LAMP-TR-143 May 2007
CS-TR-4871
UMIACS-TR-2007-24

Computing the Sentiment Polarity of Chinese Words and Sentences

Yejun Wu and Douglas W. Oard

College of Information Studies and
Computational Linguistics and Information Processing Laboratory
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742-3275
{*wuyj,oard*}@*umd.edu*

Abstract

This paper reports on experiments with a newly available sentiment classification test collection for Chinese. Detection of negation during training and classification is shown to improve the accuracy of character-based classification for the semantic orientation of individual Chinese words. Using the resulting classifier for unknown words resulted in substantial improvements in classification accuracy for positively and negatively opinionated sentences.

Keywords: sentiment polarity, Chinese words, Chinese sentences

The support of this research by DARPA under contract HR-0011-06-2-0001 (GALE) is gratefully acknowledged.

Computing the Sentiment Polarity of Chinese Words and Sentences

Yejun Wu and Douglas W. Oard
College of Information Studies and
Institute for Advanced Computer Studies
University of Maryland, College Park, MD 20742
{wuyj,oard}@umd.edu

Abstract

This paper reports on experiments with a newly available sentiment classification test collection for Chinese. Detection of negation during training and classification is shown to improve the accuracy of character-based classification for the semantic orientation of individual Chinese words. Using the resulting classifier for unknown words resulted in substantial improvements in classification accuracy for positively and negatively opinionated sentences.

1 Introduction

Accurate and efficient techniques for characterizing sentiment could serve many purposes. Supervised machine learning techniques have been applied to identification of semantic polarity at the scale of words [5], sentences [8], and documents [6].

Regardless of scale, however, words typically have provided the base feature set that those classifiers exploited. The vast majority of the reported research on sentiment classification has focused on written English, where tokenization into words is straightforward. Many forms of linguistic expression lack such easily detected word boundaries, however. Examples include spoken language (where automatic transcription can yield overlapping alternative hypotheses), written languages such as Chinese that do not expose word boundaries and freely compounding languages such as German in which some word boundaries are not clearly marked in the written form. One-best segmentation techniques offer an obvious starting point for extending prior work on written English to these more challenging cases, but there is a far larger design space to be explored. Written Chinese presents an attractive point from which to begin this exploration, both because there has been prior work on computing word semantic orientation using subword features and because a test collection for Chinese sentiment classification has recently become available.

The National Institute of Informatics (NII) in Japan conducts a series of language technology evaluations known as the NII Test Collection for IR systems (NTCIR).[1] In 2006, NTCIR-6 established an Opinion Analysis Pilot Task. Comparable (i.e., topically and temporally similar) Chinese, English and Japanese opinion analysis test collections were distributed in late 2006 that included both word and sentence annotation of news texts for training and a news text evaluation set for blind evaluation. Ground truth annotations are not yet available for the evaluation set, so in this paper, we evaluate our results at sentence scale using the training portion of the NTCIR-6 opinion analysis Chinese collection.

In this paper we reimplement the approach introduced by Ku et al. [2] for computing the semantic orientation of Chinese words based on their characters. We extend that model to account for characters that indicate negation, obtaining small improvements in word-level classification accuracy (about 2%) in cross-validation experiments with a manually annotated 17,037-word Chinese lexicon. Some improvements also result from accounting for negation characters when classifying sentiment at the scale of entire sentences when using the lexicon to identify the polarity of known words and character-based classification of polarity for unknown words.

2 Approaches

We have extended Ku et al.'s method in three ways: (1) extending their lexicon with additional hand-annotated terms, (2) automatically enhancing the training lexicon by stripping negation characters and flipping the polarity of the resulting term, and (3) augmenting the classifier with

[1] http://research.nii.ac.jp/ntcir

polarity-flipping when negation characters are encountered.

2.1 Lexicon Acquisition and Preparation

We obtained the Chinese sentiment lexicon used by Ku et al., [2] which contains 2,812 positive words and 8,276 negative words in Traditional Chinese. We also manually rekeyed two books, the Chinese Positive Dictionary and the Chinese Negative Dictionary [4, 7]. This yielded a total of 5,184 positive words and 3,116 negative words. Combining these with Ku et al.'s lexicon, removing duplicates, and cleaning up some punctuation and characters preceding an ellipsis, yielded a positive lexicon with 6,743 Chinese words and a negative lexicon with 10,294 words.

