

Incorporating topic information into sentiment analysis models

Tony Mullen

National Institute of Informatics (NII)
Hitotsubashi 2-1-2, Chiyoda-ku
Tokyo 101-8430,
Japan,
mullen@nii.ac.jp

Nigel Collier

National Institute of Informatics (NII)
Hitotsubashi 2-1-2, Chiyoda-ku
Tokyo 101-8430,
Japan,
collier@nii.ac.jp

Abstract

This paper reports experiments in classifying texts based upon their favorability towards the subject of the text using a feature set enriched with topic information on a small dataset of music reviews hand-annotated for topic. The results of these experiments suggest ways in which incorporating topic information into such models may yield improvement over models which do not use topic information.

1 Introduction

There are a number of challenging aspects in recognizing the favorability of opinion-based texts, the task known as sentiment analysis. Opinions in natural language are very often expressed in subtle and complex ways, presenting challenges which may not be easily addressed by simple text categorization approaches such as n-gram or keyword identification approaches. Although such approaches have been employed effectively (Pang et al., 2002), there appears to remain considerable room for improvement. Moving beyond these approaches can involve addressing the task at several levels. Negative reviews may contain many apparently positive phrases even while maintaining a strongly negative tone, and the opposite is also common.

This paper attempts to address this issue using Support Vector Machines (SVMs), a well-known and powerful tool for classification of vectors of real-valued features (Vapnik, 1998). The present approach emphasizes the use of a variety of diverse information sources. In particular, several classes of features based upon the proximity of the topic with phrases which have been assigned favorability values are described in order to take advantage of situations in which the topic of the text may be explicitly identified.

2 Motivation

In the past, work has been done in the area of characterizing words and phrases according to their emotive tone (Turney and Littman, 2003; Turney, 2002; Kamps et al., 2002; Hatzivassiloglou and Wiebe, 2000; Hatzivassiloglou and McKeown, 2002; Wiebe, 2000), but in many domains of text, the values of individual phrases may bear little relation to the overall sentiment expressed by the text. Pang et al. (2002)'s treatment of the task as analogous to topic-classification underscores the difference between the two tasks. A number of rhetorical devices, such as the drawing of contrasts between the reviewed entity and other entities or expectations, sarcasm, understatement, and digressions, all of which are used in abundance in many discourse domains, create challenges for these approaches. It is hoped that incorporating topic information along the lines suggested in this paper will be a step towards solving some of these problems.

3 Methods

3.1 Semantic orientation with PMI

Here, the term *semantic orientation* (SO) (Hatzivassiloglou and McKeown, 2002) refers to a real number measure of the positive or negative sentiment expressed by a word or phrase. In the present work, the approach taken by Turney (2002) is used to derive such values for selected phrases in the text. For the purposes of this paper, these phrases will be referred to as *value phrases*, since they will be the sources of SO values. Once the desired value phrases have been extracted from the text, each one is assigned an SO value. The SO of a phrase is determined based upon the phrase's *pointwise mutual information* (PMI) with the words "excellent" and "poor". PMI is defined by Church and Hanks (1989) as follows:

$$\text{PMI}(w_1, w_2) = \log_2 \left(\frac{p(w_1 \& w_2)}{p(w_1)p(w_2)} \right) \quad (1)$$

where $p(w_1 \& w_2)$ is the probability that w_1 and w_2 co-occur.

The SO for a *phrase* is the difference between its PMI with the word “excellent” and its PMI with the word “poor.” The method used to derive these values takes advantage of the possibility of using the World Wide Web as a corpus, similarly to work such as (Keller and Lapata, 2003). The probabilities are estimated by querying the AltaVista Advanced Search engine¹ for counts. The search engine’s “NEAR” operator, representing occurrences of the two queried words within ten words of each other in a text, is used to define co-occurrence. The final SO equation is

$$\text{SO}(\textit{phrase}) = \log_2 \left(\frac{\text{hits}(\textit{phrase} \text{ NEAR } \textit{“excellent”}) \text{hits}(\textit{“poor”})}{\text{hits}(\textit{phrase} \text{ NEAR } \textit{“poor”}) \text{hits}(\textit{“excellent”})} \right)$$

Intuitively, this yields values above zero for phrases with greater PMI with the word “excellent” and below zero for greater PMI with “poor”. A SO value of zero would indicate a completely neutral semantic orientation.

3.2 Osgood semantic differentiation with WordNet

Further feature types are derived using the method of Kamps and Marx (2002) of using WordNet relationships to derive three values pertinent to the emotive meaning of adjectives. The three values correspond to the *potency* (strong or weak), *activity* (active or passive) and the *evaluative* (good or bad) factors introduced in Charles Osgood’s Theory of Semantic Differentiation (Osgood et al., 1957).

