Opinion Extraction from Editorial Articles based on Context Information and Predicate Classification

Yoshimi Suzuki[*], **Fumiyo Fukumoto**[*]

[*]University of Yamanashi
Takeda 4-3-11, Kofu, Yamanashi, Japan
{ysuzuki, fukumoto}@yamanashi.ac.jp

Abstract

Opinion extraction supports various tasks such as sentiment analysis in user reviews for recommendations and editorial summarization. In this paper, we address the problem of opinion extraction from newspaper editorials. To extract author's opinion, we used context information addition to the features within a single sentence only. Context information are a location of the target sentence, and its preceding, and succeeding sentences. We defined the opinion extraction task as a sequence labeling problem, using conditional random fields (CRF). Moreover, we used the results of predicate classification. We used Japanese newspaper editorials in the experiments, and used multiple combination of features of CRF to reveal which features are effective for opinion extraction. The experimental results show the effectiveness of the method, especially, predicate expression, location and previous sentence are effective for opinion extraction.

1. Introduction

Opinion extraction supports various tasks such as sentiment analysis in user reviews for recommendations, document classification, and editorial summarization. Much of the previous work on automatic opinion extraction focused on sentiment or subjectivity classification at sentence level. However, it is not sufficient to find opinion by using features within a single sentence only. For instance, in the news documents, although the features of authors' opinions are often expressed in predicates of a sentence, it is unusual to find only one sentence containing opinion as well as factual information (Wiebe et al., 2005).

In this paper, we focused on editorials of Japanese newspaper, and present a method to extract authors' opinions. We assumed that opinion parts consist of some sentences and some of them are difficult to recognize as opinion sentence by only themselves. Therefore, we employed conditional random fields (CRF) (Lafferty et al., 2001) to use context information: the preceding and the following sentences. Context information indicates a location of the target sentence, and its preceding, and succeeding sentences. In the experiments, we used multiple combination of features of CRF to reveal which features are effective for opinion extraction. For a feature of CRF, we classified predicates which frequently appear in editorial articles.

We used Japanese newspaper editorials in the experiments. However, the features used our method are very simple. Therefore, our method can be applied easily to different languages given documents.

2. RELATED WORK

The analysis of opinions and emotions in language is a practical problem as well as the process of large-scale heterogeneous data since the World-Wide Web is widely used. Wiebe *et al.* (Wiebe et al., 2005) presented annotation scheme that identifies key components and properties of opinions and emotions in language. They described annotation of opinion, emotion, sentiment, speculations, evaluations and other private states.

Apart from the corpus annotation, there are many attempts on the automatic identification of opinions (Hu and Liu, 2004; Kim and Hovy, 2004; Kobayashi et al., 2004; Burfoot and Baldwin, 2009; Mihalcea and Pulman, 2007; Wicaksono and Myaeng, 2013). The earliest work on opinion mining is the work on Hu *et al.* (Hu and Liu, 2004). They proposed a method of summarization by using opinion extraction from customer reviews on the web. The method consists of three steps: (i) mining product features, (ii) identifying opinion sentences, and (iii) summarizing the results. They reported that the experimental results using reviews of a number of products sold online demonstrated the effectiveness of the techniques. Kim *et al.* focused on English words and sentences. They proposed a template-based approach, a sentiment classifier by using thesauri (Kim and Hovy, 2004). Kobayashi *et al.* (Kobayashi et al., 2004) proposed a semi-automatic method to extract evaluative expressions. They used review sites on the Web for car and game domains, and extracted particular co-occurrence patterns of evaluated subject, focused attribute and value expressions. Balahur *et al.* (Balahur et al., 2009) proposed a method for opinion mining from quotations in newspaper articles.

