Opinion Mining on the Web by Extracting Subject-Aspect-Evaluation Relations

Nozomi Kobayashi and **Ryu Iida** and **Kentaro Inui** and **Yuji Matsumoto**

Nara Institute of Science and Technology,
Takayama, Ikoma, Nara 630-0192, Japan
{nozomi-k,ryu-i,inui,matsu}@is.naist.jp

Abstract

This paper addresses the task of extracting opinions from a given document collection. Assuming that an opinion can be represented as a tuple ⟨*Subject*, *Aspect*, *Evaluation*⟩, we propose a computational method to extract such tuples from texts. In this method, the main task is decomposed into (a) the process of extracting *Aspect-Evaluation* pairs from a given text and (b) the process of judging whether an extracted pair expresses an opinion of the author. We apply machine-learning techniques to both subtasks. We also report on the results of our experiments and discuss future directions.

Introduction

The explosive spread of communication on the Web has attracted increasing interest in technologies for automatically mining opinions and recommendations from large amount of information on the message boards and blog pages for opinions and recommendations.

Previous approaches to the task of mining a large-scale document collection for opinions can be classified into two groups: the document classification approach and the information extraction approach. In the document classification approach, researchers have been exploring techniques for classifying documents according to semantic/sentiment orientation such as positive vs. negative (Dave, Lawrence, & Pennock 2003; Pang & Lee 2004; Turney 2002), etc. The information extraction approach, on the other hand, focuses on the task of extracting the elements which constitute opinions (Kanayama & Nasukawa 2004; Hu & Liu 2004; Gamon *et al.* 2005; Popescu & Etzioni 2005), etc.

The aim of this paper is to propose and evaluate a method for extracting opinions that represent an evaluation of a products together with the reason from documents. To achieve this, we consider our task from the information extraction viewpoint. We term the above task **opinion extraction** in this paper.

While they can be linguistically realized in many ways, opinions are in fact often expressed in the form of an aspect-evaluation pair. An aspect represents one aspect of a subject. Given this observation, we approach our goal by reducing

the task to a general problem of extracting triplet ⟨*Subject, Aspect, Evaluation*⟩ from a large-scale text collection. Technology for this opinion extraction task would be useful for collecting and summarizing latent opinions from the Web. A straightforward application might be visualization of collected opinions as suggested by Gamon *et al.*(2005) and Liu *et al.*(2005).

As we discuss later, an aspect and its evaluation may not appear in a fixed expression and may be separated in texts. Though an evaluation should have an aspect it is targeting, the aspect may be missing in a sentence. In this respect, finding the aspect of an evaluation is similar to finding the missing antecedent of an ellipsis. In this paper, we apply a machine learning-based method used for zero pronoun anaphora resolution to the opinion extraction problem and report on our experiments conducted on a domain-restricted set of Japanese texts excerpted from review pages on the Web.

Related work

In this section, we discuss previous approaches to the opinion extraction problem. Murano & Sato (2003) and Tateishi et al. (2001) proposed the method which uses pre-defined extraction patterns and a list of evaluative expressions. These extraction patterns and the list of evaluation expressions need to be created manually. However, as is the case in information extraction, manual construction of rules may require considerable cost to provide sufficient coverage and accuracy. Popescu & Etzioni (2005) also extract opinions using 10 extraction rules. however, as we show later, there are some cases where an aspect and its evaluation are not within a sentence. The pattern-based approach can not extract these cases.

Hu & Liu (2004) attempt to extract the aspects of target products on which customers have expressed their opinions using association mining, and to determine whether the opinions are positive or negative. Their aim is quite similar to ours. However, our work differs from theirs in that we extract the evaluation and its corresponding aspect at the expression level while they extract the aspects and determine their semantic orientation at the sentence level.

Taking the semantic parsing-based approach, Kanayama & Nasukawa (2004) apply the idea of transfer-based machine translation to the extraction of aspect-evaluation pairs.

