Modifying SO-PMI for Japanese Weblog Opinion Mining by Using a Balancing Factor and Detecting Neutral Expressions

Guangwei Wang
Graduate School of Information
Science and Technology
Hokkaido University
Sapporo, Japan 060-0814
wgw@media.eng.hokudai.ac.jp

Kenji Araki
Graduate School of Information
Science and Technology
Hokkaido University
Sapporo, Japan 060-0814
araki@media.eng.hokudai.ac.jp

Abstract

We propose a variation of the SO-PMI algorithm for Japanese, for use in Weblog Opinion Mining. SO-PMI is an unsupervised approach proposed by Turney that has been shown to work well for English. We first used the SO-PMI algorithm on Japanese in a way very similar to Turney's original idea. The result of this trial leaned heavily toward positive opinions. We then expanded the reference words to be sets of words, tried to introduce a balancing factor and to detect neutral expressions. After these modifications, we achieved a well-balanced result: both positive and negative accuracy exceeded 70%. This shows that our proposed approach not only adapted the SO-PMI for Japanese, but also modified it to analyze Japanese opinions more effectively.

1 Introduction

Recently, more and more websites add information in the form of personal opinions to the Web, e.g. customer reviews of products, forums, discussion groups, and blogs. Here, we use the term Weblog for these sites. This type of information is often useful. However, we have to deal with an enormous amount of unstructured and/or semi-structured data. These data are subjective, in free format and mostly textual, thus using them is difficult and time consuming. Therefore, how to mine the Weblog opinions automatically more effectively has attracted more and more attention (Gamon, 2005; Popescu, 2005; Chaovalit, 2005).

Turney (2002) has presented an unsupervised opinion classification algorithm called SO-PMI (Semantic Orientation Using Pointwise Mutual Information). The main use of SO-PMI is to estimate the semantic orientation (i.e. positive or negative) of a phrase by measuring the hits returned from a search engine of pairs of words or phrases, based on the mutual information theory. This approach has previously been successfully used on English. The average accuracy was 74% when evaluated on 410 reviews from Epinions[1].

However, according to our preliminary experiment, directly translating Turney's original idea into Japanese gave a very slanted result, with a *positive accuracy* of 95% and a *negative accuracy* of only 8%. We found that the balance between the positive and negative sides is influenced greatly by the page hits of reference words/sets, since a search engine is used. Therefore, we introduced a balancing factor according for the difference in occurrence between positive and negative words. And then we added several threshold rules to detect neutral expressions. The proposed approach is evaluated on 200 positive and 200 negative Japanese opinion sentences and yielded a well-balanced result.

In the remainder of this paper, we review the SO-PMI Algorithm in Section 2, then adapt the SO-PMI for Japanese and present the modifications in Section 3. In section 4, we evaluate and discuss the experimental results. Section 5 gives concluding remarks.

2 Details of the SO-PMI Algorithm

The SO-PMI algorithm (Turney, 2002) is used to estimate the semantic orientation (SO) of a phrase by

[1] http://www.epinions.com

Proceedings of NAACL HLT 2007, Companion Volume, pages 189–192,
Rochester, NY, April 2007.

measuring the similarity of pairs of words or phrases using the following formula:

$$PMI(word_1, word_2) = log_2\left[\frac{p(word_1 \& word_2)}{p(word_1)p(word_2)}\right] \quad (1)$$

$$SO(phrase) = PMI(phrase, \text{“}excellent\text{”}) - PMI(phrase, \text{“}poor\text{”}) \quad (2)$$

The reference words “excellent” and “poor” are used, thus SO is positive when a phrase is more strongly associated with “excellent” and negative when a phrase is more strongly associated with “poor”. Let hits(query) be the number of hits returned when using a search engine, the following estimate of SO can be derived from Formula (2) and (1) with some minor algebraic manipulation.

$$SO(phrase) = log_2[A]$$
$$A = \frac{hits(phrase\ NEAR\ \text{“}excellent\text{”}) * hits(\text{“}poor\text{”})}{hits(phrase\ NEAR\ \text{“}poor\text{”}) * hits(\text{“}excellent\text{”})} \quad (3)$$

Turney used AltaVista[2] search engine because it has a NEAR operator. This operator constrains the search to documents that contain the words within ten words of one another, in either order. Turney’s previous work has shown that NEAR performs better than AND when measuring the strength of semantic association between words.

