Sentiment Classification: Linguistic and Non-linguistic Issues

Mori Rimon
The Hebrew University
e-mail: rimon@cs.huji.ac.il

Abstract

There is a growing interest in the NLP community in methods for determining the sentiment (tone, polarity, semantic orientation) of a given piece of text (see references below). The large number of potential applications, such as quantitative summarization of customer reviews, public opinion surveys, business intelligence, trend analysis, etc. justifies this. From a research point of view, sentiment classification is an interesting and challenging problem, involving diverse linguistic and non-linguistic considerations.

A related task is extraction of representative phrases, to be used as quotes in a summary, supporting the sentiment evaluation of a review. Quotes of interest may be generally positive or negative, or pertain to a specific dimension (topic) of the subject domain.

In this presentation, I will describe the common approaches to tackle this challenge and will suggest directions that I believe to be the most appropriate. I will discuss the contribution of various levels of linguistic analysis to the effectiveness of the classification method.

Briefly:

The most common and basic approach to sentiment classification is keyword-based. In this approach, terms, mainly adjectives (e.g. *awesome*, *awful*) and fixed expressions (e.g. *dream come true*, *stay away*), are used as sentiment indicators. The list of indicators can be prepared manually (the most naïve approach), composed semi-automatically using sources such as Wordnet (Miller et al [1993]), or acquired by machine learning algorithms that infer the best indicators from tagged samples in the domain of interest.

While keyword-based approaches (manual and learning) prove to be quite effective for simple topic classification tasks, they seem to be limited when applied to more demanding tasks such as sentiment classification. Even if the classifier knows that *recommend* is a positive term, it may be unable to conclude that the following is a negative statement in a book review: *the overly detailed approach makes it a hard book to recommend enthusiastically*. This is just one of many examples that one finds in real texts. In fact, the number of simple sentences in real texts is quite low.

Adding a linguistic dimension to a sentiment classifier makes a real difference. Two enhancements are found to contribute most:
(a) The addition of "composite features" of a syntactic nature to the set of tentative indicators. Verb-Object pairs are one example: *break the law* has negative connotations, while *break a record* has positive ones; each word in itself is neutral.
(b) Using polarity markers, particularly polarity reversing adverbs like *not*, *never*, etc. To be able to extract the required information from a variety of language constructions and use it properly, one should use a parser. The verb and its complements are not necessarily adjacent; an adverb can be placed far from the verb it modifies; etc. A parser based on a dependency grammar, e.g. Connexor's Functional Dependency Grammar

(Tapanainen & Jarvinen [1997]), appears to be more appropriate for the job than conventional phrase-structure parsers (although, as shown in Golan-Lappin-Rimon [1988], basic grammatical roles can be reliably inferred from phrase-structure grammars as well).

Once each sentence in a given text is evaluated, there is still a tough problem to combine sentence-level ratings to a global sentiment score. This is particularly hard when the source text contains sections that are not actually reviews. Relatively simple discourse heuristics help assign different weights to different sections and filter out irrelevant text.

What about deeper linguistic processing, e.g. a more elaborate discourse component or semantically-oriented analysis? These measures are much more complicated and hard to implement and, in the current state of the art, do not seem to contribute enough to the quality of sentiment classification to justify the effort.

References

K. Crammer and Y. Singer. Pranking with Ranking (2001).
I. Golan, S. Lappin and M. Rimon. Computing Grammatical Functions from
    Configurational Parse Trees (1989).
V. Hatzivassiloglou and K. McKeown. Predicting the Semantic Orientation of
    Adjectives (1997).
M. Hu and B. Liu. Mining and Summarizing Customer Reviews (2004).
S.M. Kim and E. Hovy. Determining the Sentiment of Opinions (2004).
L. Lee. A Matter of Opinion: Sentiment Analysis and Business Intelligence (2005).
G. Miller et al. Introduction to Wordnet (1993).
P. Tapanainen and T. Jarvinen. A Non-Projective Dependency Parser (1997).
P. Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised
    classification of reviews (2002).
J.M. Wiebe. Learning subjective adjectives from corpora (2000).