An Opinion Detection and Classification System Using Support Vector Machines

Hiroya Ozawa Hiroya Susuki Hiroaki Saito
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa, Japan
{ozawa, susuki, hxs}@nak.ics.keio.ac.jp

Abstract

We developed an opinion detection and polarity classification system for Japanese newspapers at NTCIR-7 MOAT task. Our system detects sentences which are "opinionated" or "not opinionated" and classifies them into "positive", "negative" or "neutral". We used Support Vector Machines (SVM) as a machine learning method. To determine features, we focused on the end expression, some particular structure of opinionated sentences, and continuity of opinion.

In the formal run, the opinion detection subtask attained precision 81.15%, recall 34.16%, F-measure 48.08, and the polarity classification subtask attained precision 48.05%, recall 18.01%, F-measure 26.20.

Keywords: *Multilingual Opinion Analysis Task (MOAT), Opinionated Sentence, Opinion Detection, Opinion Classification, Polarity.*

1 Introduction

In recent years, an opinion analysis task becomes more important in the field of natural language processing. There are various opinions in web sites. By analyzing them, we are able to know public opinions or users' opinions about certain products.

We developed an opinion detection and polarity classification system for Japanese newspaper at NTCIR-7. Our system detects sentences which are "opinionated" or "not opinionated" and classifies them into "positive", "negative" or "neutral".

The following shows examples of opinionated not opinionated sentences.

Opinionated 私はここに住んでいる個人の気持ちを、市場調査をしてみるべきだと思います。
(I think you should investigate feelings of people living here.)

Not Opinionated ベルギー同様、湾岸戦争に参加しなかったイタリアにとっても劣化ウラン弾は遠い存在だった。
(Italy that did not participate in the Gulf War has been far away from depleted uranium ammunitions as well as Belgium.)

And the following shows example sentences of polarity.

Positive そのこと自体は何ら問題はないと思います。
(I think this is no problem.)

Negative そういう事実を認識しないのはよくないと思います。
(I do not think it is good that you do not recognize these facts.)

Neutral 米軍は、すべての情報を公開すべきだ。
(US military forces should reveal all the information.)

We used Support Vector Machines as a machine learning method. To determine features, we focused on the end expression, some particular structure of opinionated sentences and continuity of opinion. We used the same features for the opinion detection subtask and the polarity classification subtask.

We will describe analysis of data sets in Section 2, the outline of our system in Section 3, experiments and results in Section 4, and the conclusion in Section 5.

2 Analysis of data sets

First, we manually analyzed data sets (NTCIR6 opinion task test collection and NTCIR7 sample data) to determine features of the opinion detection subtask and found the following aspects.

2.1 End expression

The Japanese language has subject-object-verb (SOV) structure, namely a verb phrase appears at the end of a sentence. Thus the end expression of a sentence is important for the opinion detection subtask [1] and the end expression was often used in recent researches [1] [2].

In our system, we use only auxiliary verbs at the end of a sentence as a feature (ex. "べきだ"; "should" in English) because auxiliary verbs tend to express opinionatedness. Table 1 and 2 show the top 20 of the op_score in opinionated and not opinionated sentences. The op_score is calculated by the following formula.

$$\text{op\_score} = \frac{\text{fav}(o)}{O} - \frac{\text{fav}(no)}{NO},$$

where $\text{fav}(o)$ shows the frequency of auxiliary verbs in opinionated sentences, O shows the number of opinionated sentences, $\text{fav}(no)$ shows frequency of auxiliary verbs in not opinionated sentences, NO shows the number of not opinionated sentences. The op_score means the difference between the ratio of auxiliary verbs in opinionated sentence and the ratio of auxiliary verbs in not opinionated sentence.

Table 1. Auxiliary verbs at the end of opinionated sentences and their op_scores (top20)

Opinionated	op_score
ない	0.091611
だ	0.081080
だろう	0.028399
たい	0.015811
べきだ	0.012995
ます	0.012776
であり	0.011570
なら	0.008292
です	0.007567
である	0.007371
なく	0.007206
ないだろう	0.005857
で	0.005312
う	0.004332
ません	0.003087
でしょう	0.002828
べきである	0.002659
でなく	0.002505
ず	0.002424
ただろう	0.002154

Table 1 and 2 indicate an auxiliary verb will be an effective feature.

