Sentiment Classification: Linguistic and Non-linguistic Issues

Mori Rimon
The Hebrew University
rimon@cs.huji.ac.il

Extended 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 selected 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.

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 predict the best indicators on the basis of tagged samples in the domain of interest.

While keyword-based approaches (particularly those that apply learning) prove to be quite effective for simple topic classification tasks, they seem to be limited when applied to the more demanding task of 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 sentiment indicators. Verb-Object pairs are the most meaningful composite feature: *break the law*, for example, has negative connotations, while *break a record* has positive ones; each word in itself is neutral. Other useful syntactic relations include Adjective-Verb(inf) pairs, as in *hard to recommend*.

(b) Identification and marking of various types of negation: polarity-reversing adverbs like *not* and *never*; NP-internal indicators, like in *no reviewer has ever recommended this book*; etc. Once detected, negation is marked on the verb (or verb-object) feature, as in the following example: *NEG*recommend.

To be able to extract the required information from a variety of language constructions and use it properly, one should use a syntactic 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). Rules are applied to the output of the dependency parser to extract the data we need; for example, a *NEG*verb condition for a subject NP internal indicator would be: det[no]:> subj:> main[verb] (the ">" signifies a link from a dependent to a head).

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. Not all sentences have the same level of significance. The task is particularly hard when the source text contains sections that are not actually reviews. However, relatively simple discourse heuristics can 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 a semantically-oriented analysis? These are much more complicated measures which, given the current NLP technologies, are hard to implement and their expected contribution does not justify the effort. In any case, automatic sentiment classification can never be perfect: the orientation of a statement may depend on context, world knowledge, style (e.g. irony), all of which are still far beyond the current state of the art.

For the set of slides of the IATL/ISCOL talk, see:
http://cs.haifa.ac.il/~shuly/iscol/index.html

References

K. Crammer and Y. Singer. Pranking with Ranking. Proceedings of the NIPS-01 conference, 2001.

I. Golan, S. Lappin and M. Rimon. Computing Grammatical Functions from Configurational Parse Trees. Technical Report 88.268, IBM Haifa Research Lab., 1989.

V. Hatzivassiloglou and K. McKeown. Predicting the Semantic Orientation of Adjectives. Proceedings of the ACL-97 conference, 1997.

M. Hu and B. Liu. Mining and Summarizing Customer Reviews. Proceedings of the KDD-04 conference, 2004.

S.M. Kim and E. Hovy. Determining the Sentiment of Opinions. Proceedings of the COLING-04 conference, 2004.

L. Lee. A Matter of Opinion: Sentiment Analysis and Business Intelligence. Position Paper, 2005.

G. Miller et al. Introduction to Wordnet. International Journal of Lexicography, 1993.

P. Tapanainen and T. Jarvinen. A Non-Projective Dependency Parser. Proceedings of the 5th Conference on Applied Natural Language Processing, 1997.

P. Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the ACL-02 conference , 2002.

J.M. Wiebe. Learning subjective adjectives from corpora. Proceedings of the AAAI-00 conference, 2000.