreTraining
      - Autoencoders
      - Restricted Boltzmann Machines
- Distributed Representations (30 slides, 60 minutes)
  - Introduction
  - Representation Models [1][2][3][5][6][9]
    - Count based methods
    - Prediction based methods
    - Continuous Language models
  - Evaluation of representations [26][27][28]
    - Different evaluation tasks suggested in the literature
    - Different comparisons available in literature
    - Different comparisons done by us

Part 2: Applications
- Named Entity Recognition (15 slides, 30 minutes)
  - Introduction
  - Deep Learning models for Named Entity Recognition [2][22]
  - Results and Future directions
- Sentiment Analysis (15 slides, 30 minutes)
  - Introduction
  - Deep Learning models for Sentiment Analysis [15][16][25]
  - Results and Future directions
- Word Sense Disambiguation (15 slides, 30 minutes)
  - Introduction
  - Deep Learning models for Word Sense Disambiguation [11][12][13][14]
  - Results and Future directions
- WordNet Linking(15 slides, 30 minutes)
  - Overview of WordNet Linking process
  - Deep Learning models for WordNet Linking [10][11][14][20][21]
  - Results and Future directions

Part III: Hands on sessions
- Using Deep Learning tools for sample applications (15 slides, 90 minutes)
    - Theano
    - Word2vec [5][6]
    - Sent2vec [4]
    - Doc2vec [4]

Conclusions and Discussions (5/6 slides, 30 minutes)

Presenters:

Sudha Bhingardive
PhD Scholar, Dept of Computer Science & Engineering, IIT Bombay.
Sudha is a PhD student at IIT Bombay, since July 2011. She is working on various topics like Word Sense Disambiguation, Word Representations, WordNet Enrichment, *etc*. She also co-ordinates IndoWordNet related activities, and has been involved in organizing workshops such as Asian Languages Resources (ALR) at IJCNLP 2013. She has publications in top NLP conferences such as **ACL, NAACL, AAAI,** and **GWC**.

Rudramurthy V
PhD Scholar, Dept of Computer Science & Engineering, IIT Bombay.
Rudramurthy is a PhD student at IIT Bombay, since July 2013. He is working on areas of NLP like Deep Learning, MultiWord Expressions, *etc*. He has publication in **NAACL** conference.

Kevin Patel
PhD Scholar, Dept of Computer Science & Engineering, IIT Bombay.
Kevin Patel is a PhD student at IIT Bombay, since July 2014. He is working on WordNet linking, Deep Learning, Word Sense Disambiguation, MultiWord Expressions *etc*.

Prerana Singhal
M.tech Student, Dept of Computer Science & Engineering, IIT Bombay.
Prerana is currently pursuing her Masters at IIT Bombay in Dept. of Computer Science and Engineering. Her Masters thesis deals with Sentiment Analysis, Sentence Classification and Deep Learning.

All above mentioned are guided by:
Dr. Pushpak Bhattacharyya
Director, IIT Patna
Vijay and Sita Vashee Chair Professor
Department of Computer Science and Engineering
IIT Bombay
http://www.cse.iitb.ac.in/~pb

References:

[1] Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Janvin. "**A neural probabilistic language model**." *The Journal of Machine Learning Research* 3 (2003): 1137-1155.

[2] Collobert, Ronan, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. "**Natural language processing (almost) from scratch**." *The Journal of Machine Learning Research 12 (2011): 2493-2537*.

[3] Collobert, Ronan, and Jason Weston, 2008, "**A unified architecture for natural language processing: deep neural networks with multitask learning**." in I*CML, ACM International Conference Proceeding Series, Vol. 307, edited by William W. Cohen, Andrew McCallum, and Sam T. Roweis (ACM) pp. 160–167.*

[4] Quoc V, Le and Mikolov, Tomas 2014, "**Distributed representations of sentences and documents**," *arXiv preprint arXiv:1405.4053.*

[5] Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "**Efficient estimation of word representations in vector space**." *arXiv preprint arXiv:1301.3781* (2013).

[6] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "**Distributed representations of words and phrases and their compositionality.**" In *Advances in neural information processing systems*, pp. 3111-3119. 2013.

