Sentiment Document Classification Using Global and Domain Features

Youngjoong Ko

Department of Computer Engineering, Dong-A University
Busan, 604-714, Korea
Email: youngjoong.ko@gmail.com

Abstract

The goal of sentiment classification is to detect writer’s sentiment from a document. This paper investigates which features and what combination of them is more effective in sentiment classification. Experiments show that the effective combination method of global and domain features can significantly reduce classification errors relative to features which have been used in general text classification.

Key Words: Opinion Mining, Sentiment Classification, Sentiment-bearing Words, Feature Selection

1. Introduction

The research area of sentiment classification has received considerable attention from researchers in recent years [1,2,3,4]. While the focus of traditional text classification is on topic identification, that of sentiment classification is on the assessment of writer’s sentiment toward the topic.

Since a sentiment-bearing (positive/negative) word is a key indicator of sentiment classification, the determination of the orientation of each individual word is a fundamental task for sentiment classification [5]; these sentiment-bearing words are called Global Feature (GF) for sentiment classification. The conceptually simplest approach to collect them is probably Turney’s studies [4,6] which have obtained interesting results on determining orientation by summing up the orientations of words to estimate the orientation of the document they belong to. More sophisticated methods are also possible [7]. In addition, we consider that there is a different type of sentiment-bearing words. They do not have any sentimental polarity in common usage but they can show some sentimental polarity when using in some special domains; they are called Domain Feature (DF).

This paper proposes how to effectively extract and use sentiment-bearing words as global and domain features. Then their availability is evaluated in sentiment classification. Many previous studies have used a thesaurus to build up English sentiment-bearing words, whereas ones for other languages such as Korean, which do not have any thesaurus, have suffered for obtaining good-qualified sentiment-bearing words as global features. Thus we construct a
Korean sentiment-bearing word set as global features through the following three methods: English-Korean translation words, annotated words from corpus, and synonym words from Machine Readable Dictionary (MRD). For English-Korean translation words, we first extract English sentiment-bearing words by using synonym information from English thesaurus and translate them into Korean sentiment-bearing words by using English-Korean Machine Readable Dictionary (MRD). For annotated words, we discriminate sentiment-bearing words from all the words in corpus by hand. Finally, we extract synonym words from MRD using the previous extracted sentiment-bearing words (translation words and annotated words) as queries.

In this paper, we recommend using another effective feature set, domain feature set, for sentiment classification. Although many previous studies have been focused on building up and using the global features such as [4,5], we found that some words, which are not global features, can have sentimental polarity in some domains. They are easily collected by a feature selection technique such as the statistics.

In experimental results, our sentiment classification system showed more improved performance when it used the combination of global and domain features.

This paper is organized as follows: Section 2 describes our approach to extract global and domain features in detail. Section 3 reports our experiments and results. In section 4, the conclusion is presented.

2. Acquiring Global and Domain Features as Sentiment-bearing Words

2.1 Extracting Global Features

There are some big differences between effective features in general topic classification and ones in sentiment classification. The one of them is that the important POS tags for topic classification are noun and verb, but ones for sentiment classification are adjective and adverb as well as noun and verb. In this paper, global features mean that they contain some sentimental polarities regardless of their occurrence in any domain. The sentiment classification basically needs to obtain effective sentiment-bearing words as global features. To collect the sentiment-bearing words, a thesaurus have been used usefully for English but a thesaurus for Korean is not available. Thus we decide to extract global features from various available resources such as thesaurus, MRD, and corpus. The following subsections will explain how to extract global features from each resource.
(1) English-Korean translation words

Since there is no available Korean thesaurus, we try to first extract English sentiment-bearing words from an English thesaurus, and then translate them into Korean sentiment-bearing ones by using an English-Korean MRD. Our process of extraction is as follows:

1. We start with a small set of manually-selected and annotated seed words (7 English words for positive and negative respectively). We choose these words from referring other study [6].

<table>
<thead>
<tr>
<th>Positive</th>
<th>good, correct, positive, excellent, nice, fortunate, superior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>bad, nasty, negative, poor, unfortunate, wrong, inferior</td>
</tr>
</tbody>
</table>

2. We expand each sentiment-bearing words set by adding synonyms of each sentiment-bearing word, which is previously chosen in the sentiment-bearing word set, from an English thesaurus. This expansion process is continued until the number of sentiment-bearing words is converged. Here, an English thesaurus is used, which is provided by dictionary service from Naver Potal Site¹.

