

NLP ASSIGNMENT 4

Yash Aggarwal (2019480)

Bhavya Narang (2019462)

Pre Processing

Given an XML file I did the following tasks:

- Convert the XML file to the ordered dictionary and then extract all the relevant information, that is, text, aspect terms, sentence id, and mapped the polarities to negative is 0, neutral is 1, and positive is 2 rest are -1 which are not for the aspect terms.
- Also, we used a tokenizer to obtain words from the sentences.
- We have used the BIO encoding scheme to represent the tags of aspects, that is the first word of aspect is assigned 1 and the second word is 2 as our window size is 2.

Methodology

We have used the bert-base-uncased pre-trained model.

1. Aspect Term Extraction

In this task, we have to extract the Aspect Term from the sentence using the BERT layer followed by LSTM. For BERT, we use a pre-trained model, and for LSTM we have used the PyTorch implementation.

We have used PyTorch DataLoader to create mini-batches for training the data. We are using the sentences and their aspect terms to train our model and then using it to test the model.

2. Aspect Based Sentiment Analysis

For Aspect Based Sentiment Analysis we will use the Sentence and the aspect for predicting the polarity of the given aspect. We are using the BERT layer followed by 1D CNN. The model implementation is similar to the previous one, only that it is using 1D CNN.

Assumption: We have used a window size of 2 for ABSA.