2.2 Word Classifier I: Baseline

Hatzivassiloglou and McKeown [1] defined semantic orientation or polarity as the direction a word deviates from the norm for its semantic group or lexical field. Ku et al. [2] extended this idea to Chinese characters, computing the polarity of a Chinese word as a function of Chinese characters the word contains and assigned a character's *sentiment tendency* values as follows:

$$P_{c_i} = \frac{FP_{c_i}/\sum_{j=1}^{n} FP_{c_j}}{FP_{c_i}/\sum_{j=1}^{n} FP_{c_j} + FN_{c_i}/\sum_{j=1}^{m} FN_{c_j}}$$

$$N_{c_i} = \frac{FN_{c_i}/\sum_{j=1}^{m} FN_{c_j}}{FP_{c_i}/\sum_{j=1}^{n} FP_{c_j} + FN_{c_i}/\sum_{j=1}^{m} FN_{c_j}}$$

where P_{c_i} and N_{c_i} are the positive and negative tendencies of character c_i respectively; FP_{c_i} and FN_{c_j} are frequency of character c_i in positive and negative words respectively; *n* and *m* denote total number of unique characters in positive and negative words respectively, and therefore $\sum_{j=1}^{n} FP_{c_j}$ and $\sum_{j=1}^{m} FN_{c_j}$ are the total number of characters in positive and negative words respectively. Subtracting N_{c_i} from P_{c_i} yields an estimate of the net sentiment expressed by character c_i (which will be positive for positive polarity, and negative for negative polarity). The sentiment for a Chinese word is then computed as the average of the estimated net sentiment for each character in the word.

[2] http://nlg18.csie.ntu.edu.tw:8080/opinion/pub1.html

沒(no) 不(no) 無(no) 未(not yet) 否(no)
毋需 (not need) 尚難(still hard to)

Figure 1: Negation Characters and Bigrams.

傲慢無禮 (arrogant and not polite)
束手無策 (helpless and having no way)

Figure 2: Words with 2 Bigrams.

2.3 Word Classifier II: Using Negation Characters and Bigrams

There are many words in our positive lexicon that contain obvious negation characters (or sometimes negation bigrams); see Figure 1 for some examples. Removing those characters (or bigrams) often yields a word with the opposite polarity. For instance, removing the “not” in “not elegant” (negative), we get “elegant” (positive). We therefore hand coded a Perl script to remove words that contain a negation character (or bigram) from the lexicon, stripped off the negation character (or bigrams), and then added the remaining characters back to the lexicon with opposite polarity (if it is not already there). We expect the effect of this processing to be to sharpen the association of individual characters to a polarity. However, some words (typically those with 4 characters) are essentially composition compounds. These sometimes consist of one part that includes a negation character and one part that does not (see Figure 2 for examples). If we were to strip the negation character in these cases, we would often create a mixed-polarity word that belongs in neither lexicon. We therefore leave a lexicon word unchanged if it contains four or more characters.

2.4 Word Classifier III: Using Negation Rules

Roughly 9 times as many words with negation characters (or bigrams) were found in the original negative lexicon as in the original positive, so polarity flipping in the training data alone would enrich the positive examples at the expense of negative examples. This bias can be overcome, however, by detecting negation characters (or bigrams) at classification time, classifying the remainder of the word, and then flipping the resulting classification.

2.5 Sentence Classification

Character-oriented techniques offer the potential to adopt characters as the base feature set for sentence classification, thus obviating the need to partition sentences into their constituent words. Doing so raises a number of challenges we have not yet had time to address, however. For example, it is not yet clear how higher-order features (e.g., words found in a lexicon, or character n-grams found in training data) can best be incorporated. For our initial exploration of sentence classification we therefore decided to work with a one-best partition by the Stanford Segmenter [3].

We then count the words that are estimated to have positive and negative polarity. For the baseline sentence classifier we make this decision using the lexicon for Word Classifier I. If the word is found in the lexicon, it gets a score of 1 for positive or a score of -1 for negative, otherwise a score of 0. Sentence Classifier I extends this by using both Word Classifier I and the lexicon. In that case, a score of 1 (or -1) can result from finding a word in the positive (or negative) lexicon or (for out of vocabulary words) from computing any positive (or negative) score using character-based classification. Sentence Classifier III uses the same decision logic, but with Word Classifier III . Because negation characters (and bigrams) can be segmented as separate words, we added an additional rule to all the 3 sentence classifiers that flipped the polarity of any word that immediately followed such a negation word. NTCIR sentence classification ground truth includes four possible values: positive, negative, neutral, or not opinionated. We approximated this by automatically changing the classification to "not opinionated" if no word in the sentence was found in the lexicon and achieved a score < -0.5 or > 0.5 (which would typically indicate strong support for the assigned polarity from every character in the word). If a sentence **is** opinionated, it is classified as neutral if the aggregated polarity from words is 0, positive if the aggregated polarity is $>= 1$, negative if the aggregated polarity is $<= -1$.

3 Results

In this section we report word-scale and sentence-scale sentiment classification accuracy.

Approach	Positive	Negative	Average
I	0.8579	0.8690	0.8635
II	0.8735	0.8266	0.8500
III	**0.8843**	**0.8755**	**0.8799**

Table 1: Word Classification Accuracy.

