



Learning Vector Representations for Sentences

P. Lê

Learning Vector Representations for Sentences – The Recursive Deep Learning Approach

Natural language processing (NLP) systems, until recently, relied heavily on sophisticated representations and carefully designed feature sets. Now with the rise of deep learning, for the first time in the history of NLP, the importance of such manual feature engineering has started to be challenged. Deep learning systems using very few handcrafted features can achieve state-of-the-art (or nearly state-of-the-art) performance on many tasks, such as syntactic parsing, machine translation, sentiment analysis, and language modelling. However, rather than letting deep learning replace linguistically informed approaches, in this dissertation I explore how linguistic knowledge can provide insights for building even better neural network models. I tackle the problem of transforming sentences into vectors by employing a hybrid approach of symbolic NLP and connectionist deep learning based on the principle of compositionality. In this approach, the role of symbolic NLP is to provide syntactic structures whereas composition functions are implemented (and trained) by connectionist deep learning.

All of the models I develop in this dissertation are variants of the Recursive neural network (RNN). The RNN takes a sentence, syntactic tree, and vector representations for the words in the sentence as input, and applies a neural network to recursively compute vector representations for all the phrases in the tree and the complete sentence. The RNN is a popular model because of its elegant definition and promising empirical results. However, it also has some serious limitations: (i) the composition functions it can learn are linguistically impoverished, (ii) it can only be used in a bottom-up fashion, and (iii) it is extremely sensitive to errors in the syntactic trees it is presented with. Starting with the classic RNN, I propose extensions along three different directions that solve each of these problems.

The first direction focuses on strengthening the composition functions. One way to do that is making use of syntactic information and contexts, as in Chapter 3. In that

chapter, I propose composition functions, which are also one-layer feed-forward neural networks, taking into account representations of syntactic labels (e.g. N, VP), context words, and head words. Another way is to replace one-layer neural networks by more advanced networks. In Chapter 6, based on empirical results which show that the Long short term memory (LSTM) architecture can capture long range dependencies and deal with the vanishing gradient problem more effectively than Recurrent neural networks, I introduce a novel variant of the LSTM, called Recursive-LSTM, that works on trees. Empirical results on an artificial task and on the Stanford Sentiment Treebank confirm that the proposed Recursive-LSTM model is superior to the classic RNN model in terms of accuracy. Furthermore, in Chapter 7, I demonstrate how a convolutional neural network can be used as a composition function.

The second direction to extend the classic RNN is to focus on how information flows in a parse tree. In traditional compositional semantics approaches, including the RNN model, information flows in a bottom-up manner, leading to a situation where there is no way for a node to be aware of its surrounding context. As a result, these approaches are not applicable to top-down processes such as several top-down generative parsing models, and to problems requiring contexts such as semantic role labelling. In Chapter 4, I propose a solution to this, namely the Inside-Outside Semantic framework, in which the key idea is to allow information to flow not only bottom-up but also top-down. In this way, we can recursively compute representations for the content and the context of the phrase that a node in a parse tree covers. The Inside-Outside RNN model, a neural-net-based instance of this framework, is shown to work well on several tasks, including unsupervised composition function learning from raw texts, supervised semantic role labelling, and dependency parsing (Chapter 5).

The third direction is dealing with the uncertainty of the correct parse. As a result of relying on the principle of compositionality, compositional semantics uses syntactic parse trees to guide composition, which in turn makes compositional semantics approaches vulnerable to the errors of automatic parsers. The problems here are that automatic parsers are not flawless, and that they are not aware of domains to which they are applied. To overcome this problem, in Chapter 7, I propose the Forest Convolutional Network model, which takes as input a forest of parse trees rather than a single tree as in traditional approaches. The key idea is that we should give the model several options and let it select (or combine) ones that best fit its need. Empirical results show that the model performs on par with state-of-the-art models on the Stanford Sentiment Treebank and on the TREC question dataset.

The dissertation thus proposes solutions to the main shortcomings of the RNN model. It provides all components for a completely neural implementation of a syntactic-semantic parser: the three ideas above essentially yield a neural inside-outside algorithm. This represents an approach to NLP that combines the best of two worlds: all the flexibility and learning power of deep learning without sacrificing the linguistic adequacy of earlier approaches in computational linguistics.