Table 1: Examples of aspects

aspect types	examples of aspects
physical parts	engine, tire, steering, interior, seat
related object	manufacturer, dealer
attribute	size, color, design, performance, weight
aspects [of parts]	sound [of engine], stiffness [of body], power [of engine]

They regard the extraction task as translation from a text to a sentiment unit which consists of a sentiment evaluation, a predicate, and its arguments. Their idea is to replace the translation patterns and bilingual lexicons with sentiment expression patterns and a lexicon that specifies the polarity of expressions. Their method first analyzes the predicate-argument structure of a given input sentence making use of the sentence analysis component of an existing machine translation engine, and then extracts a sentiment unit from it, if any, using the transfer component.

We also consider the opinion extraction as the task of finding the arguments of an predicate. One important problem is that opinion expressions often appear with ellipses, which need to be resolved to accomplish the opinion extraction task. Our analysis of an opinion-tagged Japanese corpus (described below) showed that 30% of the aspect-evaluation pairs we found did not have a direct syntactic dependency relation within the sentence, mostly due to ellipsis. To analyze predicate argument structure robustly, we have to solve this problem. We address this problem by incorporating a machine learning-based technique for zero-anaphora resolution into our opinion extraction model.

Subject, aspect and evaluation

Our aim is to extract opinions expressed in the form of a triplet ⟨*Subject, Aspect, Evaluation*⟩. We elaborate on these notations below.

Subject Subject is a specific entity of the given class. (e.g. product name, service name, person name, organization name and so on, that is the target of evaluation).

Aspect Aspect is the particular aspect of the subject that writers have expressed evaluation on. This includes parts, aspects and aspect of parts. Here we use the notation of parts in a broad sense; parts of a subject include its physical parts, members, and various related objects. Table 1 shows examples of aspects in the automobile domain classified by aspect types.

Evaluation This includes the expressions representing writer's evaluation on the subject or the aspect (e.g. high, good, excellent, poor), or writer's emotion or mental attitude (e.g. like, dislike, satisfied). These expressions are realized typically by adjective, noun, verb or adverb phrases. Thus, we call the expressions "evaluation phrase".

Opinionhood

There are many types of "opinion" such as beliefs, evaluations, requests, etc. In this paper, we limit our focus on explicit evaluation and define an opinion as follows:

An opinion is a description that expresses the writer's subjective evaluation of a particular product or a certain aspect of it and implies the writer's sentiment orientation toward it.

By this definition, we exclude requests, factual or counter-factual descriptions and hearsay evidence from our target opinions. For example, "*The engine is powerful*" is an opinion, while a counter-factual sentence such as "*If only the engine were more powerful*" is not regarded as an opinion. Such types of opinion should be included in the scope of our future work.

Method for opinion extraction

We consider the task of extracting opinion tuples ⟨*Subject, Aspect, Evaluation*⟩ from review sites and message boards on the Web dedicated to providing and exchanging information about retail goods. On these Web pages, products are often specified clearly and it is frequently a trivial job to extract the information for the *Subject* slot. As we will show later, if the product name is given, it is not difficult to detect the *Subject* of the *Evaluation*. We therefore focus on the problem of extracting ⟨*Aspect, Evaluation*⟩ pairs.

In the process of aspect-evaluation pair identification for opinion extraction, we need to address following issues. First, the argument of the predicate may not appear in a fixed expression and may be separated. As we mentioned earlier, 30% of the pairs did not have direct dependency relations. In the following example, the aspect "*design*" and the evaluation "*like*" are not connected via a dependency relation, since "it" is elided.

⟨dezain-wa⟩$_a$ hen-dato iwarete-iruga watashi-wa ϕ ⟨suki⟩$_v$
⟨design⟩$_a$ be-weird said but I [it] ⟨like⟩$_v$
(*It is said that the design is weird, but I like it.*)

Here ϕ denotes an ellipsis, which does not appear in the sentence and is called a zero pronoun. ⟨⟩$_a$ denotes the word sequence corresponding to the Aspect. Likewise, we also use ⟨⟩$_v$ for the Evaluation.