3. All the words of collected English sentiment-bearing word sets are translated into Korean words using the English-Korean MRD. Finally, the noisy words of each sentiment-bearing word set are filtered out by hand.

We can finally build up the English-Korean translation features through this process.

(2) Annotated sentiment-bearing words from corpus

The English-Korean translation words cannot include any sentiment-bearing words which appear only in Korean. Thus we try to extract these sentiment-bearing words by an annotation task from a corpus by hand. For this, we first extract content words with noun, verb, adjective, and adverb, and annotate whether they have some sentiment polarities by two persons. If their annotation results are not same, they try to coincide in their opinions.

(3) Synonym words from a Korean MRD

Final sentiment-bearing words are from the synonym set of a Korean MRD. Since the above two methods provide us qualified sentiment-bearing words and the Korean MRD is available to us, we can easily collect synonyms of sentiment-bearing words which are previously selected by the previous two methods.

Table 2 shows the constitution of our global features. Since the number of negative words is over two times as many as that of positive words in translation words and annotated words, they need to be reduced for balancing the number of features. Thus we choose only negative words with more than 2 occurrences in translation words and more than 2 document frequencies in annotated words. The target sentiment-bearing words for synonym extraction are balanced feature set.

Table 2. The constitution of sentiment-bearing words as a global feature set

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation words</td>
<td>781</td>
<td>802 (1,834)</td>
</tr>
<tr>
<td>Annotated words</td>
<td>1,123</td>
<td>1,042 (2,155)</td>
</tr>
<tr>
<td>Synonym words</td>
<td>271</td>
<td>514</td>
</tr>
<tr>
<td>Total</td>
<td>2,175</td>
<td>2,358</td>
</tr>
</tbody>
</table>

* The numbers in parenthesis denote the number of words before normalization.

2.2 Extracting Domain Features

The sentiment-bearing words, which contain some sentimental polarity in any usage or domain, are collected as global features. They are very important features for sentiment classification because they give us fundamental information. However, there are other effective features for sentiment classification. They do not show sentimental polarity generally but they can become important sentiment-bearing words in any special domain. For example, if a movie has a bad reputation, its title or main character’s name can be a great negative feature. Since they depend on special application domains, they can be directly selected from training data by using a feature selection technique; these named entities are hard to be extracted from the Web or other general corpora [8]. The \( \chi^2 \) statistics technique is used in the study as feature selection. Using the two-way contingency table of a word \( t \) and a category \( c \), the word-goodness measure is defined as follows [9]:

\[
\text{Goodness of } t \text{ for } c = \frac{2 \times \text{Contingency Table} - (\text{Expected Counts})^2}{\text{Expected Counts}}
\]
\[
\chi^2(t, c) = \frac{N \times (AD - CB)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)}
\]

(1)

where \(i\) \(A\) is the number of times \(t\) and \(c\) co-occur, \(ii\) \(B\) is the number of times \(t\) occurs without \(c\), \(iii\) \(C\) is the number of times \(c\) occurs without \(t\), \(iv\) \(D\) is the number of times neither \(c\) nor \(t\) occurs, and \(vi\) \(N\) is the total number of documents.

To measure the goodness of a word through all the categories, we combine the category-specific scores of a word as follows [10]:

\[
\chi^2(t) = \max_{i=1}^{m} \left\{ \chi^2(t, c_i) \right\}
\]

(2)

where \(m\) denotes the number of categories.

After the goodness of words, which are not global features, is estimated by formula (1) and (2), they are ordered according to their estimated goodness score. The \(n\) words with high goodness score are chosen as domain features. The word number, \(n\), is determined by our experiment and the selected domain features as well as global features are used as important features in our sentiment classification.

3. Empirical Evaluation

3.1 Data Sets and Experimental Settings

A sentiment data set for experiments was constructed manually. It was collected from Web and it consists of 3 different categories (news articles, product reviews, and movie reviews). The total number of this data set is 2,479. Three people participated in classifying the sentiment of each document by voting mechanism for evaluation. Table 3 shows the distribution of each category in the data set.

<table>
<thead>
<tr>
<th>Category</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>News Article</td>
<td>417</td>
<td>312</td>
<td>729</td>
</tr>
<tr>
<td>Product Review</td>
<td>205</td>
<td>190</td>
<td>395</td>
</tr>
<tr>
<td>Movie Review</td>
<td>703</td>
<td>652</td>
<td>1,355</td>
</tr>
<tr>
<td>Total</td>
<td>1,325</td>
<td>1,154</td>
<td>2,479</td>
</tr>
</tbody>
</table>

Table 3. The category distribution of the data set

For fair evaluation, 5-fold cross validation was used and precision and recall are used as performance measures in this paper. Final performance is reported as a F1 measure score to
combine precision and recall scores. SVM is used as a classifier and TFIDF is used as term weighting in our experiments [11].