3 Our Proposed Approach

The first step of our approach is to extract opinion phrases using word POS (part of speech) templates based on our analysis of opinions in Japanese Weblog and the results of related work (Kobayashi, 2003; Taku, 2002; Wang, 2006). The second step is to estimate the semantic orientation of the extracted phrases, using the SO-PMI algorithm.

3.1 Adapting SO-PMI for Japanese

Following Turney’s original idea, we first translated the SO formula to the one shown in Formula (4) for Japanese.

$$SO(phrase) = log_2[B] \quad (4)$$
$$B = \left[\frac{hits(phrase\ \ AND\ \ \text{"すばらしい"}) * hits(\text{"不良"})}{hits(phrase\ \ AND\ \ \text{"不良"}) * hits(\text{"すばらしい"})}\right]$$

We used the Google search engine[3] to get the hits(query) even though Google does not have a NEAR operator. The AltaVista NEAR operator does not work well for Japanese and Google indexes more pages than AltaVista, thus we used Google and replaced the NEAR operator with the AND operator in the SO formula. “すばらしい” and “不良” were selected because they correspond to the English words “excellent” and “poor”.

For testing the performance of this trial, we used 200 positive and 200 negative Japanese opinion sentences which have been labeled by hand. The results were very slanted. Many phrases, whether positive or negative in meaning, still received a positive SO. Some possible causes could be that “不良 (poor)” has more hits than “すばらしい (excellent)”, as shown in Table 1, and that the AND operator is less useful than the NEAR operator.

3.2 Modifying SO-PMI for Japanese

In Japanese, there are many expressions when people evaluate something. For example, “いい (good)”, “良い (good)”, “満足 (satisfaction)” , “すばらしい (excellent)” are usually used when someone wants to convey a positive opinion. Hence we tried to replace the reference words “excellent” and “poor” with two reference sets: “p_basic” and “n_basic”:

$$SO(phrase) = log_2[C]$$
$$C = \frac{hits(phrase\ AND p\_basic) * hits(n\_basic)}{hits(phrase\ AND n\_basic) * hits(p\_basic)} \quad (5)$$

“p_basic” is a set of common strong positive words in Japanese. “n_basic” is a set of common weak negative words. The hit counts of these words from Google is shown in Table 1 (All data from 2007/01/12). The hits(query) was calculated by *hits(phrase AND (“いい (good)” OR “好き (like)”) OR “良い (good)” OR ...)*.

Table 1: Frequency of p_basic/n_basic words on the Web

p_basic words	Hits (K)	R(%)	n_basic words	Hits (K)	R(%)
いい(good)	372,000	36.83	不良(poor)	119,000	11.78
好き(like)	242,000	23.96	悪い(bad)	110,000	10.89
良い(good)	211,000	20.89	不安(worry)	83,000	8.22
魅力(charm)	150,000	14.85	欠点(fault)	77,900	7.71
大好き(favorite)	115,000	11.39	難しい(hard)	77,600	7.68
欲しい(want)	115,000	11.39	嫌(dislike)	65,000	6.44
楽しい(delightful)	107,000	10.59	あんまり(not good)	37,900	3.75
よい(good)	103,000	10.20	嫌い(dislike)	37,100	3.67
良く(good)	96,900	9.59	だめ(useless)	26,500	2.62
満足(satisfaction)	80,600	7.98	辛い(painful)	26,500	2.62
面白い(interesting)	79,700	7.89	不快(dissatisfaction)	26,100	2.58
嬉しい(happy)	75,500	7.48	不満(dissatisfaction)	22,200	2.20
素敵(lovely)	74,700	7.40	最悪(worst)	20,700	2.05
うれしい(happy)	59,500	5.89	不具合 (fault)	16,600	1.64
おもしろい(interesting)	28,400	2.81	あまり(not good)	15,600	1.54
すばらしい(excellent)	26,000	2.57	まずい(bad)	10,300	1.02

We evaluated this modification using the same

[2] http://www.altavista.com/sites/search/adv

[3] http://www.google.co.jp/

data as in Section 3.1. We obtained a slightly better result. However the SO values were still slanted. This time many phrases, whether positive or negative in meaning, still received a negative SO. All of these test results are shown in detail in Section 4.2.