Second, as pointed out by Hu & Liu (2004) and Popescu & Etzioni (2005), aspects may not always be explicitly expressed. Let us see two examples from the reviews of the automobile:

"⟨The seat⟩$_a$ is very ⟨comfortable⟩$_v$"
"A ⟨big⟩$_v$ car"

In the first example, both a evaluation and its corresponding aspect appear in the text, while in the second example, a evaluation appears in the text but its aspect is missing since it is inferable form the evaluation phrase and the context (in this example, "a big car" implies the "size" of the car is "big").
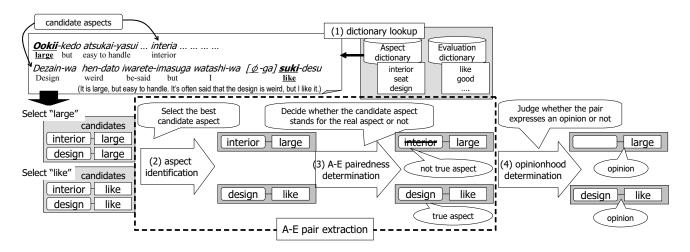
Figure 1: Process of opinion extraction

Third, recall that evaluation phrases do not always constitute opinions; the target of an evaluation may be neither a subject nor an aspect of a subject of the given domain, and furthermore we want to exclude evaluation phrases appearing, for example, in interrogative and subjunctive sentences. We therefore need to incorporate into our opinion extraction model a classifier for judging whether a given evaluation phrase constitutes an opinion. In the judgement, we expect that the information about the candidate aspect is likely to be useful for the determination. For example,

[1] *kosuto-ga   takai*   [2] *shiyou hindo-ga   takai*
    cost   high   frequency of use   high
  (*the cost is high.*)   ((*its) frequency of use is high.*)

These descriptions share the same evaluation expression "high". However, [1] is our target opinion, while [2] is not because this description describes rather a fact not a writers' subjective evaluation. As this example shows, the likelihood of the evaluation expression to be an opinion changes according to its aspect. From this observation, we expect that carrying out aspect identification before pairedness determination should outperform the counterpart model which executes the two subtasks in the reversed order.

As illustrated in Figure 1, we propose an opinion extraction model derived from the aforementioned discussion as follows:

1. **Dictionary lookup**: Assuming that we have domain-specific dictionaries of evaluation and aspect phrases, identify candidate aspects and evaluations by dictionary lookup. In Figure 1, "*large*" and "*like*" are evaluation candidates, and "*interior* and "*design*" are aspect candidates.

2. **Aspect identification**: For each candidate phrase, identify the best candidate aspect. In Figure 1, the model identifies the best candidate "*interior*" for the evaluation candidate "*large*" even if "*large*" does not have an explicit aspect.

3. **Aspect-evaluation pairedness determination**: Decide whether the candidate aspect is the true aspect of the evaluation (i.e. the evaluation has an explicit aspect in the text). In this step, we detect whether the evaluation has explicit aspect or not. Note that we do not identify what elided aspect is in the case where no explicit aspect is identified. In this example, "*design*" is the true aspect of the evaluation "*like*" and "*interior*" is not the true aspect of the evaluation "*large*".

4. **Opinionhood determination**: Judge whether the obtained aspect-evaluation pair[1] constitutes an opinion or not. In this example, both "*large*" and "*like*" constitutes an opinion, thus the model judges these are opinions.

We adopted the tournament model (Iida *et al.* 2003) for aspect identification. This model implements a pairwise comparison (i.e. a match) between two candidates in reference to the given evaluation treating it as a binary classification problem, and conducts a tournament which consists of a series of matches, in which the one that prevails through to the final round is declared the winner, namely, it is identified as the most likely candidate aspect. Each of the matches is conducted as a binary classification task in which one or the other candidate wins.

The pairedness determination task and the opinionhood determination task are also binary classification tasks. In the opinionhood determination step, we can use the information about whether the evaluation has a corresponding aspect or not. We therefore create two separate models for the cases where the evaluation does and does not have an aspect. These models can be implemented in a totally machine learning-based fashion.

