Deep Learning for Natural Language Inference

Samuel R. Bowman
New York University
bowman@nyu.edu

Xiaodan Zhu
ECE, Queen's University
zhu2048@gmail.com

1 Description

The task of natural language inference (NLI; also known as recognizing textual entailment, or RTE) asks a system to evaluate the relationships between the truth-conditional meanings of two sentences or, in other words, decide whether one sentence follows from another. This task neatly isolates the core NLP problem of sentence understanding as a classification problem, and also offers promise as an intermediate step in the building of complex systems (Dagan et al., 2005; MacCartney, 2009; Bowman et al., 2015).

The last few years have seen fast progress in NLI, with the introduction of a few large training datasets and many popular evaluation sets as well as an explosion of new model architectures and methods for using unlabeled data and outside knowledge. This tutorial will layout the motivations for work on NLI, survey the available resources for the task, and present highlights from recent research showing us what NLI can teach us about the capabilities and limits of deep learning models for language understanding and reasoning.

The tutorial will start from a brief discussion on the motivations for NLI, problem definitions, and typical conventional approaches (Dagan et al., 2013; MacCartney, 2009; Iftene and Balahur-Dobrescu, 2007).

Critical to the recent advance on NLI, the creation of larger annotated datasets (Bowman et al., 2015; Williams et al., 2018; Conneau et al., 2018) has made it feasible to train complex models that need to estimate a large number of parameters. The tutorial will present detailed discussion on the available datasets as well as the motivations for and insights from developing these datasets. Then based on more recent research on annotation artifacts, we will extend the discussion to what we should or shouldn't take away from the current datasets.

We will then focus on the cutting-edge deep learning models for NLI. We start from two basic setups for NLI modeling: sentence-embedding-based modeling (Bowman et al., 2015; Chen et al., 2017b, 2018a; Williams et al., 2018; Yoon et al., 2018; Kiela et al., 2018; Talman et al., 2018) and deep-learning approaches that utilize cross-sentence statistics (Bowman et al., 2015; Chen et al., 2017a, 2018b; Radford et al., 2018; Devlin et al., 2018; Peters et al., 2018). We will cover typical deep-learning architectures in both paradigms.

Based on this we will deepen our discussion from several perspectives. We first describe models that can further consider linguistic structures in the deep-learning NLI architectures (Chen et al., 2017a). We then advance to discuss models that utilize external knowledge, which include two typical types of approaches: those explicitly incorporating human-authorized knowledge (Chen et al., 2018b) and those based on unsupervised pretraining (Radford et al., 2018; Devlin et al., 2018; Peters et al., 2018). We will present how NLI models are sensitive or robust to different newly proposed tests (Glockner et al., 2018; Wang et al., 2018; Naik et al., 2018; Poliak et al., 2018). The tutorial will also cover the recent modeling on cross-lingual NLI (Conneau et al., 2018).

Finally we will summarize the tutorial and flesh out some discussions on future directions.

2 Tutorial Outline

- Introduction
- Background
  - Problem definition
  - Motivations
- History and conventional methods
  - Natural logic methods
  - Theorem proving methods
- Recent advance on dataset development
  - Motivations
  - Detailed discussions/insights on dataset development and available datasets
  - Recent research on annotation artifacts representation
- Cutting-edge deep learning models

Proceedings of NAACL HLT 2019: Tutorial Abstracts, pages 6–8
Minneapolis, Minnesota, June 2, 2019.

        - Sentence-embedding-based models
        - Deep learning architectures exploring cross-sentence statistics
        - Models enhanced with linguistic structures
        - Modeling external knowledge
        - Recent advance on pretrain-based models
        - Cross-lingual NLI Models
        - Revisiting data and model limitation jointly
- Applications
    - Existing and potential downstream applications
        - MNLI for evaluation
        - MNLI for pretraining (incl. RTE)
    - NLI for evaluating sentence representation
- Summary

3 Instructors

Sam Bowman, New York University.
bowman@nyu.edu
http://www.nyu.edu/projects/bowman

Sam Bowman is an Assistant Professor of Data Science and Linguistics at New York University. Sam works on building artificial neural network models for sentence understanding, with the dual goals of both improving language technology for problems like translation and facilitating basic research on human language. He co-directs the Machine Learning for Language group (with Prof. Kyunghyun Cho) and the larger CILVR applied machine learning lab. He completed a PhD at Stanford University in 2016 with advisors Chris Manning and Chris Potts. He received a 2017 Google Faculty Research Award and led a twenty-researcher team project during the summer of 2018 as part of the JSALT workshop program.

Xiaodan Zhu, Queen's University, 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. His research interests are in natural language processing and machine learning. His recent work has focused on natural language inference, sentiment analysis, semantic composition, and summarization. He has presented tutorials before at ACL-2017 and EMNLP-2014.

References

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *EMNLP*.

Qian Chen, Zhen-Hua Ling, and Xiaodan Zhu. 2018a. Enhancing sentence embedding with generalized pooling. In *COLING*.

Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Diana Inkpen, and Si Wei. 2018b. Neural natural language inference models enhanced with external knowledge. In *ACL*.

Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017a. Enhanced lstm for natural language inference. In *ACL*.

Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017b. Recurrent neural network-based sentence encoder with gated attention for natural language inference. In *EMNLP RepEval Workshop*.

Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel R. Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. Xnli: Evaluating cross-lingual sentence representations. In *EMNLP*.

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2005. The PASCAL recognising textual entailment challenge. In *MLCW*.

Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto. 2013. *Recognizing Textual Entailment: Models and Applications*. Synthesis Lectures on Human Language Technologies. Morgan Claypool Publishers.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Max Glockner, Vered Shwartz, and Yoav Goldberg. 2018. Breaking nli systems with sentences that require simple lexical inferences. *arXiv preprint arXiv: 1805.02266* .

Adrian Iftene and Alexandra Balahur-Dobrescu. 2007. *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*.

Douwe Kiela, Changhan Wang, and Kyunghyun Cho. 2018. Dynamic meta-embeddings for improved sentence representations. *arXiv preprint arXiv: 1804.07983* .

Bill MacCartney. 2009. *Natural Language Inference*. Ph.D. thesis, Stanford University.

Aakanksha Naik, Abhilasha Ravichander, Norman Sadeh, Carolyn Rose, and Graham Neubig. 2018. Stress test evaluation for natural language inference. *arXiv preprint arXiv:1806.00692* .

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365* .

Adam Poliak, Jason Naradowsky, Aparajita Haldar, Rachel Rudinger, and Benjamin Van Durme. 2018. Hypothesis only baselines in natural language inference. *arXiv preprint arXiv: 1805.01042* .

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training. *Technical Report, OpenAI* .

Aarne Talman, Anssi Yli-Jyrä, and Jörg Tiedemann. 2018. Natural language inference with hierarchical bilstm max pooling architecture. *arXiv preprint arXiv:1808.08762* .

Haohan Wang, Da Sun, and Eric P. Xing. 2018. What if we simply swap the two text fragments? a straightforward yet effective way to test the robustness of methods to confounding signals in nature language inference tasks. *arXiv preprint arXiv: 1809.02719* .

Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *NAACL*.

Deunsol Yoon, Dongbok Lee, and SangKeun Lee. 2018. Dynamic self-attention: Computing attention over words dynamically for sentence embedding. *arXiv preprint arXiv: 1808.07383* .