Several researchers have investigated the use of statistics and machine learning techniques. Burfoot *et al.* (Burfoot and Baldwin, 2009) proposed a method for extracting satirical articles from newswire using SVMs. Mihalcea *et al.* (Mihalcea and Pulman, 2007) proposed a method using Naive Bayes and SVMs for extracting humour text. However, they used features within a sentence and ignore the relationships between sentences. Wilson (Wilson, 2008) classified sentences of news documents into 6 attitude types, i.e. 'Sentiment', 'Agreement', 'Arguing', 'Intention', 'Speculation' and 'Other 'Attitude'. They used four types of machine learning, rule learning (Cohen, 1996), boosting, support vector machines, and k-nearest neighbor. Wicaksono *et al.* (Wicaksono and Myaeng, 2013) proposed a method to extract advice-revealing and their context sentences from Web forms based on Conditional random fields (CRFs) (Lafferty et al., 2001). They compared their Multiple Linear CRFs (ML-CRF) and 2 dimensional CRFs Plus (2D-CRF+) with traditional machine learning models for advice-revealing sentences, and showed that ML-CRF is the best approach among other models studied in their paper. Similar to Wicaksono *et al.* method, we used CRFs to extract opinion expressions from editorial news. The difference is that we examined the effect of multiple combination of features, while they investigated the effect of machine learning techniques.

3. OPINION EXPRESSION IN EDITORIAL ARTICLES

We assumed editors would like to argue facts which occurred in the past and make recommendation to the government or readers in editorial articles. We classified authors' opinion into 4 types, *i.e.* 'speculation', 'hope', 'proposition' and 'assertion'. Table 1 shows 4 types of opinion on two-dimensional surface. Table 2 shows an example of opinion sentence assigned to each type. Typical expression, such as "*utagawasii* (Serious questions remain)" in speculation type, and "*kadai-ha ooi* (be numerous matters)" in assertion type. The opinion defined in Table 1 are the intersection among surface expression, location, and positional relation between other opinion sentences. Most of the opinions have typical expression. We assigned each sentence illustrated in Table 2 to one of the four types by using surface expression, location, and positional relation of its preceding and succeeding sentences.

Table 1: Four types of opinion in editorial articles.

	weak opinion	strong opinion
past, present	Speculation	Assertion
future	Hope	Proposition, Suggestion

Table 2: Examples of opinion sentences.

type	example (Japanese)
Speculation	Serious questions remain as to whether a decision on whether to expand overseas activities of the Self-Defense Forces should be left to legislators who make such highly questionable remarks. (自衛隊の海外での活動を 拡大することの是非の判断を、こういう 問題発言をする 政治家たちに任せていいのか疑わしい。)
Assertion	There are numerous matters that need to be discussed in working out a new stadium plan. (新しい計画の策定にあたって議論すべき 課題は多い。)
Hope	Fees should be reduced further. (さらなる 料金引き 下げを 求めたい。)
Proposition	It is necessary to ascertain who is responsible for causing confusion over the matter. (混乱を 招いた責任はどこにあるのか 今後の検証が必要だ。)

4. FEATURES FOR OPINION EXTRACTION

The feature we defined for opinion extraction are categorized into three as shown in Table 3. From the newspaper editorial analysis, we used the following seven features to extract opinion. Predicates including verbs, adjective and adverbs are important features to extract opinion expression. It is often the case that the sentence of the last part of the editorial articles includes an opinion. Let us take a look at Japanese newspaper editorials.

Table 3: Features for opinion extraction.

Syntactic Feature		
1.	feature	predicate expression (PE)
	example	I would like to expect :*kitai shitai*
	reason	Most opinion sentences have typical predicate expression.
2.	feature	root form of predicate (PR)
	example	hope :*kitai suru*
	reason	Predicate expressions have various conjugations.
3.	feature	cluster No. of predicate (PC)
	specific	All predicates are classified into 50 clusters)
	reason	Predicate expressions can be classified into some semantic groups.
4.	feature	subject (Subj)
	example	we, government
	reason	Most opinion sentences must have typical subjects.
Location		
1.	feature	sentence location in article (LIA)
	specific	partition number is 5.
	reason	Most of opinions must appear in last half of Japanese editorials.
2.	feature	sentence location in paragraph (LIP)
	specific	partition number is 3.
	reason	Most of opinions must appear in last half of the paragraph of Japanese editorials.
Previous Sentence		
1.	feature	opinion type of preceding sentence (OTP)
	reason	Opinion sentences must concentrate in editorials.