Experiments

We conducted experiments with Japanese Web documents to empirically evaluate the performance of our opinion extraction model, focusing particularly on the validity of the method discussed in the previous section.

---

[1]For simplicity, we call a evaluation both with and without an aspect uniformly by the term *aspect-evaluation pair* unless the distinction is important.

Table 2: Features used in each model. AI: the attribute identification model, PD: the pairedness determination model, OD: the opinionhood determination model.

	AI	PD	OD(A-V)	OD(V)
a	√	√	√	√
b	√	√	√	

Opinion-tagged corpus

We created an opinion-tagged Japanese corpus consisting of 288 review articles in the automobile domain (4,442 sentences). The definitions of evaluation and aspect are based on the discussion in the previous section.

Note that if some aspects are in a hierarchical relation with each other, we asked the annotator to choose the aspect lowest in the hierarchy as the aspect of the evaluation. The hierarchical relation we mentioned includes part-of (e.g. "the switch of the air conditioner") and aspect-of (e.g. "the sound of the engine") relations. For example, in *"the sound of the engine is good"*, only *sound* is annotated as the aspect of the evaluation *good*.

The corpus contains 2,191 evaluations with explicit aspect and 420 evaluations without an explicit aspect. Most of the aspects appear in the same sentence as their corresponding evaluations or in the immediately preceding sentence (99% of the total number of pairs). Therefore, we extract aspects and their corresponding evaluations from the same sentence or from the preceding sentence.

Experimental method

As preprocessing, we analyzed the opinion-tagged corpus using the Japanese morphological analyzer *ChaSen*[2] and the Japanese dependency structure analyzer *CaboCha* [3].

We used Support Vector Machines to train the models for aspect identification, pairedness determination and opinionhood determination. We used the 2nd order polynomial kernel as the kernel function for SVMs. Evaluation was performed by 10-fold cross validation using all the data.

Dictionaries   We use dictionaries for identification of aspect and evaluation candidates. We constructed an aspect dictionary and an evaluation dictionary from review articles about automobiles (230,000 sentences in total) using the semi-automatic method proposed by (Kobayashi *et al.* 2004).

We assume that we have large dictionary which covers most of the aspect and evaluation phrases, thus we added to the dictionaries expressions which frequently appeared in the opinion-tagged corpus. The final size of the dictionaries becomes 3,777 aspect phrases and 3,950 evaluation phrases.

Features   We extracted the following two types of features from the aspect candidate and the evaluation candidate:

(a) surface spelling and part-of-speech of the target evaluation expression, as well as those of its dependent phrase and those in its depended phrase(s)

(b) relation between the target evaluation and candidate aspect (distance between them, existence of dependency relation, existence of a co-occurrence relation)

Table 2 summarizes which of the following types of features are used for each model. Existence of a co-occurrence relation is determined by reference to a predefined co-occurrence list that contains aspect-evaluation pair information such as "height of vehicle – low". We created the list from the 230,000 sentences described in previous section by applying the aspect and evaluation dictionary and extracting aspect-evaluation pairs if there is a dependency relation between the aspect and the evaluation. The number of pairs we extracted was about 48,000.

Results

Table 3 shows the results of opinion extraction. In the table, "evaluation with explicit aspect" indicates recall and precision of aspect-evaluation pairs where both a evaluation and its aspect appear in the text, and "evaluation without explicit aspect" indicate the result where the evaluation appears in the text while its aspect is missing. "aspect-evaluation pairs" is sum of above two rows.

We compared our model with the baseline model. In this model, if the candidate evaluation and a candidate aspect are connected via a dependency relation, the candidate evaluation is judged to have an aspect. When none of the candidate aspects have a dependency relation, the candidate evaluation is judged not to have an aspect.