Table 2. Auxiliary verbs at the end of not opinionated sentences and their op_scores (bottom20)

Not opinionated	op_score
た	−0.085723
だった	−0.002047
なかった	−0.001628
ました	−0.001295
であった	−0.000907
ん	−0.000616
る	−0.000370
つ	−0.000370
ませんでした	−0.000179
でなかったら	−0.000123
べきでなかった	−0.000123
や	−0.000123
ないやん	−0.000123
ないらしい	−0.000123
じ	−0.000056
らしい	0.000034
べきで	0.000034
べき	0.000034
ございません	0.000034
べく	0.000034

2.2 A particular structure in opinionated sentences

Opinionated sentences tend to have a particular structure. For example, a speaker's name is expressed after *kagikakko*[1]

> 「長いジュースを取れず、これが勝負の分かれ目だった」(杉山)
> ("I could not dominate a long deuce, so I lost that match." (*Sugiyama*))

This structure appears 1.8% in opinionated and 1.2% in not opinionated sentences.

Postposition "と" is often inserted after *kagikakko*.

> 同社は「コンタクトレンズのメーカーなので、特に培養角膜の研究に力を入れたい」と話している。
> (This company says, "We will make a lot of effort in cultivated corneas because we manufacture contact lenses".)

This structure appears 16.2% in opinionated and 8.7% in not opinionated sentences.

Therefore, to use these structure as a feature, we transformed sentences by the following rules.

[1] The Japanese quote symbol, "「" and "」".

- auxiliary verbs, postpositions and symbols → the original form is kept (not transformed)
- other part of speech (POS) → POS-tag

We kept the original form of auxiliary verbs because they will be effective from the result of Section2.1.

We used 2-gram of these words as features. For example,

> 同社は「コンタクトレンズのメーカーなので、特に培養角膜の研究に力を入れたい」と話している。

This sentence is transformed into the following:

> [Noun] [は] [「] [Noun] [の] [Noun] [なので] [、] [Adverb] [Noun] [の] [Noun] [に] [Noun] [を] [Verb] [たい] [」] [と] [Verb] [て] [Verb] [。]

In this example, [Noun は], [は「], [「Noun], [Noun の], etc are 2-grams.

2.3 Continuity of opinionated sentences

Opinionated sentences tend to appear continuously. Mizuguchi et al. used "the previous sentence is opinionated or not" as a feature [2]. In addition, we used "2nd previous sentence is opinionated or not" as a feature.

We analyzed the corpus in terms of continuity of opinionated sentences in NTCIR-7 sample data and NTCIR-6 data. The number of sentences is 32961; total of opinionated and not opinionated sentences [2]. And the number of opinionated sentence is 8,626, the number of sentences whose previous sentence is opinionated is 4,207, the number of sentences whose 2nd previous sentence is opinionated is 3,483.

Suppose opinionated sentences appear randomly (random data), probability of the previous sentence being opinionated is 0.262 by $8,626/32,961$, the probability of the 2nd previous sentence being opinionated is 0.262, and probability of both previous and 2nd previous sentences being opinionated is 0.068 by $(8,626/32,961)^2$.

On the other hand, according to the continuity of opinionated sentences (NTCIR data), probability of the previous sentence being opinionated is 0.488 by $4,207/8,626$, probability of the 2nd previous sentence being opinionated is 0.404 by $3,483/8,626$, and probability of both previous and 2nd previous sentences being opinionated is 0.256 by $2,204/8,626$. Table 3 summarizes these results.

Therefore, whether the previous / 2nd previous sentences are opinionated or not will be an effective feature.

[2] However, these sentences annotated by three annotators, we treat these as different sentences. Thus the number of distinct sentences is 10,987.

Table 3. Probability of the previous / 2nd previous / both sentences are opinionated

	random data	NTCIR data
Previous sentences	0.262	0.488
2nd previous sentences	0.262	0.404
Both	0.068	0.256

2.4 Category of characters

How often a certain category of characters (*hiragana*, *katakana*, *kanji* and others) is used is different between opinionated and not opinionated sentences. For example, not opinionated sentences tend to be more explanatory than opinionated sentences, thus not opinionated sentences have more percentage of *kanji*. Table 4 shows this.

Table 4. Percentage of character category

	Opinionated	Not Opinionated
Hiragana / Katakana	0.512	0.434
Kanji	0.388	0.424
Others	0.099	0.142

Therefore, the category of characters will be an effective feature.

3 Our system

We used SVMlight[3] and TinySVM[4] to classify sentences and we used MeCab[5] as a Japanese morphological analyzer. Figure 1 and 2 show the structure of our system.

3.1 SVM classifier

Our system uses SVMlight in the opinion detection subtask and TinySVM in the polarity classification subtask as support vector machines, because SVMlight attained a better result in the opinion detection subtask and TinySVM showed a better result in the polarity classification subtask.

