Sentic Demo: A Hybrid Concept Level Aspect Based Sentiment Analysis Toolkit

Soujanya Poria[1] and Alexander Gelbukh[2]

[1] School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore
[2] Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico
sporia@ntu.edu.sg
www.gelbukh.com

Abstract. The ways people express their opinions and sentiments have radically changed in the past few years thanks to the advent of social networks, web communities, blogs, wikis, and other online collaborative media. The automatic analysis of online opinions, in fact, involves a deep understanding of natural language text by machines, from which we are still very far. In this work, we introduce a novel paradigm to concept-level sentiment analysis that merges linguistics, common-sense computing, and machine learning for improving the accuracy of tasks such as polarity detection. By allowing sentiments to flow from concept to concept based on the dependency relation of the input sentence, in particular, we achieve a better understanding of the contextual role of each concept within the sentence and, hence, obtain a polarity detection engine that outperforms state-of-the-art statistical methods.

1 Concept Parser and Aspect Parser

1.1 Concept Parser

Given a sentence our proposed concept parser extracts the concepts from the sentence. The concept parser uses the dependency parse tree of a sentence produced by *Stanford Dependency parser available at - http://nlp.stanford.edu:8080/parser/index.jsp* and based on some hand crafted rules the concept parser extracts concepts using the dependency parse tree. Few of the rules which are used in the algorithm are given below.

Subject noun Rule

Trigger: when the active token is found to be the syntactic subject of a verb.
Behavior: if a word *h* is in a subject noun relationship with a word *t* then the concept *t-h* is extracted.
Example: In (1), *movie* is in a subject relation with *boring*.

(1) The movie is boring.

Here the concept `(boring-movie)` is extracted.

Joint Subject noun and Adjective complement rule

Trigger: when the active token is found to be the syntactic subject of a verb and the verb is on adjective complement relation with an adverb.

Behavior: if a word *h* is in a subject noun relationship with a word *t* and the word *t* is with adjective complement relationship with a word *w* then the concept *w*-*h* is extracted.

Example: In (2), *flower* is in a subject relation with *smells* and *smells* is in adjective complement relationship with *bad*.

(2) The flower smells bad.

Here the concept `(bad-flower)` is extracted.

Experiments and Results To calculate the performance, we selected 300 sentences from the *Stanford Sentiment Dataset* [2] and extracted the concepts manually. This process yielded 3204 concepts. We achieve 92.01% accuracy in this experiment.

1.2 Aspect Parser

Aspect-based opinion mining aims to model relations between the polarity of a document and its opinion targets, or aspects. Our system is able to extract both implicit and explicit aspect extraction tasks.

Formation of Implicit aspect lexicon We use the product review dataset by [1] to create the implicit aspect lexicon. First of all, the sentences having implicit aspects were extracted. Then from those sentences we extracted the implicit aspect term and also manually labeled the term using its category. For example - *The car is expensive* - in this sentence, we extract the implicit aspect term *expensive* and labeled this aspect term *expensive* by *price* category. Finally, we obtained following categories -

- functionality
- weight
- price
- appearance
- behavior
- performance
- quality
- service
- size

We then obtained the synonyms of all extracted implicit aspect terms from the wordnet and prepared a large lexicon where each implicit aspect term is categorized into the above categories. The lexicon has 1128 implicit aspect terms labeled by their corresponding aspect categories.

Opinion Lexicon We use SenticNet 2.1 as an opinion lexicon. It contains 14000 common sense knowledge concepts which is labeled by their polarity scores.

First of all, Stanford parser is used to get the dependency parse structure of the sentences. Then, hand crafted dependency rules are employed on the parse trees to extract aspects. Below are the few rules which are used in our system.

Subject Noun Rule *Trigger:* when the active token is found to be the syntactic subject of a token. *Behavior:* if a word *h* is in a subject noun relationship with a word *t* then -

1. if *t* has any adverbial or adjective modifier and the modifier exists in the SenticNet then we extract *t* as an aspect.
2. if the sentence does not have auxiliary verb i.e. *is, was, would, should, could etc.* then -
   - if the verb *h* is modified by an adjective or adverb or is in *adverbial clause modifier* relation in another token then *h* is extracted as an aspect. In (3), *battery* is in a subject relation with *lasts* so the aspects *last* and *battery* are extracted.

