Deep Learning of Hierarchical Structure

Chris Manning
Stanford University

Hierarchical and recursive structure is commonly found in inputs from the richest sensory modalities, including natural language sentences and scene images. But such hierarchical structure has traditionally been a strong point of both structured and supervised models (whether symbolic of probabilistic) and a weak point of both neural networks and unsupervised learning.

I will present some of our recent work on a recursive neural network architecture that tries to address these issues. The system was initially developed to jointly parse natural language sentences and to learn syntactico-semantic vector representations of phrases. The system is based on a context-aware recursive neural network (RNN), which can both find a globally optimal parse tree with dynamic programming and induce distributed phrase representations for unseen, variable-sized inputs. The induced representations capture interesting semantic similarities and can be used successfully as features for finding syntactically and semantically plausible paraphrases. For instance, the phrases ”declined to comment” and ”would not disclose” are close by in the induced embedding space. However, the same approach can also be applied to find hierarchical structure in images. Discovering this hierarchical structure helps us to not just identify the units that an image contains but also how they interact to form a whole. We introduce a max-margin structure prediction architecture based on a recursive neural network that can successfully recover such structure from complex scene images. For semantic scene segmentation, annotation and classification, this algorithm obtains a new level of state-of-the-art performance for segmentation on the Stanford background dataset (78.1%), outperforming Gist descriptors for scene classification by 4%.

The recursive neural network parser is trained on supervised data. I finally discuss extending this work in a less supervised direction by exploring using recursive autoencoders for predicting sentence-level sentiment distributions. The model captures compositional semantics and learns phrase feature representations sufficiently well as to outperform more traditional supervised sentiment classification methods operating on flat bags of features, even without the use of any predefined sentiment lexica or polarity shifting rules.

This talk presents joint work with Richard Socher and Andrew Ng.