Expose for student research paper

In preparation for dependency parse driven composition models for argument facet aware vector space representations of textual units

Arne Binder

August 8, 2017

1 Motivation

Large online debates contain redundant arguments and therefore hinder efficient debating. To tackle this problem Boltuzic and Šnajder (2015) proposed to cluster similar arguments and filter only the most representative ones. The Common Round[1] project (Uszkoreit et al. 2017) introduces a platform to facilitate large scale online debating with NLP technologies. The authors define the aggregation of the semantic content of debates as one of their major objectives which they intend to achieve by clustering similar arguments. Moreover, Misra, Ecker, and Walker (2016) define the Argument Facet Similarity (AFS) task as the recognition of the semantic similarity of two sentential arguments.

To achieve these goals, a semantically consistent similarity measure for arguments is necessary. Frameworks like word2vec (Mikolov et al. 2013) or glove (Pennington, Socher, and Manning 2014) lead to success in many NLP tasks by producing dense embedding vectors for single words. The cosine distance of these word embeddings enables a semantically consistent similarity measure for word tokens. Building on this, Habernal and Gurevych (2015), Boltuzic and Šnajder (2015), and Misra, Ecker, and Walker (2016) use summed or averaged word embeddings for argument similarity measures. However, Misra, Ecker, and Walker (2016) state that averaging all word embeddings may lose too much information in long sentences. Wang and Zong (2017) present an overview of different embedding composition models for phrase representations. They conclude that the recurrent long short-term memory (LSTM) model (Hochreiter and Schmidhuber 1997) just slightly outperforms the additive baseline model in this task. More recently, Tai, Socher, and Manning (2015) introduced the recursive TreeLSTM model which further reduced the vanishing gradient problem as compared to recurrent models by shortening the average distance of entities in the computation graph. The authors use dependency parse trees to construct the neural model and present promising results for the SICK phrase relatedness task[2].

However, these models are applied to sentences, but not yet to multi-sentence arguments or paragraphs. Additionally, their approach does not use dependency edge type information such as 'nsubj' (nominal subject) or 'prep' (preposition).

2 Objective

In this work we examine what degree of semantical awareness is achievable with different existing embedding composition models of varying complexity when applied to sentences.

[1] see http://commonround.dfki.de/

[2] The SICK corpus (Marelli et al. 2014) contains ∼10.000 similarity scored sentence pairs. The system by Tai, Socher, and Manning (2015) achieved a Pearson's r of 0.8676 when predicting the similarities.

We regard compositions of token embeddings as semantically aware if they enable a distance measure that matches human intuition. Furthermore, our goal is to provide access to the rapid implementation of different neural network topologies including recursive neural networks (RecNN) for semantically aware composition.

3 Approach

To achieve these goals we implement the following: (a) an averaging model, and (b) a neural sequence model. We evaluate the semantical awareness with the SICK Semantic Textual Similarity (STS) challenge (Marelli et al. 2014) and compare the model performances against a TF-IDF baseline.

To prepare for rapid comparison with various neural models, we use the Tensorflow Fold[3] framework (Looks et al. 2017) for implementation. It allows fast training of neural networks with dynamic computation graphs, i.e. networks whose structure depends on the data that is fed during training. Moreover, it is based on Tensorflow[4] which is widely used as a deep learning framework.

3.1 Model details

As neural sequence model, we use a model based on Mueller and Thyagarajan (2016) due to its simplicity. It consists of one LSTM as composition model which is applied to two input sentences and uses the Manhattan metric as distance measure. Nevertheless, the authors report a Pearson's r of 0.8822 for the SICK challenge. Furthermore, we analyze the effect of dependency type information by optionally appending edge type embeddings to the token embedding vectors.

3.2 Dataset enlargement

One major drawback of the SICK corpus is its small size. We enable pre-training with paraphrase data from the PPDB[5] corpus (Ganitkevitch, Van Durme, and Callison-Burch 2013) and evaluate its effect to the model performances. In its smallest, most accurate version the phrasal subset of the corpus contains 1,530,812 pairs of short phrase snippets like ('`maintain international peace and`', '`maintaining world peace and`'). We treat these pairs as similar (Likert score = 4.0) and extend the dataset with artificially generated negative samples (Likert score = 1.0) where the two phrases are randomly sampled from the corpus.

References

Boltuzic, Filip and Jan Šnajder (2015). "Identifying prominent arguments in online debates using semantic textual similarity". In: *Proceedings of the 2nd Workshop on Argumentation Mining*, pp. 110–115. URL: http://www.aclweb.org/anthology/W/W15/W15-05.pdf#page=122.

Ganitkevitch, Juri, Benjamin Van Durme, and Chris Callison-Burch (2013). "PPDB: The Paraphrase Database." In: *HLT-NAACL*, pp. 758–764. URL: http://www.aclweb.org/anthology/N13-1#page=796.

Habernal, Ivan and Iryna Gurevych (2015). "Exploiting Debate Portals for Semi-Supervised Argumentation Mining in User-Generated Web Discourse." In: *EMNLP*, pp. 2127–2137. URL: http://www.aclweb.org/anthology/D/D15/D15-1255.pdf.

[3] see https://github.com/tensorflow/fold

[4] see www.tensorflow.org

[5] see http://paraphrase.org

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural computation* 9 8, pp. 1735–80.

Looks, Moshe et al. (2017). "Deep Learning with Dynamic Computation Graphs". In: *arXiv:1702.02181 [cs, stat]*. arXiv: 1702.02181. URL: http://arxiv.org/abs/1702.02181.

Marelli, Marco et al. (2014). "A SICK cure for the evaluation of compositional distributional semantic models." In: *LREC*, pp. 216–223. URL: http://clic.cimec.unitn.it/marco/publications/marelli-etal-sick-lrec2014.pdf.

Mikolov, Tomas et al. (2013). "Distributed representations of words and phrases and their compositionality". In: *Advances in neural information processing systems*, pp. 3111–3119. URL: http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.

Misra, Amita, Brian Ecker, and Marilyn A. Walker (2016). "Measuring the Similarity of Sentential Arguments in Dialog". In: *17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, p. 276. URL: http://www.aclweb.org/anthology/W/W16/W16-36.pdf#page=294.

Mueller, Jonas and Aditya Thyagarajan (2016). "Siamese Recurrent Architectures for Learning Sentence Similarity." In: *AAAI*, pp. 2786–2792. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/download/12195/12023.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning (2014). "Glove: Global Vectors for Word Representation". In: *EMNLP*.

Tai, Kai Sheng, Richard Socher, and Christopher D. Manning (2015). "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks". In: *arXiv:1503.00075 [cs]*. arXiv: 1503.00075. URL: http://arxiv.org/abs/1503.00075.

Uszkoreit, Hans et al. (2017). "Common Round: Application of Language Technologies to Large-Scale Web Debates". In: *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, Spain: Association for Computational Linguistics, pp. 5–8. URL: http://aclweb.org/anthology/E17-3002.

Wang, Shaonan and Chengqing Zong (2017). "Comparison Study on Critical Components in Composition Model for Phrase Representation". en. In: *ACM Transactions on Asian and Low-Resource Language Information Processing* 16.3, pp. 1–25. ISSN: 23754699. DOI: 10.1145/3010088. URL: http://dl.acm.org/citation.cfm?doid=3041821.3010088.