In the experiments above, we obtained heavily slanted results. We consider that the large difference in page hits between the positive and negative reference words/sets are the main cause for this phenomenon. To mitigate this problem, we decided to introduce a balancing factor to adjust the balance between the positive and negative sides. The SO formula was modified from (5) to (6).

$$SO(phrase) = log_2 [C] + f(\alpha) \quad (6)$$

The balancing factor $f(\alpha)$ was calculated by Formula (7).

$$f(\alpha) = \alpha * log_2 \left[\frac{hits(p\_basic)}{hits(n\_basic)} \right] \quad (7)$$

The log_2 of "p_basic" and "n_basic" is a factor that adjusts the balance of the similarity of "p_basic"/"n_basic" and phrases automatically by the hits of "p_basic"/"n_basic" itself. α is a weight value. We evaluated different values of α from "0.0" to "1.0" on the benchmark dataset, which is shown in detail in Section 4.2.

From these preliminary trials, we also found that many neutral phrases often receive positive or negative SO. Therefore we added detection of neutral expressions. The idea is that if the phrase is strongly or faintly associated with both "p_basic" and "n_basic", it is considered a neutral phrase. Because this means that this phrase has an ambiguous connection with both "p_basic" and "n_basic". We use the following rules (Figure 1) to separate neutral phrases from positive/negative phrases. The threshold values **ta, tb** and **tc** are obtained from a small, hand-labeled corpus.

*1. hits(phrase AND p_basic) > **ta** AND hits(phrase AND n_basic) > **ta***
*2. hits(phrase AND p_basic) < **tb** AND hits(phrase AND n_basic) < **tb***
*3. | hits(phrase AND p_basic) – hits(phrase AND n_basic) | < **tc***
4. SO(phrase) = 0

Figure 1: Rules for Detecting Neutral Expressions

4 Experimental Performance Evaluation

4.1 Gold Standard and Evaluation Metrics

As a gold standard, we collected a benchmark dataset which has 200 positive opinion sentences and 200 negative opinion sentences from the reviews about Electronic Dictionary and MP3 Player products that have been labeled as either positive or negative reviews in "Kakaku.com"[4]. "Kakaku.com" is the largest Japanese Weblog specializing in product comparison of consumer goods, including price and user opinions, etc. Lots of people exchange miscellaneous product information and reviews. These reviews are classified as questions, positive reviews, negative reviews, rumors, sale information or "other" category.

To classify a sentence as positive (P) or negative (N), the average SO of the phrases in the sentence is used. If the average SO is P, the sentence is a positive sentence; otherwise it is a negative sentence. As evaluation metrics, we measured our proposed approach's performance by *accuracy*. *accuracy* was measured as the number of sentences correctly classified as P/N sentences to the total number of P/N sentences in the benchmark dataset (200). **PA** means *positive accuracy*, **NA** means *negative accuracy*, i.e. the accuracy on only positive or negative sentences respectively.

4.2 Experiments and Results

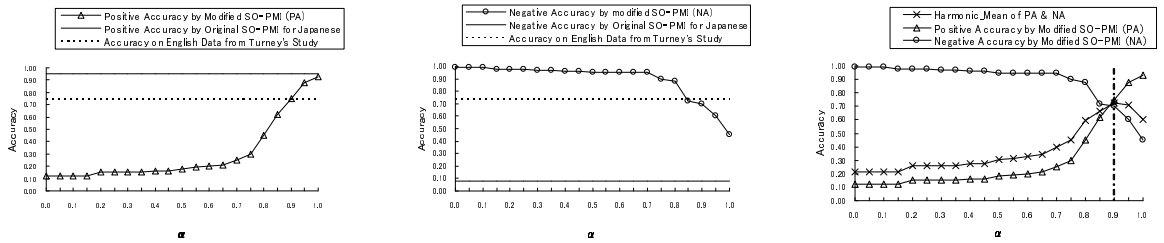
First we did the balancing factor experiment to determine the value of "α", using the benchmark dataset. The results are shown in Figure 2. (a) and (b) show the dashed line indicates average accuracy (74%) on English Data from Turney's Study (2002). Turney didn't evaluate positive and negative accuracy respectively. The full drawn line indicates the result after translating the original SO-PMI to Japanese (PA:95%, NA: 8%). **PA** series (the line with triangle mark)/**NA** series (the line with circle mark) when values of "α" from "0.0" to "1.0" were used.

