Day 105: NLP Papers Summary – Aspect Level Sentiment Classification With Attention-Over-Attention Neural Networks

By Ryan 14th April 2020 No Comments

Objective and Contribution

Introduce an attention-over-attention (AOA) neural network for aspect-based sentiment analysis. The AOA module jointly learns the representations for aspects and sentences and explicitly captures the interaction between aspects and context sentences. The results on laptop and restaurant datasets outperforms previous LSTM-based architectures.

Datasets

Experimented with two domain-specific datasets from SemEval 2014 Task 4: laptop and restaurant. Accuracy is the evaluation metric. Dataset summary is shown in the figure below:

Dataset	Positive	Neutral	Negative
Laptop-Train	994	464	870
Laptop-Test	341	169	128
Restaurant-Train	2164	637	807
Restaurant-Test	728	196	196

Table 1. Distribution by sentiment polarity category of the datasets from SemEval 2014 Task 4. Numbers in table represent numbers of sentence-aspect pairs.

Methodology

In this task, we are given a sentence and an aspect target and our goal is to classify the sentiment polarity of the aspect target in the sentence. There are 4 main components in the architecture shown below: word embedding, bi-LSTM, attention-over-attention (AOA), and final prediction.

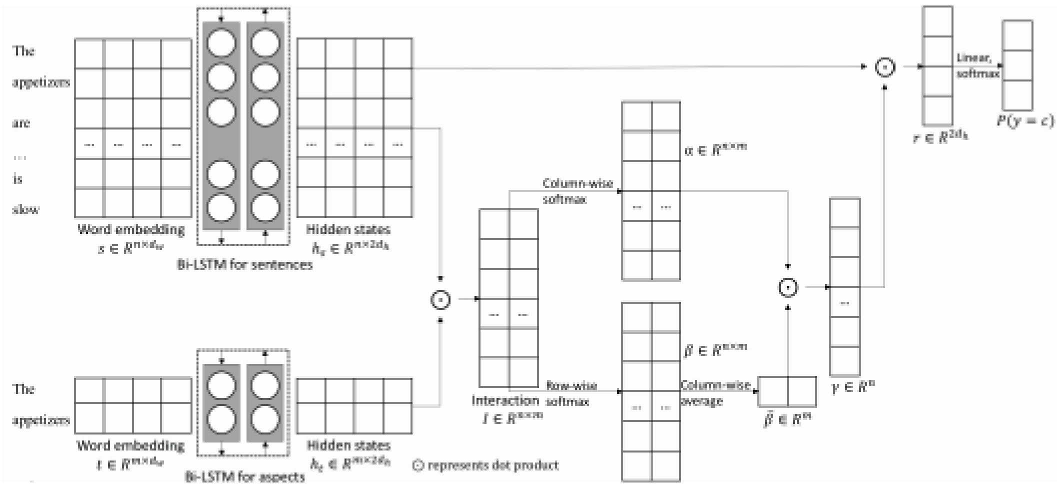

Fig. 1. The overall architecture of our aspect level sentiment classification model.

WORD EMBEDDING AND BI-LSTM

Word embedding is a standardised step where we convert the text sentence and aspect target into its numerical representations. Nothing special here. Once we get the word vectors, we feed them into two bi-LSTM respectively, to learn the hidden semantics of words in the sentence and aspect target.

ATTENTION-OVER-ATTENTION (AOA)

The next step is to calculate the attention weights for the text using the AOA module. Here he following steps:

1. Calculate a pair-wise interaction matrix between the two hidden states, where the value of each entry represents the correlation of a word pair between sentence and target
2. Perform column-wise softmax to get α, target-to-sentence attention
3. Perform row-wise softmax to get β, sentence-to-target attention
4. Calculate the column-wise average of β to get a target-level attention $\bar{\beta}$, which tells us the important parts in an aspect target
5. The final sentence-level attention γ is the weighted sum of each individual target-to-sentence attention α as follows: $\gamma = \alpha\bar{\beta}^T$.

FINAL PREDICTION

The final sentence representation is a weighted sum of sentence hidden states using sentence attention from AOA module, as follows: $r = h_s^T\gamma$. This final sentence representation is feed into a linear layer with a softmax function to output probabilities of sentiment classes. The sentiment class with the highest probability is the predicted label for the sentence, given the aspect target.

Experiments and Results

MODEL COMPARISONS

- Majority
  - Simple baseline that assigns the largest sentiment polarity in the training set to each sample in the test set
- LSTM
- TD-LSTM
  - Uses two LSTM to model the preceding and following contexts surrounding the aspect term
- AT-LSTM
  - Models the sentence via LSTM and combine the output hidden states with the aspect term embedding to generate attention vector
- ATAE-LSTM
  - Extends AT-LSTM by appending the aspect embedding into each of the word vector

- IAN
  - Uses two LSTM to model the sentence and aspect term respectively. It uses the hidden states from the sentence to generate an attention vector for the target and vice versa

RESULTS

Methods	Restaurant	Laptop
Majority	0.535	0.650
LSTM	0.743	0.665
TD-LSTM [23]	0.756	0.681
AT-LSTM [28]	0.762	0.689
ATAE-LSTM [28]	0.772	0.687
IAN [11]	0.786	0.721
AOA-LSTM	**0.812** (0.797±0.008)	**0.745** (0.726±0.008)

Table 2. Comparison results. For our method, we run it 10 times and show "best (mean±std)". Performance of baselines are cited from their original papers.

- AOA-LSTM performed the best in comparisons to other baseline methods according to the result table
- We also included the table that showcase which word contributes the most to the aspect sentiment polarity by visualising the sentence attention vectors γ.

Conclusion and Future Work

In the error analysis, there are cases that the model can't handle efficiently. One is complex sentiment expression. Another is uncommon idioms. In future work, we could incorporate sentences' grammar structures or feed prior language knowledge to the AOA neural network.

Source: https://arxiv.org/pdf/1804.06536v1.pdf

Ryan

Data Scientist

Previous Post

Day 104: NLP Papers Summary - SentiHood: Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods

Next Post

Day 106: NLP Papers Summary - An Unsupervised Neural Attention Model for Aspect Extraction

Data Science Natural Language Processing NLP Papers Summary

Day 365: NLP Papers Summary – A Survey on Knowledge Graph Embedding

Ryan
30th December 2020

Data Science Implementation Natural Language Processing

Day 364: Ryan’s PhD Journey – OpenKE-PyTorch Library Analysis + code snippets for 11 KE models

Ryan
29th December 2020

Data Science Ryan's PhD Journey

Day 363: Ryan's PhD Journey – Literature Review – Knowledge Acquisition – 1st Passes XIV

Ryan
28th December 2020