Enhance Japanese Opinionated Sentence Identification using Linguistic Features: Experiences of the IISR Group at NTCIR-8 MOAT Task

Guo-Hau Lai, Jyun-Wei Huang, Chia-Pei Gao, Richard Tzong-Han Tsai
Department of Computer Science and Engineering, Yuan Ze University
135, Far-East Rd. ,Chung-Li, Taoyuan, Taiwan, ROC
+886-3-4638800 #2367 & #7062
{s986005, s976017}@mail.yzu.edu.tw, {chia, thtsai}@saturn.yzu.edu.tw

ABSTRACT
This paper presents our work in NTCIR-8 workshop of Multilingual Opinion Analysis Task (MOAT). We describe a feature-based system that is designed to detect the opinion sentence or not. The system utilizes various features: headlines in newspapers, Japanese sentence patterns, dependency pairs, numeral features and some related to newspapers opinionated words. The experiments show that our feature-based system is feasible and effective.

Categories and Subject Descriptors
H.3.1 [Information Storage and Retrieval]: linguistic processing

General Terms
Experimentation

Keywords
Multilingual Opinion Analysis Task (MOAT), Opinionated Sentence Detection, Japanese Newspapers Features

1. INTRODUCTION
In recent years, more and more websites add information in the form of opinions. Users can now post reviews of products at e-commerce sites and express their views on almost anything in Internet forums, discussion groups, and blogs. By extracting and analyzing them, we are able to know public opinions or personal opinions about specific products.

In order to solve these problems, the sentiment analysis has become a very important research field recently. The sentiment analysis uses the theories and techniques, such as artificial intelligence, text mining, natural language processing and information retrieval. However, the opinionated sentence detection is one of the major tasks in sentiment analysis.

NTCIR-8 Multilingual Opinion Analysis Task (MOAT) [1] is also focused on the subjectivity and sentiment analysis in newspaper article with various subtasks based on the past NTCIR workshop [2, 3]. In this year, the organizers defined a new subtask: cross-lingual subtask to evaluate answer opinion extraction accuracy in the languages different from the question.

As the first year we participated, we decided to join the opinion judgment subtask: which required a binary classification of sentences for subjectivity, determining whether the sentence contains opinions or not. Our system takes the form of a general feature-based opinion detection system. The system utilizes various features: headlines in newspapers, Japanese sentence patterns, dependency pairs, numeral features and some related to newspapers opinionated words.

In this paper, we describe and evaluate a feature-based system. In Section 2, we discuss the Japanese newspapers features. Section 3 exposes our approach to detect the opinion sentences. Section 4 presents evaluation result of our approach, based on these features of Japanese newspapers. Finally, the conclusions are given in Section 5.

2. DATASET ANALYSIS
We analyzed NTCIR7 and NTCIR8 sample data (contains 1,601 sentences in total and 499 opinion sentences) to find useful features for opinion judgment subtask. We founded these features as follow.

2.1 Headline
The headlines or titles are usually brief and sometimes sentence-final expressions will be omitted, for example, “◇米國は因果関係否定 (United States denied the causal relationship)” and “◇各國がテロ非難 (Nations have expressed condemnation of terrorist acts)”. Therefore, we used some special structures and opinion words for opinion detection in headlines. First, we defined these headlines or titles by using following symbols: “★□■◇◆●○▲▽ [] 【】” and the rate of opinion sentences was about 20 percent. Then, we analyzed these headlines only opinionated and summarized as follow.

1. A person says something in the headline
   ◇豪州人も標的に−−松野明久・大阪外大助教授（インドネシア専攻）の話
   (◇Australians are also targets−−Akihisa Matsuno (A professor of Osaka University of Foreign Studies / research Indonesia's Affairs))
2. Opinion words in the headline
   ◇湾岸戦争で使われた劣化ウラン弾と「湾岸症候群」との関係を医学的に証明するのはまだ難しい。
   (◇Medically explain the relationship between the depleted uranium ammunition used in Gulf War and the Gulf War syndrome is still difficult.)

In the sample 2, where “難しい (difficult)” is the opinion word.

