Natural Language Processing
Recursive Networks and Paragraph Vectors

Felipe Bravo-Marquez

November 20, 2018

Recursive Neural Networks over Sentiment Treebank

- A recursive neural tensor network for learning the sentiment of pieces of texts of different granularities, such as words, phrases, and sentences, was proposed in [Socher et al., 2013].
- The network was trained on a sentiment annotated treebank `http://nlp.stanford.edu/sentiment/treebank.html` of parsed sentences for learning compositional vectors of words and phrases.
- Every node in the parse tree receives a vector, and there is a matrix capturing how the meaning of adjacent nodes changes.
- The network is trained using a variation of backpropagation called Backprop through Structure.
- The main drawback of this model is that it relies on parsing.

Recursive Neural Networks over Sentiment Treebank

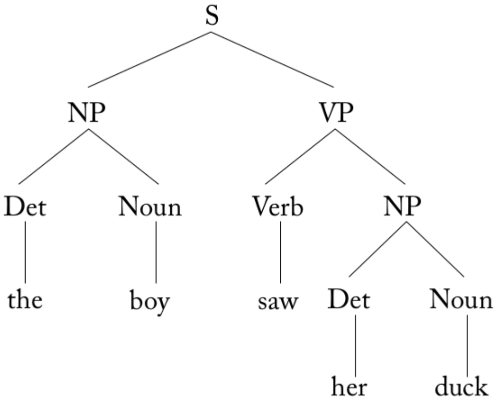

Figure: A parse tree

Recursive Neural Networks over Sentiment Treebank

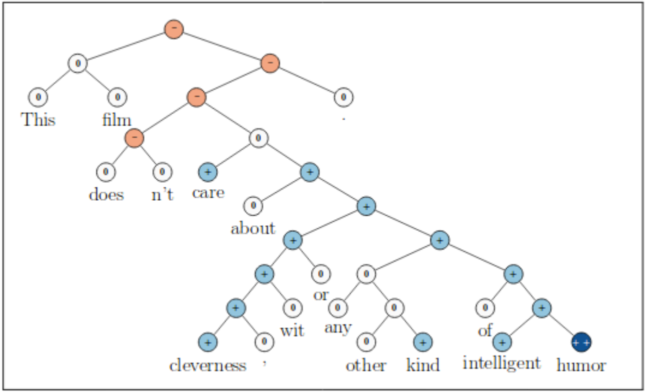

Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

Recursive Neural Networks over Sentiment Treebank

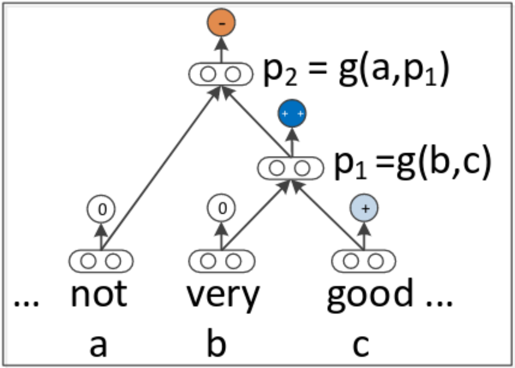

Figure 4: Approach of Recursive Neural Network models for sentiment: Compute parent vectors in a bottom up fashion using a compositionality function g and use node vectors as features for a classifier at that node. This function varies for the different models.

Paragraph vector

- A paragraph vector-embedding model that learns vectors for sequences of words of arbitrary length (e.g, sentences, paragraphs, or documents) without relying on parsing was proposed in [Le and Mikolov, 2014].
- The paragraph vectors are obtained by training a similar network as the one used for training the CBOW embeddings.
- The words surrounding a centre word in a window are used as input together with a paragraph-level vector for predict the centre word.
- The paragraph-vector acts as a memory token that is used for all the centre words in the paragraph during the training the phase.
- The recursive neural tensor network and the paragraph-vector embedding were evaluated on the same movie review dataset used in [Pang et al., 2002], obtaining an accuracy of 85.4% and 87.8%, respectively.
- Both models outperformed the results obtained by classifiers trained on representations based on bag-of-words features.
- Many researchers have have struggled to reproduce these paragraph vectors [Lau and Baldwin, 2016].

Paragraph vector

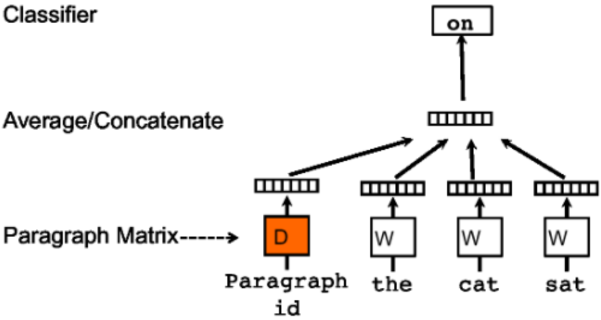

Figure 2. A framework for learning paragraph vector. This framework is similar to the framework presented in Figure 1; the only change is the additional paragraph token that is mapped to a vector via matrix D. In this model, the concatenation or average of this vector with a context of three words is used to predict the fourth word. The paragraph vector represents the missing information from the current context and can act as a memory of the topic of the paragraph.

Summary

- Neural networks are making improvements across many NLP tasks (e.g., sentiment analysis, machine translation).
- Deep Learning ! $=$ Feature Engineering.
- Word embeddings provide a practical framework for semi-supervised learning (i.e., leveraging unlabeled data).
- Character-level embeddings are worth paying attention to!
- Convolutional neural networks can capture useful features (e.g., n-grams) regardless of the position.
- Recurrent Neural Networks are very useful for learning temporal patterns, especially for long dependencies.
- We just scratched the surface!!

Questions?

Thanks for your Attention!

References I

Lau, J. H. and Baldwin, T. (2016).
An empirical evaluation of doc2vec with practical insights into document embedding generation.
arXiv preprint arXiv:1607.05368.

Le, Q. V. and Mikolov, T. (2014).
Distributed representations of sentences and documents.
In *Proceedings of the 31th International Conference on Machine Learning*, pages 1188–1196.

Pang, B., Lee, L., and Vaithyanathan, S. (2002).
Thumbs up? Sentiment classification using machine learning techniques.
In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pages 79–86. Association for Computational Linguistics.

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., and Potts, C. (2013).
Recursive deep models for semantic compositionality over a sentiment treebank.
In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642. Association for Computational Linguistics.