

Baseline System Description

This document is intended to provide a high level description of the baseline system for the ABSA (Aspect Based Sentiment Analysis) SemEval 2014 task. The baseline system is available for download from the ABSA SemEval 2014 task description page.

The purpose of the baseline system is to provide ABSA participants with a mechanism to test their own purpose-built systems against and to demonstrate what can be achieved using simple methods.

For each of the ABSA subtasks, the method employed in the baseline system is described below.

Aspect Term Extraction

Build a dictionary of aspect terms by collecting all aspect terms present in the training data. To identify aspect terms in the test data, match with collected aspect terms from the dictionary.

Clearly, this method will only identify previously tagged aspect terms in the training data.

Aspect Category Detection

For every test sentence, the k training sentences that are most similar to the test sentence are first retrieved. The similarity between two sentences is measured as the Dice coefficient of the sets of (distinct) words of the two sentences. For example, the similarity between the sentences "this is a demo" and "that is yet another demo" is $2 * 2 / (4 + 5) = 0.44$.

The most frequent number (let us call it m) of aspect categories per sentence among the k retrieved training sentences is then computed. Assume, for example, that $k = 5$ and that the retrieved training sentences have the following aspect categories; then $m = 2$.

sentence 1: food
sentence 2: food, service
sentence 3: price
sentence 4: food, price
sentence 5: food, ambience

The test sentence is then assigned the m most frequent aspect categories of the k training sentences. In the example above, the test sentence would be assigned the $m = 2$ aspect categories "food" (frequency 4) and "price" (frequency 2).

By default $k = 5$. The implementation of this baseline includes a parameter *multi*. If *multi* is set to True, then m is computed as above; if *multi* = False, then m is always set to 1.

Aspect Term Polarity

For each aspect term in a test sentence, check if the aspect term had been encountered in the training sentences. If it had been encountered, assign to the aspect term (in the test sentence) the most frequent polarity (positive, negative, neutral, or conflict) that the aspect term had in the k most similar training sentences (as in the aspect category detection baseline) that contained the aspect term. If the aspect term had not been encountered in the training sentences, assign it (in the test sentence) the most frequent polarity of all the aspect terms in the training sentences.

Aspect Category Polarity

For each test sentence and for each aspect category of the test sentence, assign to the aspect category (of the test sentence) the most frequent polarity that the aspect category had in the k most similar training sentences (as in the aspect category detection baseline) that had the same aspect category.