Neural Architectures for Modeling Compositionality in Natural Language

John Wieting
Toyota Technological Institute at Chicago

Mohit Bansal
UNC Chapel Hill

Kevin Gimpel
Toyota Technological Institute at Chicago

Karen Livescu
Toyota Technological Institute at Chicago

Abstract

We relate our experience in designing and experimenting with neural network architectures for modeling compositionality in natural language. The goal is to represent a natural language text such as a word or sentence in a fixed-length embedding that is useful for downstream applications. We use a broad range of neural architectures, including those based on bag-of-word models (Wieting et al. 2016b), bag-of-character-n-gram models (Wieting et al. 2016a), recurrent neural networks on words (Wieting et al. 2016b) or characters (Wieting et al. 2016a), convolutional neural networks using character n-gram filters (Wieting et al. 2016a), and hybrids that combine the strengths of bag-of-words and recurrent architectures (Wieting and Gimpel 2017). We consider several tasks, including word similarity, sentence similarity, sentiment analysis, part-of-speech tagging, and textual entailment. We find that the best compositional architecture varies depending on the task. Table 1 compares several word-based architectures for three tasks. While simple bag-of-words models perform strongly on semantic similarity and entailment tasks, recurrent networks perform best on sentiment analysis:

	sent. sim.	entailment	sentiment
word averaging	**86.4**	**84.6**	83.0
DAN(Iyyer et al. 2015)	86.0	84.5	83.4
RNN	73.1	76.4	86.5
LSTM (no output gate)	85.5	83.2	86.6
LSTM (with output gate)	83.4	82.0	**89.2**

Table 1: Results with word-based neural network architectures: Pearson's $r \times 100$ for sentence similarity using the SICK dataset (Marelli et al. 2014), accuracy for entailment again using the SICK dataset, and accuracy for binary sentiment classification using the Stanford Sentiment Treebank (Socher et al. 2013).

Table 2 compares several subword-based architectures. A simple architecture that sums vectors for character n-grams (CHARAGRAM) outperforms models that use recurrent or convolutional networks to compose character sequences into word or sentence representations (Wieting et al. 2016a).

	word sim.	sent. sim.	POS tagging
charCNN	16.7	59.2	97.0
charLSTM	48.8	41.9	96.9
CHARAGRAM	**60.0**	**68.7**	**97.1**

Table 2: Results with subword-based neural network architectures: Spearman's $\rho \times 100$ for word similarity using the SimLex-999 dataset (Hill, Reichart, and Korhonen 2015), Pearson's $r \times 100$ using 24 semantic textual similarity (STS) datasets, and accuracy for part-of-speech tagging using the standard WSJ portion of the Penn Treebank.

We also consider a transfer learning setting, finding that simpler models show more stability when crossing into new domains (Wieting et al. 2016b). We develop a novel architectural variation—the gated recurrent averaging network—that combines the strengths of averaging and recurrent models, improving results in both supervised and transfer settings (Wieting and Gimpel 2017).

References

Hill, F.; Reichart, R.; and Korhonen, A. 2015. SimLex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics* 41(4).

Iyyer, M.; Manjunatha, V.; Boyd-Graber, J.; and Daumé III, H. 2015. Deep unordered composition rivals syntactic methods for text classification. In *Proceedings of ACL*.

Marelli, M.; Menini, S.; Baroni, M.; Bentivogli, L.; bernardi, R.; and Zamparelli, R. 2014. A sick cure for the evaluation of compositional distributional semantic models. In *Proceedings of LREC*.

Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C. D.; Ng, A.; and Potts, C. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of EMNLP*.

Wieting, J., and Gimpel, K. 2017. Revisiting recurrent networks for paraphrastic sentence embeddings. In *Proceedings of ACL*.

Wieting, J.; Bansal, M.; Gimpel, K.; and Livescu, K. 2016a. Charagram: Embedding words and sentences via character n-grams. In *Proceedings of EMNLP*.

Wieting, J.; Bansal, M.; Gimpel, K.; and Livescu, K. 2016b. Towards universal paraphrastic sentence embeddings. In *Proceedings of ICLR*.