A Rule-Based Approach to Aspect Extraction from Product Reviews

1 Introduction

Aspect extraction is necessary to first deconstruct sentences into product features and then assign a separate polarity value to each of these features.

There are two types of aspects defined in aspect-based opinion mining: explicit and implicit.

Explicit aspects are concepts that explicitly denote targets in the opinionated sentence. On the other hand, an aspect can also be expressed indirectly through an implicit aspect clue.

Explicit aspect extraction has been widely researched and there exists several approaches for this task. Still, limited work has been done in extracting implicit aspects. This task is very difficult yet very important because the phenomenon of implicit aspects is present in nearly every opinionated document.

In contrast to the task of identification of explicit aspects, the general scheme for identification of implicit aspects, a task called implicit aspect extraction, typically involves two steps:

1. Identify IACs (e.g., “sleek”) in the opinionated document.

2. Map them to the corresponding aspects (e.g., appearance).

This phenomenon is known in opinion mining as desirable fact: communicating fact that by commonsense are good or bad, which indirectly implies polarity.

In this paper, we present a rule-based approach that exploits common-sense knowledge and sentence dependency trees to detect both implicit and explicit aspects.

2 Related Work

Aspect extraction can be seen as a general information extraction problem, for which techniques based on sequential labeling are generally used.

3 Method

3.1 Corpus for aspect extraction

3.2 Pre-Processing

Pre-processing is a key step for aspect parsing. The pre-processing module of the proposed framework consists of two major steps: firstly, the sentence dependency tree is obtained through Stanford Dependency Parser3; secondly, dependency structure elements are processed by means of Stanford Lemmatizer for each sentence. It is important to build the dependency tree before lemmatization as swapping the two steps results in several imprecisions caused by the lower grammatical accuracy of lemmatized sentences.

3.3 Aspect Parser

3.3.1 Implicit aspect lexicon

We use the implicit aspect corpus developed by Cruz-Garcia et al. (Cruz-Garcia et al., 2014), where IACs are indicated and manually labeled by their corresponding aspect categories.

For our task, we extracted the sentences having implicit aspects and then extracted IACs for each of them, along with their corresponding labeled categories. For example, in “The car is expensive” the IAC is expensive and it is labeled by the category price.

3.3.2 Opinion Lexicon

We use SenticNet 3 as a concept-level opinion lexicon. The proposed aspect parser is based on two general rules:

Rules for the sentences having subject verb.

Rules for the sentences which do not have subject verb.

A dependency relation is a binary relation characterized by the following features:

The type of the relation that specifies the nature of the (syntactic) link between the two elements in the relation.

The head of the relation: this is the element that is the pivot of the relation. Core syntactic and semantics properties (e.g., agreement) are inherited from the head.

The dependent is the element that depends on the head and which usually inherits some of its characteristics (e.g., number, gender in the case of agreement).

First of all, Stanford parser is used to obtain the dependency parse structure of each sentence. Then, hand-crafted dependency rules are employed on the parse trees to extract aspects.

**3.3.3 Subject Noun Rule**

Trigger: when the active token is found to be the syntactic subject of a token.

Behavior: if an active token 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 SenticNet, then t is extracted as an aspect.
2. if the sentence does not have auxiliary verb, i.e., is, was, would, should, could, then:

if the verb t is modified by an adjective or an adverb or it is in adverbial clause modifier relation with another token, then both h and t are extracted as aspects.

if t has any direct object relation with a token n and the POS of the token is Noun and n is not in SenticNet, then n is extracted as an aspect.

if t has any direct object relation with a token n and the POS of the token n is Noun and n exists in SenticNet, then the token n extracted as aspect term.

if t is in open clausal complement relation with a token t1, then the aspect t-t1 is extracted if t-t1 exists in the opinion lexicon. If t1 is connected with a token t2 whose POS is Noun, then t2 is extracted as an aspect.

1. A copula is the relation between the complement of a copular verb and the copular verb. If the token t is in copula relation with a copular verb and the copular verb exists in the implicit aspect lexicon, then t is extract as aspect term.
2. If the token t is in copula relation with a copular verb and the POS of h is Noun, then h is extracted as an explicit aspect.
3. If the token t is in copula relation with a copular verb and the copular verb is connected to a token t1 using any dependency relation and t1 is a verb, then both t1 and t are extracted as implicit aspect terms, as long as they exist in the implicit aspect lexicon.

**3.3.4 Sentences which do not have subject noun relation in their parse tree**

For sentences that do not have noun subject relation in their parse trees, aspects are extracted using the following rules:

1. if an adjective or adverb h is in infinitival or open clausal complement relation with a token t and exists in the implicit aspect lexicon, then h is extracted as an aspect.
2. if a token h is connected to a noun t using a prepositional relation, then both h and t are extracted as aspects.
3. if a token h is in a direct object relation with a token t, t is extracted as aspect.
   * 1. **Additional Rules**

For each aspect term extracted above, if an aspect term h is in co-ordination or conjunct relation with another token t, then t is also extracted as an aspect.

A noun compound modifier of an NP is any noun that serves to modify the head noun. If t is

extracted as an aspect and t has noun compound modifier h, then the aspect h-t is extracted and t is removed from the aspect list.

4 Novelty of the proposed work

First of all, the proposed method is fully unsupervised and depends on the accuracy of the dependency parser and the opinion lexicon, rather then a training corpus and supervised learning accuracy.

5 Experiments and Results

6 Conclusion

We have illustrated a method for extracting both explicit and implicit aspects from opinionated text.