We evaluated the results by recall R and precision P defined as follows (For simplicity, we substitute "A-E" for aspect-evaluation pair):

$$R = \frac{\text{correctly extracted A-E opinions}}{\text{total number of A-E opinions}},$$

$$P = \frac{\text{correctly extracted A-E opinions}}{\text{total number of A-E opinions found by the system}}.$$

We also used the F-measure, which is the harmonic mean of precision and recall:

$$F-measure = \frac{2 \times recall \times precision}{(recall + precision)}$$

In order to demonstrate the effectiveness of the information about the candidate aspect, we evaluated the results of pair extraction and opinionhood determination separately. Table 4 shows the results. In the pair extraction, we assume that the evaluation is given, and evaluate how successfully aspect-evaluation pairs are extracted.

Discussions

From Table 3, we can see that recall of our model outperforms the baseline model, since this method can extract pairs which are not connected via a dependency relation in the sentence. Moreover, the precision of our method outperforms the baseline model.

In what follows, we discuss the results of pair extraction and opinionhood determination.

Table 3: The precision and the recall for opinion extraction

procedure		evaluation with explicit aspect		evaluation without explicit aspect		aspect-evaluation pairs	
baseline	precision	60.5%	(1130/1869)	10.6%	(249/2340)	32.8%	(1379/4209)
	recall	51.6%	(1130/2191)	59.3%	(249/420)	52.8%	(1379/2611)
	F-measure	55.7		21.0		40.5	
proposed model	precision	80.5%	(1175/1460)	30.2%	(150/497)	67.7%	(1325/1957)
	recall	53.6%	(1175/2191)	35.7%	(150/420)	50.7%	(1325/2611)
	F-measure	64.4		32.7		58.0	

Table 4: The result of pair extraction and opinionhood determination
AI: the aspect identification model, PD: the pairedness determination model, OD: the opinionhood determination model.

	procedure	precision		recall	
	baseline (dependency)	71.1%	(1385/1929)	63.2%	(1385/2191)
	PD→AI	65.3%	(1579/2419)	72.1%	(1579/2191)
pair extraction	AI→PD	76.6%	(1645/2148)	75.1%	(1645/2191)
	(dependency)	87.7%	(1303/1486)	79.6%	(1303/1637)
	(no dependency)	51.7%	(342/ 662)	61.7%	(342/ 554)
opinionhood determination	OD	74.0%	(1554/2101)	60.2%	(1554/2581)
	AI→OD	82.2%	(1709/2078)	66.2%	(1709/2581)

Pair extraction: From Table 4, we can see that carrying out aspect identification before pairedness determination outperforms the reverse ordering by 11% in precision and 3% in recall. This result supports our expectation that knowledge of aspect information contributes to aspect-evaluation pair extraction. Focusing on the rows labeled "(dependency)" and "(no dependency)" in Table 4, while 80% of the aspect-evaluation pairs in a direct dependency relation are successfully extracted with high precision, the model achieves only 51.7% recall with 61.7% precision for the cases where an aspect and evaluation are not in a direct dependency relation.

According to our error analysis, a major source of errors lies in the aspect identification task. In this experiment, the precision of aspect identification is 78%. A major reason for this problem was the coverage of the dictionary. In addition, the system causes a false decision the aspect appears in the preceding sentence. We need to conduct further investigations in order to resolve these problems.

Opinionhood determination: Table 4 shows that carrying out aspect identification followed by opinionhood determination outperforms the reverse ordering, which supports our expectation that knowing the aspect information helps opinionhood determination.

While it produces better results, our proposed method still has room for improvement in both precision and recall. Our current error analysis has not identified particular error patterns — the types of errors are very diverse. However, we need to address the issue of modifying the feature set to make the model more sensitive to modality-oriented distinctions such as subjunctive and conditional expressions.

Subject detection

As mentioned in the method for opinion extraction, we have so far put aside the task of filling the *Subject* slot assuming that it is not a bottle-neck problem. Here, we provide a piece of evidence for this assumption by briefly reporting on the results of another experiment.