     (3) The battery lasts little.

   - the subject *t* is also extracted as an opinion target.
   - if *h* has any direct object relation with a token *n* and the part of speech of the token is *Noun* and *h* is not in opinion lexicon i.e. SenticNet then *n* is extracted as an aspect. In (4), *like* is in direct object relation with *lens* so the aspect *lens* is extracted.

     (4) I like the lens of this camera.

   - if *h* has any direct object relation with a token *n* and the part of speech of the token is *Noun* and *h* is in opinion lexicon i.e. SenticNet then in an addition to the *n* the token having *Noun* as its part of speech is extracted as an aspect. In (5), *like* is in direct object relation with *beauty* which is connected to *screen* via a preposition relation. So the aspects *screen* and *beauty* are extracted.

     (5) I like the beauty of the screen.

3. A copula is the relation between the complement of a copular verb and the copular verb. If the token *h* is in coupla relation of a couplar verb and the couplar verb exists in the implicit aspects lexicon then we extract *h* as an aspect. In (6) *expensive* is extracted as an aspect.

   (6) The car is expensive.

Sentences which do not have subject noun relation in their parse tree For sentences which do not have noun subject relation in their parse trees, we extract the aspects from those sentences using following rules -

1. if an adjective or adverb *h* is in infinitival or open clausal complement relation with a token *t* and *h* exists in the implicit aspects lexicon then we extract *h* as an

aspect. In (7) we extract *big* as an aspect as it is connected to *hold* using a clausal complement relation.

(7) Very big to hold.

2. if a token *h* is connected to a noun *t* using a prepositional relation then we extract both *h* and *t* as aspects. In (8) *sleekness* is extracted as an aspect.

(8) love the sleekness of the player.

Obtaining implicit aspect categories After obtaining the aspects using the abovementioned rules we extract the categories of the implicit aspects from the implicit aspects lexicon. We first search each aspect in the lexicon and if it is found the corresponding aspect category is extracted.

Experiments and Results We carried out our experiment on Semeval 2014 aspect based sentiment analysis data obtained from `http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools`. On this dataset we got 91.25% precision and 88.12% recall.

2 Commonsense knowledge based Sentiment Analysis

This section describes the algorithm used to compute the polarity score of a sentence. We use an ensemble algorithm for detecting sentiment from the sentences. Below we first describe few rules used in the system followed by the ensemble classification step.

2.1 Dependency rules

In this section we go over all of the rules that have been implemented to deal with specific dependency patterns. The main goal of these rules is to drive the way concepts are searched in SenticNet.

Subject nouns

Trigger: when the active token is found to be the syntactic subject of a verb.
Behavior: if the complex concept `(h,d)` is found it is used to calculate the polarity of the relation, otherwise nothing is done: subsequent relations will be activated later.
Example 1: In (9), *movie* is in a subject relation with *boring*.

(9) The movie is boring.

If the concept `(movie, boring)` is in SenticNet, its polarity will be used.

Adjective and clausal complements These rules deal with verbs having as complements either an adjective or a closed clause (i.e. a clause, usually finite, with its own subject).

Modifiers Modifiers, by definition, affect the interpretation of the head they modify. This explains why in most of the following rules the dependent is the guiding element for the computation of polarity.

2.2 Machine Learning Technique

For each sentence, we extracted concepts from it and looked them up in SenticNet. If we found at least one concept in SenticNet, then we used our knowledge-based method to detect sentiment. Otherwise, we relied on our machine learning-based technique. The Machine Learning module is trained on blitzer dataset using some relevant features described below.

Sentic Feature The polarity scores of each concept extracted from the sentence were obtained from the SenticNet and summed up to produce a single scalar feature.

Part of Speech Features This feature was defined by number of adjectives, adverbs, and nouns in the sentence, which gave three distinct features. in, adjective, or adverb. The modification feature was set to 1 if we found any modification relation in the sentence; otherwise it was set to 0.

2.3 Results on the Blitzer-derived Dataset

On the sentence level Blitzer dataset 87.00% accuracy was achieved.

3 Demos

The demos of aspect and concept parsing and the sentiment analysis are freely available at `http://www.soujanyaporia.com/software/`.

References

1. Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27, 2011.
2. Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. EMNLP, 2013.