2.2 Sentence Patterns
The Japanese sentence pattern is the sentence structure in Japanese that includes word order (Japanese is SOV structure),

tense and expletive. By understanding the Japanese sentence patterns, reader can rapidly understand what the content that author intended to convey is. Since Japanese newspaper is a formal article, the sentence structure is more regular and complete than a normal article, using the sentence pattern can easily achieve our goal. We analyzed the sample data provided by organizer and observed that some sentence patterns were frequently used in opinion sentences. For example:

1. Sentence pattern appears within a sentence
   - について
     この兵士の死とウラン弾の因果関係について、イタリア国防省は否定した。
     (The Italian Ministry of Defense denied that this soldier's death related to the uranium ammunition.)
2. Sentence pattern appears in the end of a sentence
   - そうだ
     パーキンソン病治療や移植医療に貢献しそうだ。
     (Its contribute to both treatment of Parkinson's disease and organ transplant.)
   - べきだ
     米軍は、すべての情報を公開すべきだ。
     (The U.S. military should make public all information.)

2.3 Keywords

We first examined the frequency of each word in opinion sentences and non-opinion sentences, and then selected a word list for opinion detection. For counting the frequency of each word, we used the Japanese morphological tool, MeCab [4]. Since the word list may not contain all of the words relevant to the opinion judgment subtask, we expanded our word list manually by using Japanese WordNet [5]. Table 1 shows the size of word list with POS type after expanded.

Table 1: The size of word list with POS type after expanded

POS Type	Size
Adjective	40
adjectival verb (na-adjective)	69
verbal noun (sahen-noun)	48
Noun	45
Verb	25
Adverb	16
Conjunction	9
Other	20
Total	272

2.4 Dependency pairs

In the past NTCIR workshop, many groups used dependency pair or syntactic pair for opinion detection [6, 7]. We also applied this method for opinion judgment subtask. First, the opinion sentences were analyzed by the Japanese dependency parser CaboCha [8] that output a dependency tree. Then, we extracted all dependency pairs and the dependency distance was set to 1. We defined these dependency pairs as following four types:

1. “～を～する”
   分割を主張する
   (advocate the division)
   重要性を強調する
   (stress the importance)
2. “～に～”
   ワシントン連邦高裁に控訴する
   (appeal to Washington Federal Court)
   同地裁に要請した。
   (had request to the district court)
3. “～と～”
   望んでいる」と示唆した。
   (implies that somebody hope)
   同じだ」と反対している
   (against that something is the same)
4. “～は、～” or “～は～”
   首相は心配している
   (prime minister concerned that)
   マイクロソフト社は、表明した。
   (the Microsoft said)

The dependency pair consists of the source element and the sink element. We can see from the above examples, source element is the named entity (person, organization, etc.) or noun type semantic primitive. However, we only focused on the sink element and prepared a verbal noun (sahen-noun) list for sink element matching.

2.5 Statistics and Date

We also observed that the non-opinion sentences tend to have numbers by analyzed the sample data. These numbers could present a statistics data, period or specific time, we summarized these sentences as follow:

1. Sentence contains a *statistics data*
   バリ島駐在の日本人向け旅行社によると、バリを訪れる年間約３０万人の日本人観光客のうち６割以上が女性客。
   (According to the travel agency for Japanese in island of Bali said: "There were about 30 million Japanese tourists travel in island of Bali every year, which more than 60 percent were female.")
2. Sentence contains a *period of time*
   日本でも在日米軍が９５年１２月から９６年１月に沖ノ鳥島の演習で劣化ウラン弾１５２０発を誤射したことが判明。
   (Japan also confirmed that the U.S. forces in Japan launch 1520 depleted uranium ammunition in error during the Okinotorishima military exercise in December 1995 to January 1996.)
3. Sentence contains a *specific time*
   昨年１０月１２日、インドネシア・バリ島のディスコで爆弾テロが発生、日本人夫婦を含む外國人観光客ら２００人以上が死亡、３００人以上が負傷した。

(Oct. 12 the year before last, terrorist bombing in island of Bali, including the Japanese couples a total of more than 200 foreign tourists killed and more than 300 casualties.)