[3] http://svmlight.joachims.org/
[4] http://chasen.org/∼taku/software/TinySVM/
[5] http://mecab.sourceforge.net/

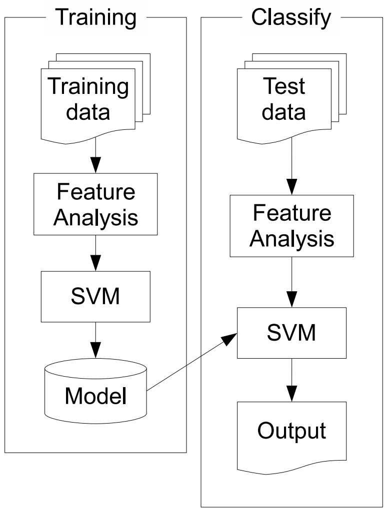

Figure 1. Opinion detection

In the polarity classification subtask, we use three SVMs adapted to the classification of three types. These SVMs classified positive or not, negative or not, neutral or not respectively, and our system selected the polarity that produced the biggest value as output.

3.2 Features

Based on the analysis in Section 2, we determined features as follows:

1. All part of speech (baseline)
2. All POS-tags (baseline)
3. Auxiliary verbs at the end of a sentence (Section 2.1)
4. POS-tag of the end of a sentence (supplement of feature 3)
5. Particular structure of opinionated sentences : transformed 2-gram (Section 2.2)
6. Category of characters (Section 2.4)
7. The previous sentence is opinionated or not (Section 2.3)
8. The 2nd previous sentence is opinionated or not (Section 2.3)

We set the system with features 1, 2 as a baseline. Features 4, 5, 6 are similar features in previous research. Feature 4 uses auxiliary verbs, changed from previous research using all part of speech at the end of a sentence.

We use features 1,2,3,4,5,6 to classify the polarity because we investigate how these features can classify the polarity of sentences.

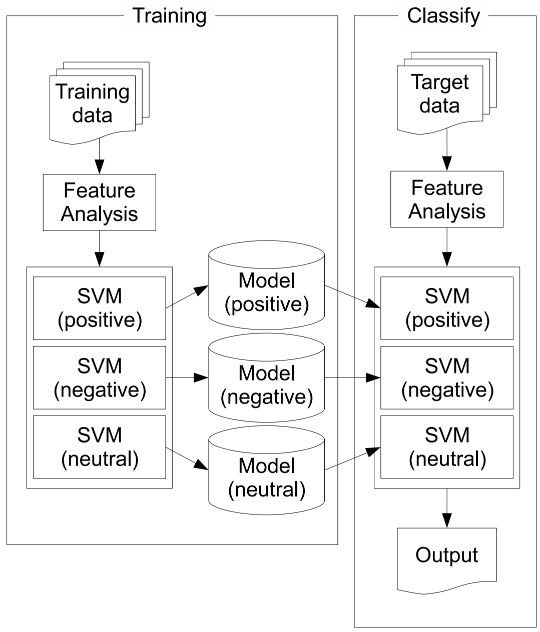

Figure 2. Polarity classification

4 Evaluation Results

This section shows the experiments to verify features, and results of the formal run.

4.1 Formal Run

We used NTCIR-7 sample data and NTCIR-6 data as training data which consists of 8,626 opinionated sentences, 24,335 not opinionated sentences on the opinion detection subtask. And we used same data as training data (Table 5) on the polarity classification subtask.

Table 5. The number of positive / negative examples in the training data of the polarity classification subtask

	positive examples	negative examples
Positive or not	1,758	7,077
Negative or not	3,834	5,001
Neutral or not	3,155	5,680

We submitted three runs as the formal run; the highest priority, the second priority, the lowest priority. The following shows these three runs.

run1 Using all features as our method

run2 Using features 1, 2, 4, 5, 7 as previous research.

run3 Using features 1, 2 as baseline

Table 6 and Table 7 shows the results [3].

Table 6. Opinion detection results

	Precision	Recall	F-measure
Run1	81.15	34.16	48.08
Run2	78.86	30.92	44.42
Run3	78.13	36.33	49.60

Table 7. Polarity classification results

	Precision	Recall	F-measure
Run1	48.05	18.01	26.20
Run2	48.63	16.70	24.86
Run3	48.31	19.36	27.64

In all results, our method attained higher precision than baseline, but lower recall than the baseline. We believe some features are not effective to the formal run data.

However, our system had some bugs, thus if we correct them, these results will change.