5. CONDITIONAL RANDOM FIELD (CRF)

Based on the extracted features, we identified opinion expression by using CRFs. CRFs is a well known technique for solving sequence labeling problems. CRFs are discriminative models and can deal with many correlated features in the inputs. CRFs have a single exponential model for the joint probability of the entire paths given the input sentence. Given a sequence $\mathbf{X} = (x_1, x_2, \cdots, x_n)$, where n is the number of sentences in the inputs, the goal is to find the sequence of hidden labels $\mathbf{Y} = (y_1, y_2,$

$\cdots, y_n)$. The sequence of hidden labels are obtained by a conditional distribution function given by:

$$p(\mathsf{y}\mathsf{x}) \quad = \tfrac{1}{Z_{\mathsf{x}}}(\prod_{i=1}^{n-1} f(y_i, y_{i+1}, \mathsf{x}, i)), \qquad (1)$$

where Z_{x} is a normalization factor, and f indicates an arbitrary feature function over i-th sequence. We used CRFs to extract opinion expressions.

6. PREDICATE CLASSIFICATION USING WORD2VEC AND SPECTRAL CLUSTERING

It is effective to use information of predicate in each sentence for opinion extraction, but there are many variation in editorial article. In order to tackle this problem, we classified predicate into some classes using word2vec (Mikolov et al., 2013) and spectral clustering. Firstly, we calculated similarity between predicates which frequently appear in editorial articles of Japanese newspaper. Then, we classified the predicates into 50 classes by using spectral clustering.

Table 4 shows a part of results of word2vec and spectral clustering. For this experiment, we used all editorial articles in Japanese newspaper (Mainichi Shimbun newspaper written in Japanese) from 1991 to 2012. We collected 587 predicates which appeared more than 80 times in the editorial articles. As we can see from Table 4, words which are similar to "required for (*hitsuyouda*)", and many predicates which are a sense of "Assertion" are extracted. In Table 4, "cluster#" means that each predicate is classified into the class.

Table 4: Results of Word2Vec and spectral clustering (best 15 words which are similar to "required for (*hitsuyouda*)").

	predicate	similarity	cluster#
1.	required for (*hitsuyoudearu*)	0.874947	9
2.	It will be necessary (*hitsuyoudarou*)	0.845753	49
3.	required (*motomerareru*)	0.829809	9
4.	be indispensable (*kakasenai*)	0.807845	15
5.	be required (*hitsuyouninaru*)	0.789526	9
6.	indispensable (*fukaketsuda*)	0.781698	49
7.	necessary (*hitsuyoudehanaika*)	0.780348	28
8.	is desirable (*nozomareru*)	0.772844	26
9.	are required (*motomerareteiru*)	0.756382	49
10.	indispensable (*fukaketsudearu*)	0.752851	28
11.	important (*taisetsuda*)	0.712825	33
12.	important (*daizida*)	0.698118	12
13.	it is vital (*kannyouda*)	0.697411	12
14.	it would be necessary (*hitsuyoudearou*)	0.679321	12
15.	it is hurried (*isogareru*)	0.677352	38

7. EXPERIMENTS

7.1. Experimental Setup

We selected editorial articles of Japanese newspaper (Mainichi Shimbun newspaper written in Japanese) for opinion extraction. We used one year (2011) Mainichi Japanese Newspaper corpus for training and test data. Table 5 shows the number of editorial articles, sentences and each opinion type in the editorial articles. In Table 5, "# of edi" means "the number of editorials" and "# of sen" means "the number of sentences".

