Aspect Level Sentiment Classification with Deep Memory Network

Abstract

We introduce a deep memory network for aspect level sentiment classification. Experiments on laptop and restaurant datasets demonstrate that our approach performs comparable to state-of-art feature based SVM system, and substantially better than LSTM and attention-based LSTM architectures.

1 Introduction

Aspect level sentiment classification is a fundamental task in the field of sentiment analysis. Given a sentence and an aspect occurring in the sentence, this task aims at inferring the sentiment polarity (e.g. positive, negative, neutral) of the aspect. For example, in sentence “great food but the service was dreadful!”, the sentiment polarity of aspect “food” is positive while the polarity of aspect “service” is negative.

Our approach is data-driven, computationally efficient and does not rely on syntactic parser or sentiment lexicon. As every component is differentiable, the entire model could be efficiently trained end-to-end with gradient descent, where the loss function is the cross-entropy error of sentiment classification.

2 Background: Memory Network

Memory network is a general machine learning framework introduced by Weston et al. (2014). Its central idea is inference with a long-term memory component, which could be read, written to, and jointly learned with the goal of using it for prediction.

3 Deep Memory Network for Aspect Level Sentiment Classification

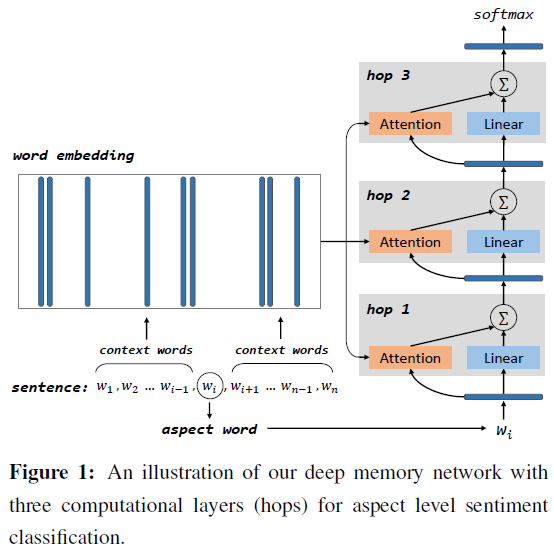
3.1 Task Definition and Notation

Given a sentence s = {w1, w2, … , wi , … wn} consisting of n words and an aspect word wi occurring in sentence s, aspect level sentiment classification aims at determining the sentiment polarity of sentence s towards the aspect wi.

3.2 An Overview of the Approach

Our approach consists of multiple computational layers (hops), each of which contains an attention layer and a linear layer.

It is helpful to note that the parameters of attention and linear layers are shared in different hops. Therefore, the model with one layer and the model with nine layers have the same number of parameters.



3.3 Content Attention

We believe that such an attention model has two advantages. One advantage is that this model could adaptively assign an importance score to each piece of memory mi according to its semantic relatedness with the aspect. Another advantage is that this attention model is differentiable, so that it could be easily trained together with other components in an end-to-end fashion.

3.4 Location Attention

We have described our neural attention framework and a content-based model in previous subsection. However, the model mentioned above ignores the location information between context word and aspect. Such location information is helpful for an attention model because intuitively a context word closer to the aspect should be more important than a farther one. In this work, we define the location of a context word as its absolute distance with the aspect in the original sentence sequence.

3.5 The Need for Multiple Hops

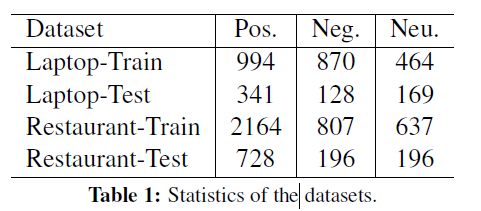
With the composition of enough such transformations, very complex functions of sentence representation towards an aspect can be learned.

3.6 Aspect Level Sentiment Classification

The model is trained in a supervised manner by minimizing the cross entropy error of sentiment classification, whose loss function is given below, where T means all training instances, C is the collection of sentiment categories, (s; a) means a sentence-aspect pair.

4 Experiment

4.1 Experimental Setting



4.2 Comparison to Other Methods

We compare with the following baseline methods on both datasets.

1. Majority
2. Feature-based SVM
3. LSTM, TDLSTM, TDLSTM+ATT
4. ContextAVG

4.3 Runtime Analysis

4.4 Effects of Location Attention

4.5 Visualize Attention Models

We visualize the attention weight of each context word to get a better understanding of the deep memory network approach.

4.6 Error Analysis

The first factor is non-compositional sentiment expression. This model regards single context word as the basic computational unit and cannot handle this situation.

The second factor is complex aspect expression consisting of many words, such as “ask for the round corner table next to the large window.” This model represents an aspect expression by averaging its constituting word vectors, which could not well handle this situation.

The third factor is sentimental relation between context words such as negation, comparison and condition.

5 Related Work

This work is connected to three research areas in natural language processing. We briefly describe related studies in each area.

5.1 Aspect Level Sentiment Classification

The aspect word in this work is given as a part of the input.

5.2 Compositionality in Vector Space

5.3 Attention and Memory Networks

6 Conclusion