Ritika Chakradhar ritika@ritikachakradhar.com

Review of the paper

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank
by
Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts

This paper highlights the importance and the use of compositionality in sentiment detection and presents a new dataset called Stanford Sentiment Treebank for sentiment detection on different levels of granularity, e.g., words, phrases and sentences. They, then, go on to introduce a novel tensor based recursive neural network to obtain robust compositional embeddings for sentiment analysis.

Major ideas discussed as well as contributions of the paper are :

- Stanford Sentiment Treebank
  - To tackle the lack of corpora that could aid in effectively making use of compositionality effects for sentiment detection, the authors introduce Stanford Sentiment Treebank.
  - It contains parse tree of sentences where each node is assigned a sentiment ranging from very positive to very negative. This also helps in assigning a sentiment to every node in the parse tree. This allows to learn how the semantics in general and sentiment in particular, percolate in a bottoms-up fashion.
  - While analyzing the annotated treebank, authors find that stronger sentiments come out more powerfully in longer n-grams showing that only a few words(unigrams) play an important role in assigning the sentiment to their parent n-grams.

- Recursive Neural Tensor Network(**RNTN**)
  - Authors introduce augment Recursive Neural Networks(RNNs) by adding a tensor $V^{[1:d]}$ to the weight matrix parameter W. These RNNs can be used to learn n-gram representations in a bottoms-up manner in the phrase-tree.
  - Given the representation of the children nodes with representations a and b, the new network now calculates the representation of the parent node, p as $f(c^T V^{[1:d]}c + Wc)$ instead of $f(Wc)$ where c is the concatenation of vectors a and b and f is a typical non-linear function ($tanh$ in this case).
  - Sentiment of each node is then obtained by using a softmax classifier. For a node represented by the vector a, it's sentiment probability distribution is given by the probability vector $y = softmax(W_{softmax}a)$.
  - The network is trained using backpropagation on a KL-divergence error function.

- Experimental observations
    - Non-linearity helps in improving the results of recursive models.
    - Training on the Stanford Sentiment Treebank boosts the performance of already existing methods by a significant margin.
    - RNTNs outperform other models for sentiment classification both at the sentence level as well as on the n-gram level.
    - On two specialized tasks, one involving ‘X but Y’ type of sentences and the other involving sentences with negations, superiority of RNTN is reported.

Advantages of the ideas discussed in this paper are -

- The paper does a thorough literature review in the related work section.
- The sentiment treebank provides a novel dataset which claims to capture compositionality better and this claim is backed by the experiments.
- RNTN outperform both vanilla RNNs and (Matrix Vector RNN)MV-RNN model while making sure that the storage overhead in case of MV-RNNs does not blow up as well as taking interactions between the vectors representing the children node into account to obtain the representation of the n-gram at the parent node (unlike the RNNs) in order to obtain an n-gram representation composed in a better manner.
- Experimental results and Demonstration via different examples on the two specialized tasks involving phrases joined by ‘but’ and negations also give a crucial insight into the working of the RNTN and how it captures compositionality better than other models. The graph showing the change in activations for negations also substantiates the claim that RNTNs are good at capturing negations and interpreting different modifiers like “but”.
- Recursive models especially RNTNs pose a strong alternative to bag of word models as well as sequential models like Recurrent Neural Nets and LSTMs(not discussed in this paper) and are relatively unexplored.

Disadvantages of ideas discussed in this paper are -

- Apart from Bag-of-word models as well as average of word vectors, the only models that are considered are Recursive Neural Networks. A better comparison would include other models as well.
- Some more light could be shed on how does the model perform on movie reviews with complex sentences like sarcasm.
- RNTN proposed in the paper rely on parse tree information. The parse trees might not be correct in case of ambiguous sentences as well as on sentences like tweets where grammatical rules are not strictly adhered to. In those cases, the model might not be able to capture compositional features well.