We used 12-fold cross validation. More precisely, we divided editorial articles into twelve months shown in Table 5. We used eleven folds to train the classifier, and the remaining fold to test the classifier. The process is repeated 12 times, and we obtained the average classification accuracy over 12 folds. We applied CaboCha (Kudo and Matsumoto, 2002) for morphological analysis, and CRF++ [1]. CRF++ is an customizable implementation of CRFs for labeling sequential data. It can be applied to a variety of NLP tasks. We used feature sets described in Sec4..

7.2. Results

We examined which feature combination is effective for opinion extraction. We thus conducted an experiment using combination of seven features. The results are shown in Table 6. In Table 6, E, R, C, S, A, P and T illustrate PE, PR, PC, Subj, LIA, LIP and OTP of the list in Sec 4. , respectively. In the columns of E, R, C, S, A, P and T, "1" means that the feature is used for opinion extraction, while "0" means that the feature is not used for opinion extraction.

We can see from Table 6 that the best result was when we use all features, and the F-measure was 0.59. These results indicate that the combination of these three features are especially effective for opinion extraction. Table 7 refers to the result of opinion type classification when we used the best results shown in Table 6.

8. DISCUSSION

We can see from Table 6 that the results using only predicate expression achieved 0.83 precision, while recall was 0.31. This shows that when the predicate expression appeared in the test data does not appear in the training data, the system can not extract opinion sentences correctly. When we added root form of predicate (PR) to predicate expression (PE), precision was decreased to 0.53. This is because inflection of some predicates is important for opinion extraction. When we added cluster No. of predicate (PC), precision was decreased. However, recall

[1] CRF++ : http://crfpp.sourceforge.net/

Table 6: Results of opinion extraction.

Features							Rec	Pre	F1	Features							Rec	Pre	F1
E	R	C	S	A	P	T				E	R	C	S	A	P	T			
1	1	1	1	1	1	1	0.84	0.46	0.59	1	0	1	1	1	1	1	0.86	0.37	0.51
1	1	1	1	1	1	0	0.83	0.46	0.59	1	0	1	1	1	1	0	0.85	0.36	0.51
1	1	1	1	1	0	1	0.84	0.46	0.59	1	0	1	1	1	0	1	0.86	0.37	0.51
1	1	1	1	1	0	0	0.84	0.46	0.59	1	0	1	1	1	0	0	0.85	0.36	0.51
1	1	1	1	0	1	1	0.82	0.44	0.58	1	0	1	1	0	1	1	0.86	0.36	0.50
1	1	1	1	0	1	0	0.83	0.44	0.57	1	0	1	1	0	1	0	0.87	0.35	0.50
1	1	1	1	0	0	1	0.81	0.44	0.57	1	0	1	1	0	0	1	0.86	0.35	0.50
1	1	1	1	0	0	0	0.82	0.45	0.58	1	0	1	1	0	0	0	0.87	0.36	0.51
1	1	1	0	1	1	1	0.84	0.46	0.59	1	0	1	0	1	1	1	0.86	0.37	0.51
1	1	1	0	1	1	0	0.84	0.46	0.59	1	0	1	0	1	1	0	0.86	0.37	0.51
1	1	1	0	1	0	1	0.84	0.46	0.59	1	0	1	0	1	0	1	0.86	0.36	0.51
1	1	1	0	1	0	0	0.83	0.45	0.58	1	0	1	0	1	0	0	0.86	0.36	0.51
1	1	1	0	0	1	1	0.82	0.44	0.57	1	0	1	0	0	1	1	0.86	0.36	0.50
1	1	1	0	0	1	0	0.83	0.44	0.58	1	0	1	0	0	1	0	0.88	0.36	0.51
1	1	1	0	0	0	1	0.83	0.44	0.57	1	0	1	0	0	0	1	0.86	0.36	0.51
1	1	1	0	0	0	0	0.83	0.44	0.58	1	0	1	0	0	0	0	0.86	0.36	0.51
1	1	0	1	1	1	1	0.84	0.45	0.59	1	0	0	1	1	1	1	0.86	0.36	0.50
1	1	0	1	1	1	0	0.84	0.45	0.59	1	0	0	1	1	1	0	0.86	0.35	0.50
1	1	0	1	1	0	1	0.84	0.45	0.59	1	0	0	1	1	0	1	0.87	0.35	0.50
1	1	0	1	1	0	0	0.84	0.44	0.58	1	0	0	1	1	0	0	0.86	0.35	0.50
1	1	0	1	0	1	1	0.83	0.44	0.57	1	0	0	1	0	1	1	0.87	0.34	0.49
1	1	0	1	0	1	0	0.82	0.43	0.57	1	0	0	1	0	1	0	0.87	0.34	0.49
1	1	0	1	0	0	1	0.82	0.43	0.57	1	0	0	1	0	0	1	0.87	0.34	0.49
1	1	0	1	0	0	0	0.70	0.49	0.57	1	0	0	1	0	0	0	0.58	0.44	0.50
1	1	0	0	1	1	1	0.83	0.45	0.59	1	0	0	0	1	1	1	0.87	0.36	0.51
1	1	0	0	1	1	0	0.84	0.45	0.58	1	0	0	0	1	1	0	0.86	0.35	0.50
1	1	0	0	1	0	1	0.84	0.45	0.59	1	0	0	0	1	0	1	0.88	0.35	0.50
1	1	0	0	1	0	0	0.84	0.45	0.58	1	0	0	0	1	0	0	0.87	0.35	0.50
1	1	0	0	0	1	1	0.82	0.43	0.57	1	0	0	0	0	1	1	0.87	0.35	0.50
1	1	0	0	0	1	0	0.82	0.44	0.57	1	0	0	0	0	1	0	0.88	0.35	0.50
1	1	0	0	0	0	1	0.83	0.43	0.57	1	0	0	0	0	0	1	0.87	0.35	0.50
1	1	0	0	0	0	0	0.53	0.53	0.53	1	0	0	0	0	0	0	0.31	0.83	0.45

