Cross-Lingual Word Vectors for Deep Sentiment Analysis

Xin Dong, Gerard de Melo
Rutgers University
xd48@rutgers.edu, gdm@demelo.org

There is a long history of inducing vector-based representations of linguistic utterances. In our work, we induce *sentiment embeddings* custom-tailored for the task of sentiment analysis. Deep neural networks normally require large annotated training corpora for each combination of language, domain, and genre. We conjecture that encoding the prior sentiment polarity of words in different domains into their word vectors may mitigate this challenge.

Our approach consists of three main ingredients. First, we induce monolingual sentiment embeddings. Given *n* binary polarity classification tasks for different domains in a resource-rich language such as English, we learn *n* corresponding linear binary classification models using bag-of-words features. Then, each word present in the vocabulary is assigned a new word vector consisting of the series of linear coefficients for the respective word across the *n* linear models.

Next, we cross-lingually extend these via graph-based propagation (de Melo 2015, de Melo 2017). Given the initial sentiment embedding vectors, we make use of a lexical knowledge graph containing multilingual words and weighted directed arcs between them. Our objective function ensures that sentiment embeddings of words accord with those of their connected words, in terms of the dot product. At the same time, a second component of the objective function ensures that the deviation from the initial word vectors computed earlier does not grow too large for words that have such vectors. For optimization, we rely on stochastic gradient descent steps.

Finally, we propose to rely on a modified convolutional neural network architecture (Dong and de Melo 2018), in which the input utterance is both fed through regular convolutional filters, while as well being processed in a second network module that makes use of our sentiment-specific word vectors.

Our experiments suggest that such a dual-module approach is best suited to exploit the automatically induced multilingual sentiment vectors. The overall approach leads to gains on datasets for seven different languages from different domains, and is most useful when training data is scarce.

References: • de Melo, Gerard (2015): Wiktionary-based Word Embeddings. *Proceedings of MT Summit XV. Vol 1: MT Researchers' Track*. Washington, DC, USA: AMTA. • de Melo, Gerard (2017): Inducing Conceptual Embedding Spaces from Wikipedia. *Proceedings of WWW 2017*. ACM • Dong, Xin & de Melo, Gerard (2018): A Helping Hand: Transfer Learning for Deep Sentiment Analysis. *Proceedings of ACL 2018*. Stroudsburg, PA, USA: ACL.