For the experiment, we created a corpus annotated with subject-evaluation pairs. The corpus consisted of 308 we-blog articles in the automobile domain (3,037 sentences) containing 870 subject-evaluation pairs.

We assumed that for each given article, all the subject expressions and evaluation expressions had been properly identified. The task was to identify the subject corresponding to a given evaluation expression. For this task, we implemented simple heuristics as follows:

1. If there are any subject expressions preceding the given evaluation expressions, choose the nearest one to the evaluation

2. Otherwise, choose the first one of those following the evaluation expression

The precision was 0.92 (822/890), and the recall was 0.94 (822/870). A major error was that the heuristics could not appropriately handle opinions that exhibited a comparison between a subject and its counterpart. However, this problem was not a big deal in terms of frequency. The results suggest that the problem of identifying subject-evaluation pairs is solvable with reasonably high precision and recall provided that subject expressions are properly identified. Subject expression identification is a subclass of named entity recognition, which has been actively studied for a decade. We are planning to incorporate state-of-the-art techniques for named entity recognition to the overall opinion mining system we are new developing.

Conclusion

In this paper, we proposed to regard the task of opinion extraction as extraction of triplets ⟨Subject, Aspect, Evaluation⟩. We proposed a machine learning-based method for the extraction of opinions on consumer products by reducing the problem to that of extracting aspect-evaluation

pairs from texts. The experimental results show that identifying the corresponding aspect for a given evaluation expression is effective in both pairedness determination and opinionhood determination.

We have so far considered the approach relies on the dictionaries in detecting evaluation and aspect candidates. However, the aspect expressions are heavily depend on the domain, while the evaluation phrases are not so domain dependent. Therefore, we are exploring an approach which does not use the aspect dictionary.

As the next step, we have started to undertake the task of identifying the relations between aspects. This task is to identify part-of, role-of, and aspect-of relations between aspects. For example, there is a aspect-of relation "sound–engine" in "the sound of the engine". For this task, we have already built an opinion-tagged corpus includes the relation informations between aspects, and started the experiment. The results of the experiments will be reported elsewhere.

References

Dave, K.; Lawrence, S.; and Pennock, D. M. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proc. of the 12th International World Wide Web Conference*, 519–528.

Gamon, M.; Aue, A.; Corston-Oliver, S.; and Ringger, E. K. 2005. Pulse: Mining customer opinions from free text. In *Proceedings of the 2005 Conference on Intelligent Data Analysis* (*IDA*), 121–132.

Hu, M., and Liu, B. 2004. Mining and summarizing customer reviews. In *Proc. of the Tenth International Conference on Knowledge Discovery and Data Mining*, 168–177.

Iida, R.; Inui, K.; Takamura, H.; and Matsumoto, Y. 2003. Incorporating contextual cues in trainable models for coreference resolution. In *Proc. of the EACL Workshop on the Computational Treatment of Anaphora*, 23–30.

Kanayama, H., and Nasukawa, T. 2004. Deeper sentiment analysis using machine translation technology. In *Proceedings of the 20th International Conference on Computational Linguistics*, 494–500.

Kobayashi, N.; Inui, K.; Matsumoto, Y.; Tateishi, K.; and Fukushima, T. 2004. Collecting evaluative expressions for opinion extraction. In *Proc. of the 1st International Joint Conference on Natural Language Processing*, 584–589.

Murano, S., and Sato, S. 2003. Automatic extraction of subjective sentences using syntactic patterns. In *Proc. of the Ninth Annual Meeting of the Association for Natural Language Processing*, 67–70. (in Japanese).

Pang, B., and Lee, L. 2004. A sentiment education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics*, 271–278.

Popescu, A.-M., and Etzioni, O. 2005. Extracting product features and opinions from reviews. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), 339–346.

Tateishi, K.; Ishiguro, Y.; and Fukushima, T. 2001. Opinion information retrieval from the internet. In *IPSJ SIGNL Note 144-11*, 75–82. (in Japanese).

Turney, P. D. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. of the 40th Annual Meeting of the Association for Computational Linguistics*, 417–424.