Deep Learning for Semantic Composition

Xiaodan Zhu
Queen's University
National Research Council Canada
zhu2048@gmail.com

Edward Grefenstette
DeepMind
etg@google.com

1 Tutorial Description

Learning representations to model the meaning of text has been a core problem in natural language understanding (NLP). The last several years have seen extensive interests on distributional approaches, in which text spans of different granularities are encoded as continuous vectors. If properly learned, such representations have been shown to help achieve the state-of-the-art performances on a variety of NLP problems.

In this tutorial, we will cover the fundamentals and selected research topics on neural network-based modeling for semantic composition, which aims to learn distributed representations for larger spans of text, e.g., phrases (Yin and Schütze, 2014) and sentences (Zhu et al., 2016; Chen et al., 2016; Zhu et al., 2015b,a; Tai et al., 2015; Kalchbrenner et al., 2014; Irsoy and Cardie, 2014; Socher et al., 2012), from the meaning representations of their parts, e.g., word embedding.

We begin by briefly introducing traditional approaches to semantic composition, including logic-based formal semantic approaches and simple arithmetic operations over vectors based on corpus word counts (Mitchell and Lapata, 2008; Landauer and Dumais, 1997).

Our main focus, however, will be on distributed representation-based modeling, whereby the representations of words and the operations composing them are jointly learned from a training objective. We cover the generic ideas behind neural network-based semantic composition and dive into the details of three typical composition architectures: the convolutional composition models (Kalchbrenner et al., 2014; Zhang et al., 2015), recurrent composition models (Zhu et al., 2016), and recursive composition models (Irsoy and Cardie, 2014; Socher et al., 2012; Zhu et al., 2015b; Tai et al., 2015). After that, we will discuss several unsupervised approaches (Le and Mikolov, 2014; Kiros et al., 2014; Bowman et al., 2016; Miao et al., 2016).

We will then advance to discuss several selected topics. We first cover the models that consider compositional with non-compositional (e.g., holistically learned) semantics (Zhu et al., 2016, 2015a). Next, we discuss composition models that integrate multiple architectures of neural networks. We also discuss semantic composition and decomposition (Turney, 2014). In the end we briefly discuss sub-word neural-network-based composition models (Zhang et al., 2015; Sennrich et al., 2016)

We will then summarize the tutorial, flesh out limitations of current approaches, and discuss future directions that are interesting to us.

2 Tutorial Outline

- Introduction
  - Definition of semantic composition
  - Conventional and basic approaches
    - Formal semantics
    - Bag of words with learned representations (additive, learned projection)
- Parametrising Composition Functions
  - Convolutional composition models
  - Recurrent composition models
  - Recursive composition models
    - TreeRNN/TreeLSTM
    - SPINN and RL-SPINN
  - Unsupervised models
    - Skip-thought vectors and paragraph vectors
    - Variational auto-encoders for text
- Selected Topics
  - Incorporating compositional and non-compositional (e.g., holistically learned) semantics
  - Integrating multiple composition architectures
  - Semantic composition and decomposition
  - Sub-word composition models
- Summary

Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics-Tutorial Abstracts, pages 6–7
Vancouver, Canada, July 30 - August 4, 2017.
https://doi.org/10.18653/v1/P17-5003

3 Instructors

Xiaodan Zhu, Queen's University; National Research Council Canada.
zhu2048@gmail.com
http://www.xiaodanzhu.com
Xiaodan Zhu is an assistant professor of the Department of Electrical and Computer Engineering of Queen's University, Canada. Before that, he was a Research Officer of the National Research Council Canada. His research interests are in Natural Language Processing and Machine Learning. His recent work has focused on deep learning, semantic composition, sentiment analysis, and natural language inference.

Edward Grefenstette, Senior Research Scientist, DeepMind.
etg@google.com
http://www.egrefen.com
Edward Grefenstette is a Staff Research Scientist at DeepMind. His research covers the intersection of Machine Learning, Computer Reasoning, and Natural Language Understanding. Recent publications cover the topics of neural computation, representation learning at the sentence level, recognising textual entailment, and machine reading.

References

Samuel R. Bowman, Jon Gauthier, Abhinav Rastogi, Raghav Gupta, Christopher D. Manning, and Christopher Potts. 2016. A fast unified model for parsing and sentence understanding. In *ACL*.

Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, and Hui Jiang. 2016. Enhancing and combining sequential and tree lstm for natural language inference. In *arXiv:1609.06038v1*.

Ozan Irsoy and Claire Cardie. 2014. Deep recursive neural networks for compositionality in language. In *NIPS*.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *ACL*.

Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. 2014. Skip-thought vectors. In *arXiv:1506.06726*.

Thomas K Landauer and Susan T. Dumais. 1997. A solution to platos problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2):211–240.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *ICML*.

Yishu Miao, Lei Yu, and Phil Blunsom. 2016. Neural variational inference for text processing. In *ICML*.

Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In *ACL*.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *ACL*.

Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *EMNLP*.

Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In *ACL*.

Peter Turney. 2014. Semantic composition and decomposition: From recognition to generation. In *arXiv:1405.7908*.

Wenpeng Yin and Hinrich Schütze. 2014. An exploration of embeddings for generalized phrases. In *ACL 2014 Student Research Workshop*.

Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. In *NIPS*.

Xiaodan Zhu, Hongyu Guo, and Parinaz Sobhani. 2015a. Neural networks for integrating compositional and non-compositional sentiment in sentiment composition. In **SEM*.

Xiaodan Zhu, Parinaz Sobhani, and Hongyu Guo. 2015b. Long short-term memory over recursive structures. In *ICML*.

Xiaodan Zhu, Parinaz Sobhani, and Hongyu Guo. 2016. Dag-structured long short-term memory for semantic compositionality. In *NAACL*.