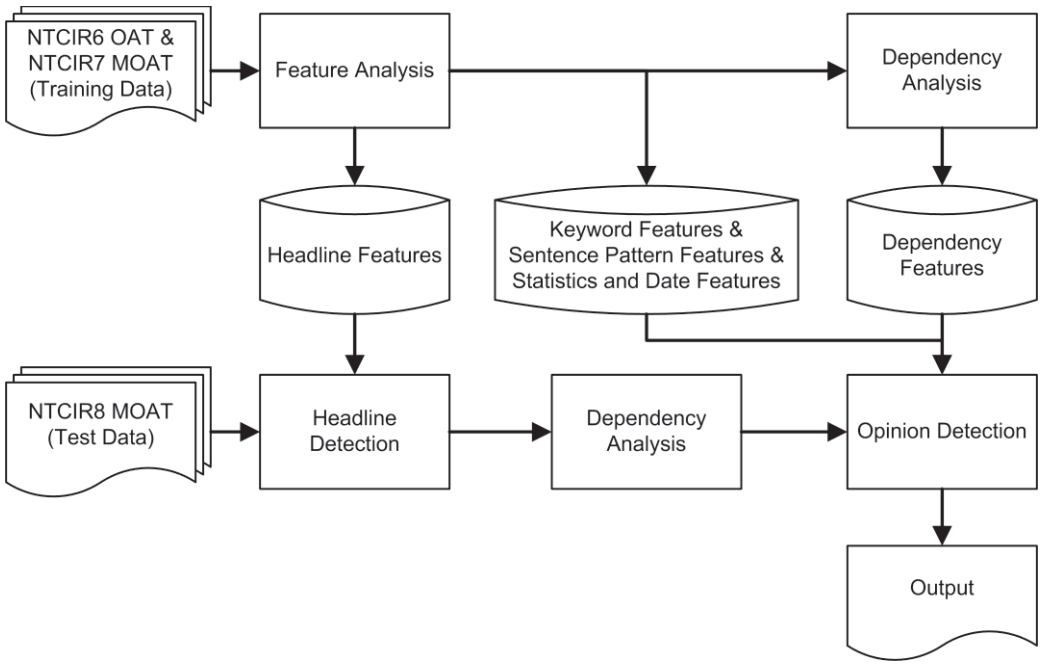

Figure 1: Overall system architecture

3. SYSTEM OVERVIEW

In this section, we will describe our system for opinionated sentence detection. Our system contains two parts: 1) generate the features for opinion detection, 2) detect the opinion sentences. Figure 1 shows our system architecture.

3.1 Dataset for training

We selected NTCIR6 OAT, NTCIR7 MOAT corpus and sample data as dataset for training. Table 2 shows the number of opinion / non-opinion sentences in the training dataset.

Table 2: The number of opinion / non-opinion sentences in the training dataset

	Opinion	Non-opinion
NTCIR6 OAT	3,026	6,682
NTCIR7 MOAT	1,562	4,323
NTCIR7&8 Sample	499	1,102
Total	5,087	12,107

3.2 Opinion score

Based on the dataset described in Section 3.1, we determined the features in the feature sets (described in Section 2) that was useful for opinion detection or not. These feature sets consist of a list of patterns or words and we were calculating the score for each item by following equation:

$$score(item_i) = \frac{Op(item_i) - On(item_i)}{Op(item_i) + On(item_i)}$$

where $item_i$ is a list item, $score(item_i)$ is opinion score of $item_i$, $Op(item_i)$ is frequency of $item_i$ appeared in opinion sentences, and $On(item_i)$ is frequency of $item_i$ appeared in non-opinion sentences. Finally, we constructed four dictionaries that included opinion score for sentence patterns, keywords, dependency pairs, and numeral features.

3.3 Opinion detection

In this section, we describe how our system works to detect the opinion sentence. In the first step, the system extracts headlines and determines whether is an opinion sentence or not based on Section 2.1. For other non-headline sentences, the second step, the sentence will be analyzed by the Japanese dependency parser CaboCha for dependency analysis. The third step, the system looks up the features in the dictionaries described in Section 3.2 and sums up the score of each feature appeared in a sentence. Final step, if the score calculated by third step is greater than 0, the systems determines this sentence as opinion.

4. EVALUATION RESULT

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

4.1 Formal Run

We finally submitted three runs as the formal run. The following briefly summarizes the features of the submitted three runs for the opinion judgment subtask.