4.2 Effective features in the opinion detection subtask

We investigated features' effectiveness of the features in Section 3.1 for the opinion detection subtask using the formal run data. Training data is the same of Section 4.1 and the formal run data is used for the test.

We made comparison between a result of using all features and results of deleting one or two features (feature 1, 2, 3, 4, 5, 6, 7, 8, 7 and 8). We combined feature 7 and 8 because they are both indicate opinion continuity. Table 8 shows deleted features, each result and difference of F-measure. Larger difference means more effective feature. Thus table 8 shows features 7, 8 are not effective and features 1,5 are more effective on the formal run data.

Next, we evaluated features 1, 2, 3, 4, 5, 6 without features 7, 8. Table 9 shows features 1, 5 are more effective too.

According to these results, we got the best result using features 1, 2, 3, 5, 6.

4.3 Effective features in the polarity classification subtask

Next, we made experiments to investigate features' effectiveness for the polarity classification subtask using the formal run data too. According to Section 3, we used three SVMs to classify the polarity. Training data is the same of Section 4.1 and the formal run data is used for the test.

We made comparison between the result of using all features and results of deleting each feature (features 1, 2, 3, 4, 5, 6). Table 10 shows feature 4 are not effective. Therefore, we got the best result using features 1,

Table 8. Feature evaluation in the opinion detection subtask

deleted feature	P	R	F	diff
none	79.95	34.45	48.15	
feature 1	81.05	22.63	35.38	12.77
feature 2	79.70	33.92	47.59	0.56
feature 3	80.28	33.98	47.75	0.40
feature 4	81.55	33.51	47.50	0.65
feature 5	77.53	31.04	44.33	3.82
feature 6	80.56	32.39	46.20	1.95
feature 7	82.20	36.39	50.45	−2.30
feature 8	81.78	35.10	49.12	−0.97
feature 7,8	82.87	38.39	52.47	−4.32

Table 9. Feature1–6 evaluation in the opinion detection subtask

deleted feature	P	R	F	diff
feature 7,8	82.87	38.39	52.47	
feature 1,7,8	82.78	27.69	41.50	10.97
feature 2,7,8	82.55	38.10	52.14	0.33
feature 3,7,8	82.54	37.80	51.85	0.62
feature 4,7,8	82.98	38.68	52.76	−0.29
feature 5,7,8	82.06	36.04	50.08	2.39
feature 6,7,8	82.82	37.98	52.08	0.39

Table 10. Feature1–6 evaluation in the polarity classification subtask

deleted feature	P	R	F	diff
none	49.62	20.41	28.92	
feature 1	47.34	13.94	21.54	7.38
feature 2	49.74	20.30	28.83	0.09
feature 3	48.85	19.89	28.27	0.65
feature 4	49.56	20.46	28.96	−0.04
feature 5	50.53	19.73	28.38	0.54
feature 6	49.29	20.04	28.49	0.43

2, 3, 5, 6.

5 Conclusion

We developed an opinion detection and polarity classification system for Japanese newspapers at NTCIR-7 MOAT task.

In the formal run, the opinion detection subtask attained precision 81.15%, recall 34.16%, F-measure 48.08, and the polarity classification subtask attained precision 48.05%, recall 18.01%, F-measure 26.20.

In the post formal run analysis, the best results in the opinion detection subtask was precision 82.98%, recall 38.68%, F-measure 52.76 using all part of speech (feature 1), all POS-tags (feature 2), auxiliary verbs at the end of a sentence (feature 3), particular structure of opinionated sentences : transformed 2-gram (feature 5), category of characters (feature 6). The best results in the polarity classification subtask was precision 49.56%, recall 20.46%, F-measure 28.96 using all part of speech (feature 1), all POS-tags (feature 2), auxiliary verbs at the end of a sentence (feature 3), particular structure of opinionated sentences : transformed 2-gram (feature 5), category of characters (feature 6).

References

[1] Toshiyuki Kanamaru, Masaki Murata, Hitoshi Isahara, “Japanese Opinion Extraction System for Japanese Newspaper Using Machine-Learning Method,” *In Proceedings of the sixth NTCIR Workshop*, pp 301-307, 2007.

[2] Hironori Mizuguchi, Masaaki Tsuchida, Dai Kusui, “Three-Phase Opinion Analysis System at NTCIR-6,” In *Proceedings of the sixth NTCIR Workshop*, pp 330-335, 2007.

[3] Yohei Seki, David Kirk Evans, Lun-Wei Ku, “Overview of Multilingual Opinion Analysis Task at NTCIR-7”, 2008.