was increased to 0.86 and we had an improvement of F-measure. This is because we could extract many expressions as opinion expression using predicate classification.

When we added subject (Subj) to predicate expression (PE), precision was slightly decreased. However, recall was significantly increased, and we had an improvement of F-measure. The observation shows that the integration of features is effective for opinion extraction. When we used sentence position, recall was worse while precision was better. A sentence position which is effective to find opinion depends on the opinion types. For further improvement, it is necessary to investigate an effective sentence position according to each type of the opinion. A sentence location within a paragraph, and a paragraph location appeared in the sentence significantly improve recall. Similarly, When we used opinion type of the preceding sentence, recall was improved. These features are also effective to improve overall performance.

The experimental results show that the best result was the combination of E, R, C, S, A, P and T, and the F-measure attained at 0.59. From the above observations, we conclude that multiple combination of features are effective for opinion extraction.

Next, we examined how the method correctly assigned a sentence to each type of opinion. As can be seen clearly from Table 7 that the best result was" hope" and F-measure was attained at 0.62. In contrast, it is difficult to identify opinion to" speculation" as the F-measure was only 0.30. It is not surprising because the training data assigned to " speculation" have various expressions, and it is not easy to classified into" speculation" manually.

For future work, we will extend our framework to improve overall performance against a small number of training data. We note that we used surface information, *i.e.*, noun and verb words in articles as a feature. Therefore, the method ignore the sense of terms such as synonyms and antonyms. The earliest known technique for smoothing the term distributions through the use of latent classes is the Probabilistic Latent Semantic Analysis (PLSA) (Hofmann, 1999), and it has been shown to improve the performance of a number of information access such as text classification (Xue et al., 2008). It is definitely worth trying with

Table 5: Number of editorials and sentences.