Run1 Using all features and training without the sample data as our method

Run2 Delete the headline features and training without the sample data

Run3 Using all features and all training dataset as our method

The evaluation results are presented in Table 3. Seen from the table, we find that our recall rate is relatively lower than precision.

Table 3: Opinion detection results

	Precision	Recall	F-measure
Run1	67.30	49.86	57.28
Run2	67.74	47.65	55.95
Run3	67.86	51.53	58.58

In Section 3.2, we have determined the features in the feature sets. Since these feature sets are different, we could probably use weighting value to tuning these feature sets. We fixed some bugs of our system and did the following experiment to confirm the effects of each feature sets.

4.2 Effects of feature sets for opinion judgment subtask

To confirm the effects of each feature sets described in Section 2 for the opinion judgment subtask, we using the formal run data. Training data is the same of Section 3.1 and the formal run data is used for the test.

According to the results of formal run, we have confirmed that the headline features are effective. We made comparison between a result of using all feature sets and results of deleting one feature set (feature set 2.2, 2.3, 2.4, and 2.5). Table 4 shows deleted feature sets, each result and difference of F-measure. Larger difference means more effective feature sets.

Table 4 shows feature set 2.5 improve precision but didn't improve recall and F-measure. We concluded that the lower recall scores cause the F-measure fell. In some opinion sentences, these numeral features appear more times than other features and our system output these sentences as non-opinion sentences.

Table 4: Feature set evaluation in the opinion judgment subtask

Deleted feature	P	R	F	Diff
None	68.52	51.28	58.66	
Feature set 2.2	67.81	37.04	47.91	10.75
Feature set 2.3	73.26	38.61	50.57	8.09
Feature set 2.4	68.37	47.91	56.34	2.32
Feature set 2.5	66.54	53.86	59.53	-0.87

5. CONCLUSION

We developed an opinion detection system for Japanese newspapers at NTCIR-8 MOAT task. In the formal run, the opinion detection subtask attained precision 67.86, recall 51.53, F-measure 58.58.

Despite the simple approach of our system in using some linguistic features into the system of analyzing opinions in newspaper articles. Overall, it has achieved high performance in opinionated sentence detection task.

6. REFERENCES

[1] Y. Seki, L. W. Ku, L Sun, H. H. Chen, and N. Kando. Overview of Multilingual Opinion Analysis Task at NTCIR-8. In Proc. of the Eighth NTCIR Workshop Meeting on Evaluation of Information Access Technologies, NII, Japan, June 2010.

[2] Y. Seki, D. K. Evans, L. W. Ku, H. H. Chen, N. Kando, and C. Y. Lin. Overview of Opinion Analysis Pilot Task at NTCIR-6. In Proc. of the Sixth NTCIR Workshop Meeting on Evaluation of Information Access Technologies, pages 265-278, NII, Japan, May 2007.

[3] Y. Seki, D. K. Evans, L. W. Ku, L. Sun, H. H. Chen, and N. Kando. Overview of Multilingual Opinion Analysis Task at NTCIR-7. In Proc. of the Seventh NTCIR Workshop Meeting on Evaluation of Information Access Technologies, pages 185-203, NII, Japan, December 2008.

[4] Mecab. Yet Another Part-of-Speech and Morphological Analyzer. Available from: <http://chasen.org/~taku/software/mecab>.

[5] H. Isahara, F. Bond, K. Uchimoto, M. Utiyama and K. Kanzaki. Development of Japanese WordNet. In LREC-2008, Marrakech.

[6] Y. Seki. A Multilingual Polarity Classification Method using Multi-label Classification Technique Based on Corpus Analysis. In Proc. of the Seventh NTCIR Workshop Meeting on Evaluation of Information Access Technologies, pages 284-291, NII, Japan, December 2008.

[7] H. Mizuguchi, M. Tsuchida, K. Tateishi, D. Kusui. Opinion sentence and topic relevant sentence extraction by using coherent structure among the sentences. In Proc. of the Seventh NTCIR Workshop Meeting on Evaluation of Information Access Technologies, pages 272-277, NII, Japan, December 2008.

[8] K. Taku. CaboCha: Yet Another Japanese Dependency Structure Analyzer [online]. Technical report, Nara Institute of Science and Technology, 2004. [cited 2004-10-26]. Available from: <http://chasen.org/~taku/software/cabocha>.