These values are derived by measuring the relative minimal path length (MPL) in WordNet between the adjective in question and the pair of words appropriate for the given factor. In the case of the *evaluative* factor (EVA) for example, the comparison is between the MPL between the adjective and “good” and the MPL between the adjective and “bad”.

Only adjectives connected by synonymy to each of the opposites are considered. The method results in a list of 5410 adjectives, each of which is given a value for each of the three factors referred to as EVA, POT, and ACT. Each of these factors’ values are averaged over all the adjectives in a text, yielding three real-valued feature values for the text, which will be added to the SVM model.

3.3 Topic proximity and syntactic-relation features

In some application domains, it is known in advance what the topic is toward which sentiment is to be evaluated. Incorporating this information is done by creating several classes of features based upon the semantic orientation values of phrases given their position in relation to the topic of the text. The approach allows secondary information to be incorporated where available, in this case, the primary information is the specific record being reviewed and the secondary information identified is the artist.

Texts were annotated by hand using the Open Ontology Forge annotation tool (Collier et al., 2003). In each record review, references (including co-reference) to the record being reviewed were tagged as THIS_WORK and references to the artist under review were tagged as THIS_ARTIST.

With these entities tagged, a number of classes of features may be extracted, representing various relationships between topic entities and value phrases similar to those described in section 3.1. The classes looked at in this work are as follows:

Turney Value The average value of all value phrases’ SO values for the text. Classification by this feature alone is not the equivalent of Turney’s approach, since the present approach involves retraining in a supervised model.

In sentence with THIS_WORK The average value of all value phrases which occur in the same sentence as a reference to the work being reviewed.

¹www.altavista.com

Following THIS_WORK The average value of all value phrases which follow a reference to the work being reviewed directly, or separated only by the copula or a preposition.

Preceding THIS_WORK The average value of all value phrases which precede a reference to the work being reviewed directly, or separated only by the copula or a preposition.

In sentence with THIS_ARTIST As above, but with reference to the artist.

Following THIS_ARTIST As above, but with reference to the artist.

Preceding THIS_ARTIST As above, but with reference to the artist.

The features used which make use of adjectives with WordNet derived Osgood values include the following:

Text-wide EVA The average EVA value of all adjectives in a text.

Text-wide POT The average POT value of all adjectives in a text.

Text-wide ACT The average ACT value of all adjectives in a text.

TOPIC-sentence EVA The average EVA value of all adjectives which share a sentence with the topic of the text.

TOPIC-sentence POT The average POT value of all adjectives which share a sentence with the topic of the text.

TOPIC-sentence ACT The average ACT value of all adjectives which share a sentence with the topic of the text.

The grouping of these classes should reflect some common degree of reliability of features within a given class, but due to data sparseness what might have been more natural class groupings—for example including **value-phrase_preposition_topic-entity** as a distinct class—often had to be conflated in order to get features with enough occurrences to be representative.

4 Experiments

The dataset consists of 100 record reviews from the Pitchfork Media online record review publication,² topic-annotated by hand. Features used include word unigrams and lemmatized unigrams³ as well as the features described in 3.3 which make use of topic information, namely the broader PMI derived SO values and the topic-sentence Osgood values. Due to the relatively small size of this dataset, test suites were created using 100, 20, 10, and 5-fold cross validation, to maximize the amount of data available for training and the accuracy of the results.

SVMs were built using Kudo's TinySVM software implementation.⁴

5 Results

Experimental results may be seen in figure 1. It must be noted that this dataset is very small, and although the results are not conclusive they are promising insofar as they suggest that the use of incorporating PMI values towards the topic yields some improvement in modeling. They also suggest that the best way to incorporate such features is in the form of a separate SVM which may then be combined with the lemma-based model to create a hybrid.

²<http://www.pitchforkmedia.com>

³We employ the Conexor FDG parser (Tapanainen and Järvinen, 1997) for POS tagging and lemmatization

⁴<http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM>

Model	5 folds	10 folds	20 folds	100 folds
All (THIS_WORK and THIS_ARTIST) PMI	70%	70%	68%	69%
THIS_WORK PMI	72%	69%	70%	71%
All Osgood	64%	64%	65%	64%
All PMI and Osgood	74%	71%	74%	72%
Unigrams	79%	80%	78%	82%
Unigrams, PMI, Osgood	81%	80%	82%	82%
Lemmas	83%	85%	84%	84%
Lemmas and Osgood	83%	84%	84%	84%
Lemmas and Turney	84%	85%	84%	84%
Lemmas, Turney, text-wide Osgood	84%	85%	84%	84%
Lemmas, PMI, Osgood	84%	85%	84%	86%
Lemmas and PMI	84%	85%	85%	86%
Hybrid SVM (PMI/Osgood and Lemmas)	86%	87%	84%	89%

Figure 1: Accuracy results (percent of texts correctly classed) for 5, 10, 20 and 100-fold cross-validation tests with Pitchforkmedia.com record review data, hand-annotated for topic.