Changing the α tends to be a tradeoff, lowering **PA** when **NA** is improved and vice versa. Therefore, we used $Harmonic\_Mean$ by the following formula to find a proper value of "α".

$$Harmonic\_Mean = \frac{2 * PA * NA}{PA + NA} \quad (8)$$

Figure 2, (c) shows **PA**, **NA** and $Harmonic\_Mean$ curves for different values

[4] http://www.kakaku.com/



(a) Positive Accuracy (PA) (b) Negative Accuracy (NA) (c) Harmonic-Mean of PA/NA

Figure 2: Experiment for α in Balance Factor

of "α". We selected the "α=0.9" giving the highest $Harmonic\_Mean$ value, thus giving a good balance between **PA** (75%) and **NA** (70%).

The comparative experiment results between the SO-PMI for Japanese (Test 1), and our modifications (Test 2, 3, 4) are shown in Table 2.

Table 2: Comparative Experiment Results

	Test Content	PA(%)	NA(%)
Test 1	Naive translation of Turney's Approach for Japanese	95	8
Test 2	Modification 1: Two Reference Sets	12	99
Test 3	Test 2 + Modification 2: Balancing Factor [α = 0.9]	75	70
Test 4	Test 3 + Modification 3: Neutral Phrase Detection	**78**	**72**

PA: Positive Accuracy NA: Negative Accuracy

In Test 1 and 2, we obtained extreme results, leaning to the positive or negative end, whether using the Turney's original approach or expanding the reference word as "p_basic" and "n_basic". In Test 3, we added a balancing factor as described in section 3.2, and obtained a comparatively well-balanced result. Finally, after adding the neutral expressions detection, we achieved a **PA** of 78% and **NA** of 72% (Test 4). The balance between positive and negative sides was quite improved by contrast with Test 1 and 2.

5 Conclusions

This study first proposed a modified unsupervised approach (SO-PMI) for Japanese Weblog Opinion Mining. Some parts of Turney's approach, such as the NEAR operator, does not work for Japanese, thus some modifications must be done. In a preliminary experiment, the *negative accuracy* (8%) was very poor while the *positive accuracy* (95%) was high. To deal with this phenomenon, we presented three modifications based on the characteristics of Japanese and the results of related work. The experiment results (*positive accuracy*: 78%, *negative accuracy*: 72%) show that our proposal achieved a considerably improved performance, comparing with directly translating the SO-PMI. Hence it would be expected that the balancing factor and neutral expressions detection would work effectively also for other reference words or languages. In the future, we will evaluate different choices of words for the sets of positive and negative reference words. We also plan to appraise our proposal on other languages.

References

Peter D. Turney. 2002. *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. Proceedings 40th Annual Meeting of the ACL, pp. 417-424.

Popescu, Ana-Maria, and Oren Etzioni. 2005. *Extracting Product Features and Opinions from Reviews*. Proceedings of HLT-EMNLP.

Michael Gamon, Anthony Aue, Simon Corston-Oliver and Eric K. Ringger. 2005. *Pulse: Mining Customer Opinions from Free Text*. Proceedings of the 2005 Conference on Intelligent Data Analysis (IDA), pp.121-132.

Pimwadee Chaovalit and Lina Zhou. 2005. *Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches*. Proceedings of the 38th Annual HICSS.

Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi and Toshikazu Fukushima. 2003. *Collecting evaluative expressions by a text mining technique*. IPSJ SIG NOTE, Vol.154, No.12, In Japanese.

Taku Kudoh and Yuji Matsumoto. 2002. *Applying Cascaded Chunking to Japanese Dependency Structure Analysis*. Information Processing Society of Japan (IPSJ)Academic Journals, Vol 43, No 6, pp. 1834-1842, In Japanese.

Guangwei Wang and Kenji Araki. 2006. *A Decision Support System Using Text Mining Technology*. IEICE SIG Notes WI2-2006-6, pp. 55-56.