

ASPECT-BASED SENTIMENT ANALYSIS

ENAYAT ULLAH

ANKIT SINGH

PROBLEM STATEMENT

- Aspect : An explicit reference of an entity towards which an opinion is expressed in the given text.

eg: 'This is a fun restaurant to go to.'

Aspect : Restaurant.

- The task is focussed on two domains: restaurants and laptops.
- Dataset:

Two datasets of 550 reviews of laptops and restaurants annotated with opinion tuples are provided by SemEval in XML format.

PROBLEM STATEMENT

- Subtask 1: Aspect term extraction.

Identification of an opinion target expression (explicit reference) towards which an opinion is expressed.

- Subtask 2: Sentiment Analysis

Each identified aspect of the given text is then assigned a polarity, from the set $P = \{\text{positive, negative, neutral}\}$

PREVIOUS WORK

- Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns, Yejin Choi and Claire Cardie, Ellen Riloff and Siddharth Patwardhan.

Used Conditional Random Fields (CRF) with features extracted using named entity recognition, POS tagging, parsing, and semantic analysis.

- Mining and summarizing customer reviews, M. Hu and B. Liu. 2004. In Proceedings of KDD, pages 168–177, Seattle, WA, USA.

Word2Vec

- A neural network implementation that learns representations for words using a skip-gram model.
- Using large amounts of unannotated plain text, word2vec learns relationships between words within a context window of fixed length.
- The words are realized as vectors in an N-dimensional space.

Approach: Aspect Term Extraction

- Extracts all the distinct nouns and noun phrases from the reviews and consider them candidate distinct aspect terms.
- Longer candidate distinct aspect terms are also included, in the order they appear in the sentence.
- The *p-support* of a candidate distinct aspect term t is the number of sentences that contain t , excluding sentences that contain another candidate distinct aspect term t' that subsumes t .

eg : “The battery life was good” is counted in the p -support of ‘battery life’, but not in the p -support of ‘battery’.

Approach: Aspect Term Extraction

- Pruning:

Discard multi-word distinct aspect terms that appear in 'non-compact' form in more than one sentences.

If the p-support of a candidate distinct aspect term t is smaller than 3 and t is subsumed by another candidate distinct aspect term t' , then t is discarded

- W2V:

- The English Wikipedia is used to train the language model and obtain word vectors.

Approach: Aspect Term Extraction

- Centroid of the ten most frequent candidate distinct aspect term vectors is computed, dubbed the *domain centroid*.
- Similarly, centroid of the 20 most frequent words of the Brown Corpus, excluding stop-words is computed to form the common language centroid.
- Any candidate distinct aspect term whose vector is closer to the common language centroid than the domain centroid is discarded and vice-versa (using cosine similarity as a measure of distance).

Approach: Sentiment Analysis

- A vector representation of the sentence is computed(linear sum of the word vectors or using paragraph2vec).
- Compute the cosine similarity with a certain number of words from SentiWordnet.
- The polarity of the nearest word is adjudged as the polarity of the sentence.

References

1. SemEval-2015 Task 12: Aspect Based Sentiment Analysis
<http://alt.qcri.org/semeval2015/task12/>
2. M. Hu and B. Liu. 2004. Mining and summarizing customer reviews. In Proceedings of KDD, pages 168–177, Seattle, WA, USA.
3. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns, Yejin Choi and Claire Cardie, Ellen Riloff and Siddharth Patwardhan
<https://www.cs.cornell.edu/home/cardie/papers/hlt-emnlp05-yejin.pdf>