5.1 Discussion

At the level of the phrasal SO assignment, it would seem that some improvement could be gained by adding domain context to the AltaVista Search. Many—perhaps most—terms’ favorability content depends to some extent on their context. As Turney notes, “unpredictable,” is generally positive when describing a movie plot, and negative when describing an automobile or a politician. Likewise, such terms as “devastating” might be generally negative, but in the context of music or art may imply an emotional engagement which is usually seen as positive. Likewise, using “excellent” and “poor” as the poles in assessing this value seems somewhat arbitrary, especially given the potentially misleading economic meaning of “poor.” Nevertheless, cursory experiments in adjusting the search have not yielded improvements. One problem with limiting the domain (such as adding “AND music” or some disjunction of such constraints to the query) is that the resultant hit count is greatly diminished. The data sparseness which results from added restrictions appears to cancel out any potential gain. It is to be hoped that in the future, as search engines continue to improve and the Internet continues to grow, more possibilities will open up in this regard. As it is, Google returns more hits than AltaVista, but its query syntax lacks a “NEAR” operator, making it unsuitable for this task. As to why using “excellent” and “poor” works better than, for example “good” and “bad,” it is not entirely clear. Again, cursory investigations have thus far supported Turney’s conclusion that the former are the appropriate terms to use for this task.

It also seems likely that the topic-relations aspect of the present research only scratches the surface of what should be possible. Although performance in the mid-80s is not bad, there is still considerable room for improvement. The present models may also be further expanded with features representing other information sources, which may include other types of semantic annotation (Wiebe, 2002; Wiebe et al., 2002), or features based on more sophisticated grammatical or dependency relations, or perhaps upon such things as zoning (e.g. do opinions become more clearly stated towards the end of a text?). In any case, it is hoped that the present work may help to indicate how various information sources pertinent to the task may be brought together.

6 Conclusion

Further investigation using larger datasets is necessary for the purposes of fully exploiting topic information where it is available, but the present results suggest that this is a worthwhile direction to investigate.

References

- K.W. Church and P. Hanks. 1989. Word association norms, mutual information and lexicography. In *Proceedings of the 27th Annual Conference of the ACL*, New Brunswick, NJ.
- N. Collier, K. Takeuchi, A. Kawazoe, T. Mullen, and T. Wattarujekrit. 2003. A framework for integrat-

- ing deep and shallow semantic structures in text mining. In *Proceedings of the Seventh International Conference on Knowledge-based Intelligent Information and Engineering Systems*. Springer-Verlag.
- V. Hatzivassiloglou and K.R. McKeown. 2002. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL*.
- V. Hatzivassiloglou and J. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity.
- Jaap Kamps, Maarten Marx, Robert J. Mokken, and Marten de Rijke. 2002. Words with attitude. In *In Proceedings of the 1st International Conference on Global WordNet*, Mysore, India.
- Frank Keller and Mirella Lapata. 2003. Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics*, 29(3). Special Issue on the Web as Corpus.
- Charles E. Osgood, George J. Succi, and Percy H. Tannenbaum. 1957. *The Measurement of Meaning*. University of Illinois.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Empirical Methods in Natural Language Processing [and Very Large Corpora]*.
- P. Tapanainen and T. Järvinen. 1997. A non-projective dependency parser. In *Proceedings of the 5th Conference on Applied Natural Language Processing, Washington D.C., Association of Computational Linguistics*.
- P.D. Turney and M.L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):315–346.
- P.D. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia.
- Vladimir Vapnik. 1998. *Statistical Learning Theory*. Wiley, Chichester, GB.
- J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin. 2002. Learning subjective language. Technical Report TR-02-100, University of Pittsburgh, Pittsburgh, PA.
- Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *Proc. 17th National Conference on Artificial Intelligence (AAAI-2000)*, Austin, Texas, July.
- J Wiebe. 2002. Instructions for annotating opinions in newspaper articles. Technical Report TR-02-101, University of Pittsburgh, Pittsburgh, PA.