3.2 Experimental Results

3.2.1 Comparing the global features to traditional text classification features

Here, we try to verify our global features (GF) through experiments. The content words used in traditional text classification are baseline features to compare the proposed global features. There are two different types of content word sets: words with noun and verb POS tags and words with noun, verb, adjective, and adverb POS tags. The former one is denoted by NV (Noun and Verb content word) and latter one is by NVAA (Noun, Verb, Adjective, and Adverb content word). The experimental results are shown in the Table 4.

<table>
<thead>
<tr>
<th>Feature Types</th>
<th>NV</th>
<th>NVAA</th>
<th>GF</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 score</td>
<td>74.56</td>
<td>77.37</td>
<td>77.49</td>
</tr>
</tbody>
</table>

As you can see Table 4, an adjective and adverb word must be an effective feature because addition of them made significant improvement, nearly 3%. However, the usage of global feature did not lead to much improvement. We think that the reason is that global features are also included in the NVAA feature set; the NVAA feature set is always a superset of the global feature set at a training phase. But the acquisition of global features is very important to sentiment classification because they can provide good fundamental features regardless of their usage in any domain. To prove this fact, we did an additional experiment. Since our data set consists of 3 different domains (news article, product review, and movie review), it can be divided into two parts (the training set and the test set) by the hold-out procedure; only whole documents of one domain are chosen as test data and remaining documents are as training data. By this method, 3 different experimental settings are possible; the first case is that news articles are chosen as test data, the second one chose product reviews as test data, and the final one chose movie reviews as test data. Since the document of any test data does not have the same domain as training data, we can ascertain whether the global features can maintain their effectiveness when they are applied to new domain.
Table 5. The experimental results at each different test domain

<table>
<thead>
<tr>
<th>The Domain of Test Data</th>
<th>NVAA</th>
<th>GF</th>
</tr>
</thead>
<tbody>
<tr>
<td>News Article</td>
<td>63.57</td>
<td>64.69</td>
</tr>
<tr>
<td>Product Review</td>
<td>69.56</td>
<td>71.71</td>
</tr>
<tr>
<td>Movie Review</td>
<td>64.5</td>
<td>66.55</td>
</tr>
<tr>
<td>Average</td>
<td>65.87</td>
<td>67.65</td>
</tr>
</tbody>
</table>

The better F1 scores were obtained at all the domains when global features are used. It is a great proof that the global features are more effective in any new domain.

3.2.2 Observing the performance changes according to the number of domain features

We focus on the usefulness of domain features as well as global features in this paper. As we present in the previous 2.2 section, the goodness of domain features is measured by the statistics and they are ordered according to their goodness scores. As the top positioned n features are added to the global feature set, the combination feature set is constructed and it is our final feature set. To fix up n, we did experiments according to the number of added domain features. The results are shown in Figure 1.

Figure 1. The performance changes according to the number of added domain features
The best performance is obtained when 1,600 domain features are added to the global feature set. As the final F1 score is 78.97, it is relative improvement as high as 1.9% over that of using only the global feature set. The whole performance scores are summarized in Table 6.

Table 6. The comparison of performance scores at each feature type

<table>
<thead>
<tr>
<th>Feature Types</th>
<th>NV</th>
<th>NVAA</th>
<th>GF</th>
<th>Combined Features (GF+DF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 score</td>
<td>74.56</td>
<td>77.37</td>
<td>77.49</td>
<td>78.97</td>
</tr>
</tbody>
</table>

4. Conclusions

This paper presents how to extract and use the effective sentiment-bearing features. First, we built up the global feature set using thesaurus, MRD, and corpus, and extract domain features using a feature selection method, the statistics. As a result, the combination method of global features and domain features showed the best performance in our experiment. Therefore, the proposed combined features are more effective for sentiment classification.

5. Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. NRF-2013R1A1A2009937).

References


*Corresponding author: Youngjoong Ko, Ph.D.
Department of Computer Engineering,
Dong-A University,
840 Hadan 2-dong, Saha-gu, Busan, 604-714, Korea.
E-mail: youngjoong.ko@gmail.com