3.1 Word-Scale Results

In order to assess the word classification accuracy we adopted a cross-validation strategy. We first randomly partition the aggregated positive and negative lexicons into 10 positive and 10 negative sets. We then trained a Word Classifier I on the first 9 partitions and evaluated classification accuracy using the remaining partition. We repeated this process 10 times, each with a different evaluation partition, reporting the average classification accuracy across those 10 runs. We repeated this process for Word Classifier II and Word Classifier III.

Table 1 compares word classification accuracy results. This is a two-way classification task with the average evenly balanced between the two conditions, so random guessing would yield 50% average accuracy. All three character-based classifiers do much better than that, with Word Classifier III showing slight gains over Word Classifier I in both conditions (3% for positive, 1% for negative). Word Classifier II performs less well on words with negative polarity, as might be expected from the disproportionate removal of negative words.

3.2 Sentence-Scale Results

Manual annotation of semantic orientation for the sentences in a small collection of news stories was provided as training data for the NTCIR-6 Opinion Analysis Pilot Task. Each sentence was assigned one of four possible values (positive, negative, neutral, not opinionated) by 3 annotators. Interannotator agreement on this task has not yet been characterized, so we employed a voting scheme to automatically construct consensus annotations. We extracted every sentence for which at least 2 annotators agreed for use in our evaluation. This resulted in 838 positive, 286 negative, 172 neutral, and 1,004 not-opinionated sentences.

Table 2 shows our sentence classification results. Our simple heuristics clearly do poorly

Categories	Baseline	I	III
POS+NEG	**0.2767**	**0.3550**	**0.4973**
POS	0.3425	0.3186	**0.5919**
NEG	0.0839	**0.4650**	0.2203
NEU	0.1395	0.1046	0.0930
NOT-OP	0.2191	0	0

Table 2: Sentence classification accuracy. POS: positive; NEG: negative; NEU: neutral; NOT-OP: not opinionated.

at discriminating neutral sentences from not-opinionated ones. For the roughly half of the sentences that express a positive or a negative opinion, Sentence Classifier III chooses correctly about half the time (random performance on what is in essence a 3-way classification task would be 33%). Interestingly, Sentence Classifier I works better than Classifier III for negative instances, suggesting that a combination of classifiers might be effective in this case.

4 Conclusion and Future Work

We have extended a previous method for computing the polarity of Chinese words, achieving modest improvements in classification accuracy by accounting for negation characters and bigrams. Of course, some Chinese words are not constructed in that way. For example, the Chinese word for "shameless," "Wu2 Chi3," would be literally translated as "no shame," a combination of a negation character and a second negative character that retains a negative polarity. Manually freezing the polarity of such words in the sentiment lexicon might be helpful.

Our overly simple approach to sentence-scale sentiment classification did not distinguish well between the "neutral" and "no sentiment" conditions in the NTCIR Opinion Analysis Pilot Task training annotations, we saw clear evidence that computing a polarity for unknown words can help to distinguish sentences with positive and negative polarity. In future work we are interested in exploring both linguistically motivated techniques (e.g., leveraging syntactic analysis to model contextual polarity) and a broader range of shallow approaches (e.g., unifying segmentation and classification in a single model rather than the two-stage cascade that we used here).

5 Acknowledgments

This work has been supported in part by the DARPA contract HR-0011-06-2-0001 (GALE).

References

[1] Hatzivassiloglou, Vasileois, and McKeown, Katheleen, 1997. Predicting the semantic orientation of adjectives. *ACL-97*, 1997, 174-181

[2] Ku, Lun-Wei, Liang, Yu-Ting, and Chen, Hsin-Hsi, 2006. Opinion extraction, summarization and tracking in news and blog corpora. *AAAI Spring Symposium Technical Report SS-06-03*, Palo Alto, California, 2006.

[3] Tseng, Huihsin, Chang, Pichuan, Andrew, Galen, Jurafsky, Daniel and Manning, Christopher. A Conditional Random Field Word Segmenter. In *Fourth SIGHAN Workshop on Chinese Language Processing*. 2005.

[4] Shi, Jilin and Zhu, Yinggui (Ed.), 2005. *Bao Yi Ci Ci Dian (Positive Dictionary)*. Sichuan Dictionary Press.

[5] Turney, Peter and Littman, Michael, 2003. Measuring praise and criticism: inference of semantic orientation from association. *ACM TOIS*, 21. 315-346.

[6] Wilson, Theresa, Wiebe, Janyce, and Hoffmann, Paul, 2005. Recognizing contextual polarity in phrase-level sentiment analysis. *HLT-EMNLP 2005*.

[7] Yang, Ling and Zhu, Yinggui (Ed.), 2005. *Bian Yi Ci Ci Dian (Negative Dictionary)*. Sichuan Dictionary Press, Chengdu, Sichuan, China

[8] Yu, Hong and Hatzivassiloglou, Vasileios, 2003. Toward answering opinion questions *EMNLP-2003*, 129-136