Mo.	# of edi (# of sen)	S	A	H	P	O
Jan.	52 (1,437)	75	97	46	20	1,199
Feb.	52 (1,282)	65	89	27	25	1,077
Mar.	57 (1,612)	24	101	55	46	1,386
Apr.	56 (1,579)	27	155	48	47	1,302
May	59 (1,476)	17	151	49	34	1,225
Jun.	57 (1,397)	12	171	47	35	1,132
Jul.	60 (1,433)	21	148	39	20	1,205
Aug.	53 (1,507)	53	129	38	31	1,256
Sep.	55 (1,388)	46	165	44	31	1,102
Oct.	60 (1,418)	41	151	40	20	1,166
Nov.	57 (1,370)	57	162	36	38	1,077
Dec.	58 (1,442)	67	140	32	39	1,164
Total	676 (17,341)	505	1,659	501	386	14,291

Table 7: The results of opinion type classification.

Type	Recall	Precision	F-measure
Speculation	0.46	0.22	0.30
Assertion	0.58	0.34	0.43
Hope	0.65	0.59	0.62
Proposition	0.50	0.23	0.31

our method to achieve type classification accuracy.

9. CONCLUSIONS AND FUTURE WORK

We proposed a method for opinion expression of editorial articles. Although training data and test data are not so large, this study led to the following conclusions: (i) predicate expression, location and previous sentence are effective for opinion extraction. (ii) results of opinion extraction are depend on the types of opinion. Future work will include (i) incorporating smoothing technique to use a sense as a feature, (ii) applying the method to a large number of editorial articles for quantitative evaluation, (iii) comparing our method with other methods.

Acknowledgments

The authors would like to thank anonymous reviewers for their valuable comments. This work was supported by the Grant-in-aid for the Japan Society for the Promotion of Science (JSPS), No.26330247.

10. References

Balahur, Alexandra, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen, and Mijail Kabadjov, 2009. Opinion mining on newspaper quotations. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops*.

Burfoot, Clint and Timothy Baldwin, 2009. Automatic satire detection: Are you having a laugh? In *Proceedings of ACL-IJCNLP 2009*.

Cohen, W., 1996. Learning trees and rules with set-valued features. In *the 13th National Conference on Artificial Intelligence*.

Hofmann, T., 1999. Probabilistic Latent Semantic Indexing. In *Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Hu, Minqing and Bing Liu, 2004. Mining and summarizing customer reviews. In *KDD'04: Proceedigs of the tenth SIGKDD international conference on Knowledge discovery and data mining*.

Kim, Soo-Min and Eduard Hovy, 2004. Determining the sentiment of opinions. In *COLING 2004*.

Kobayashi, Inui, Matsumoto, Tateishi, and Fukushima, 2004. Collecting evaluative expression for opinion extraction. In *IJCNLP 2004: Proceedings of the 6th International Conference on Natural Language Processing*.

Kudo, Taku and Yuji Matsumoto, 2002. Japanese dependency analysis using cascaded chunking. In *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops)*.

Lafferty, John, Andrew McCallum, and Fernando Pereira, 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML 2001: Proceedings of the 18th International Conference on Machine Learning*.

Mihalcea, Rada and Stephen G. Pulman, 2007. Characterizing humour: An exploration of features in humorous texts. In *CICLing 2007: Proceedings of the 8th International Conference on Intelligent Text Processing and Computational Linguistics*.

Mikolov, T., K. Chen, G. Corrado, and J. Dean, 2013. Efficient Estimation of Word Representations in Vector Space. In *Proc. of the International Conference on Learning Representations Workshop*.

Wicaksono, Alfan Farizki and Sung-Hyon Myaeng, 2013. Toward advice mining: Conditional random fields for extracting advice-revealing text units. In *CIKM 2013: Proceedings of ACM International Conference on Information and Knowledge Management*.

Wiebe, Janyce, Theresa Wilson, and Claire Cardie, 2005. Annotating expressions of opinions and emotions in language. In *Language Resources and Evaluation*.

Wilson, Theresa Ann, 2008. Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states. In *Doctoral dissertation (University of Pittsburgh)*.

Xue, G. R., W. Dai, Q. Yang, and Y. Yu, 2008. Topic-bridged PLSA for Cross-Domain Text Classification. In *Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.