Neural network based method to compute fine-grained sentiment in movie reviews

Akella Ravi Tej

1 Abstract

This is a paper review of the scientific article "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank" by Socher et al., 2013. The paper introduces a novel approach for fine grained sentiment classification of movie reviews using neural networks and a dataset that helps this model achieve state of the art performance over various metrics.

2 Introduction

Semantic Compositionality principle states that the meaning of any complex expression can be expressed as a function of the meanings of its syntactic components together with their syntactic mode of combination. Humans are able to easily understand sentences that they have never come across if they are aware of the comprising words, rules of language and grammar. This ability to reason over language may seem straightforward but is extremely impractical to manually lay down the rules of a language to build computer models. For this reason, it is best we train a neural network to figure out a model to represent the semantic compositionality of the language as a function of simpler semantic components. The biggest obstacles to this approach are the need for a large labelled dataset to accurately capture the details in semantics and a composition model that can encompass the semantic compositionality of the language.

This paper introduces *Sentiment Treebank* dataset, which offers richer supervised training and evaluation resources. *Stanford Parser* (Klein and Manning, 2003) and *Amazon Mechanical Turk* were used to parse and label (5 classes) the dataset respectively.

3 Summary

Sentiment Treebank reveals many insights about sentiment content in n-grams. Longer sentences seem to build stronger sentiment value whereas most short phrases are neutral. The new dataset offers rich supervised training for sentiment compositionality of phrases giving a significant boost to the accuracies of several known models.

Many methods have been developed to capture compositionality but recursive methods have clearly outperformed their competition. In a Recursive Neural Network architecture, an n-gram sentence is first parsed into a binary tree and the vector representation of parent node is calculated recursively in a bottom up fashion using the composition of child nodes. Further, sentiment of each vector is derived from itself using a 5 class softmax function. The models are trained using backpropagation through gradient descent in supervised learning settings. The objective function is to minimize the cross entropy error at the softmax classifier at each step of recursion. It is observed that without the non linearity, the recursive model's performance deteriorates significantly.

RNN: Recursive Neural Network: The composition function is a *tanh* nonlinearity over homogeneous additive operation of the child semantic vectors.

MV-RNN: Matrix-Vector RNN: In this method every word and phrase is represented by a vector and a matrix. The composition is a non-linear function of sum of cross products of the vector and matrix of the children.

RNTN:Recursive Neural Tensor Network: The proposed method has a tensor-based composition function in which each slice computes a specific composition of the semantic vectors. This composition function has a fixed number of parameters and performs slightly better than MV-RNN for fine grained sentiment classification.

4 Review

In simple RNN, even though the non-linearity allows to express a wider range of functions, it is almost certainly too much to expect of a single linear operation to be able to capture the meaning combination effects of all natural language operators. MV-RNN boasts a powerful composition function that learns to identify operators (such as "extremely") and thus magnify the meaning of the phrase in both positive and negative directions accordingly. The biggest disadvantage to this method is the additional storage cost of matrix along with the semantic vectors. Though the original paper suggests a low rank matrix factorization to alleviate the storage costs, the overhead is still significantly larger than that of a simple RNN model. RNTN method is designed to tackle the high number of parameters in MV-RNN, without losing the ability to capture compositionality of semantics. The base assumption behind the working of this model is, including a multiplicative interactions over semantic vector space will make the model more powerful and ensures greater interaction between semantic vectors.

Though the composition function of RNTN is stronger than previously defined RNN's, it still doesn't encompass all aspects of semantics, as it fails to do significantly better than bag of words for longer sentences. Deeper Networks have shown to outperform multiplicative and shallow networks because of their high representational power. The paper also mentions using a 2 layer RNN could be more powerful than a RNTN model but dismisses the possibility as it was hard to achieve convergence with such models. I believe that building a deeper RNN model and a stronger composition function are two promising directions to take this research further.

5 Conclusion

RNTN method proposed in the paper, provides state of the art classifiers for binary and fine grained sentiment classification when trained on the Sentiment Treebank. It also outperforms other models at different metrics (Contrastive Conjunction, High Level Negation) and is by far the only model that accurately captures negation and its scope in a sentence.