[7] Xavier, Glorot and Bengio, Yoshua, 2010, "**Understanding the difficulty of training deep feedforward neural networks**," *in International conference on artificial intelligence and statistics, pp. 249–256.*

[8] James, Bergstra and Bengio, Yoshua 2012, "**Random search for hyper-parameter optimization**," *The Journal of Machine Learning Research 13, 281–305.*

[9] Huang, Eric H., Richard Socher, Christopher D. Manning, and Andrew Y. Ng, 2012, "**Improving word representations via global context and multiple word prototypes,**" *ACL 2012.*

[10] Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. "**Linguistic Regularities in Continuous Space Word Representations**." In *HLT-NAACL*, pp. 746-751. 2013.

[11] Faruqui, Manaal, Jesse Dodge, Sujay K. Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. "**Retrofitting word vectors to semantic lexicons.**" *arXiv preprint arXiv:1411.4166* (2014).

[12] Chen, Xinxiong, Zhiyuan Liu, and Maosong Sun, 2014, "**A unified model for word sense representation and disambiguation**," *in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1025–1035.*

[13] Wiriyathammabhum, Peratham, Boonserm Kijsirikul, Hiroya Takamura, and Manabu Okumura. "**Applying Deep Belief Networks to Word Sense Disambiguation**", *arXiv preprint arXiv:1207.0396 (2012).*

[14] Sascha Rothe and Hinrich Schütze, "**AUTOEXTEND: Extending Word Embeddings to Embeddings for Synsets and Lexemes**"., *ACL 2015.*

[15] Richard Socher, and Jeffrey Pennington, and Eric H. Huang, and Andrew Y. Ng, and Christopher D. Manning "**Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions**", *EMNLP 2011.*

[16] dos Santos, Cıcero Nogueira, and Maıra Gatti. "**Deep convolutional neural networks for sentiment analysis of short texts**." In *Proceedings of the 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland.* 2014

[17] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "**A fast learning algorithm for deep belief nets**." *Neural computation 18, no. 7 (2006): 1527-1554.*

[18] Bengio, Yoshua, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. "**Greedy layer-wise training of deep networks.**" *Advances in neural information processing systems 19 (2007): 153.*

[19] Bengio, Yoshua. "**Learning deep architectures for AI**." *Foundations and trends® in Machine Learning 2, no. 1 (2009): 1-127.*

[20] Joshi Salil, Chatterjee, Arindam Arun Karthikeyan Karra, and Pushpak Bhattacharyya. "**Eating your own cooking: automatically linking wordnet synsets of two languages.**" *COLING, 2012.*

[21] Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever. "**Exploiting similarities among languages for machine translation**", *arXiv preprint arXiv:1309.4168* (2013).

[22] Joseph Turian, Lev Ratinov, and Yoshua Bengio. 2010. "**Word representations: a simple and general method for semi-supervised learning**". In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (ACL '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 384-394.

[23] Sutskever, I. Vinyals, O. & Le. Q. V. "**Sequence to sequence learning with neural networks**". *In Proc. Advances in Neural Information Processing Systems 27 3104–3112* (2014).

[24] Bengio, Y., Courville, A. & Vincent, P. "**Representation learning: a review and new perspectives**". *IEEE Trans. Pattern Anal. Machine Intell. 35, 1798–1828 (2013).*

[25] Quoc V, Le and Mikolov, Tomas 2014, "**Distributed representations of sentences and documents**," *arXiv preprint arXiv:1405.4053.*

[26] Baroni, Marco, Georgiana Dinu, and Germán Kruszewski. "**Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors.**" In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 238-247. 2014.

[27] Hill, Felix, Roi Reichart, and Anna Korhonen. "**Simlex-999: Evaluating semantic models with (genuine) similarity estimation.**" *arXiv preprint arXiv:1408.3456* (2014).

[28] Hill, Felix, KyungHyun Cho, Sebastien Jean, Coline Devin, and Yoshua Bengio. "**Not All Neural Embeddings are Born Equal.**" *arXiv preprint arXiv:1410.0718* (2014).