

Extracting Condition-Opinion Relations Toward Fine-grained Opinion Mining

Yuki Nakayama

Atsushi Fujii

Department of Computer Science
Graduate School of Information Science and Engineering
Tokyo Institute of Technology
{nakayama.y.aj@m, fujii@cs}.titech.ac.jp

Abstract

A fundamental issue in opinion mining is to search a corpus for opinion units, each of which typically comprises the evaluation by an author for a target object from an aspect, such as “This hotel is in a good location”. However, few attempts have been made to address cases where the validity of an evaluation is restricted on a condition in the source text, such as “for traveling with small kids”. In this paper, we propose a method to extract condition-opinion relations from online reviews, which enables fine-grained analysis for the utility of target objects depending the user attribute, purpose, and situation. Our method uses supervised machine learning to identify sequences of words or phrases that comprise conditions for opinions. We propose several features associated with lexical and syntactic information, and show their effectiveness experimentally.

1 Introduction

Reflecting the rapid growth in the use of opinionated texts on the Web, such as comments on news articles and customer reviews, opinion mining has been explored to facilitate utilizing opinions mainly for improving products and decision-making purposes. While in a broad sense opinion mining refers to a process to discover useful knowledge latent in a corpus of opinionated texts, fundamental issues involve modeling an unit of opinions and searching the corpus for those units, each of which typically comprises the evaluation by an author for a target object from an aspect. Other elements, such as when the opinion was submitted, can optionally be included in an opinion unit. We take the following review sentence as an example opinionated description.

(1) I think hotel A offers a reasonable price if you take a family trip with small kids.

From the above example, existing methods (Pang and Lee, 2008; Seki et al., 2009; Jin et al., 2009; Zhao et al., 2010; He et al., 2011; Liu and Zhang, 2012; Liu et al., 2013; Yang and Cardie, 2013; Liu et al., 2014) are intended to extract the following quintuple as an opinion unit.

Target = “hotel A”, Aspect = “price”,
Evaluation (Polarity) = “reasonable”
(positive), Holder = “I (author)”, Time
= N/A

Depending on the application, “Evaluation” can be any of a literal opinion word (e.g., “reasonable”), a polarity (positive/negative), or a value for multi-point scale rating.

Given those standardized units extracted from a corpus, it is feasible to overview the distribution of values for each element or a combination of elements. For example, those who intend to improve the quality of hotel A may investigate representative values for “Aspect” in the units satisfying “Target=hotel A & Polarity=negative”, while those who look for accommodation may collect the opinion units for one or more candidate hotels and investigate the distribution of values for “Polarity” on an aspect-by-aspect basis.

However, in the above example (1), the evaluation for hotel A (“a reasonable price”) is valid for “if you take a family trip with small kids”, and thus it is not clear whether this evaluation is valid irrespective of the condition. For example, the price may not be reasonable for a single customer intending for business purposes. In this paper, we shall call such a condition “condition for opinion (CFO)”. We define CFO as a condition for which an opinion unit has a polarity.

The existing methods for opinion mining, which do not consider whether a target opinion is conditional, potentially overestimate or underestimate

the utility of hotel A and consequently decrease the quality of opinion mining. We manually analyzed the first 7000 sentences in the Rakuten Travel data, which consists of 348564 Japanese reviews for hotels in Japan (see Section 4 for details of this data) and found that 2272 sentences are opinions, of which 630 opinions are conditional and thus the result for an existing method includes up to 28% (630/2272) errors.

Motivated by the above discussion, in this paper we propose a method to extract pairs of a CFO and its corresponding opinion unit from online reviews. This method provides two solutions to the above problem. First, a passive solution is detecting whether an opinion includes a CFO and, if any, isolating that opinion from the target of opinion mining. As a result, we can avoid potential errors as much as possible but the coverage is decreased.

Second, an active solution is identifying the span of each CFO in conditional opinions and classify them according to semantic categories, such as purpose, situation, and user attribute so that finer-grained opinion mining can be realized. For example, the distribution of positive and negative opinions can be available on a category-by-category basis. However, in this paper we focus only on the identification for CFOs and leave the semantic classification future work.

To produce a practical model for CFOs, it is important to investigate them from a grammar point of view. It can easily be predicted that a typical grammatical unit for CFOs is a conditional clause as in example (1). Additionally, restrictive modifiers in general can potentially be CFOs because they restrict the validity of an opinion unit from a specific perspective. A restrictive modifier comprises a word, phrase, or clause. The CFO in example (1), which is a dependent clause functioning as a condition, is also a restrictive modifier.

Example (2), which has the same meaning as example (1), includes a CFO as a prepositional phrase.

- (2) Hotel A offers a *reasonable* price **for taking a family trip with small kids**.

We denote CFOs and opinion words in bold and italic faces, respectively. Examples (3) and (4) also include a CFO as a prepositional phrase. Unlike example (2), the validity of “reasonable” is restricted from time and comparison points of view, respectively.

- (3) Hotel A offers a *reasonable* price **during this holiday season**.

- (4) Hotel A offers a *reasonable* price **for a four star hotel**.

In example (5), which has a similar meaning to example (1), the CFO is a dependent clause functioning as a reason.

- (5) Hotel A offers a *reasonable* price **because we take a family trip with small kids**.

Finally, as in example (6), an opinion holder can also be a CFO because the evaluation is restricted from a perspective of that specific person.

- (6) **My mother** regarded hotel A as a *reasonable* choice.

If the restriction by a CFO is associated with a user-related perspective, we call such CFOs “user-restrictive CFOs (U-CFOs)”. In other words, target users to whom an opinion unit is relevant are restricted by its corresponding U-CFO, although those users may agree or disagree with the opinion. The CFOs in examples (1), (2), and (5) are U-CFOs because the target users are mainly those who intend to travel with their children.

The CFO in example (3) is also U-CFO because the target users are those who intend to travel during a specific holiday season. The CFO in example (6) is also U-CFO because the opinion holder (“my mother”) implies the opinion is relevant mainly to adult females. However, opinion holders who do not represent user-related perspectives, such as “I” without any profile, are not U-CFOs.

The CFO in example (4) is not a U-CFO because the relevance of the opinion is not restricted to specific customers. It may be argued that in example (4) the target users are restricted to those who are interested in the price. However, in example (4) the price restricts the aspect of the opinion unit, and should not be confused with U-CFOs and even CFOs, which restrict the validity of the opinion unit.

If we fully utilize U-CFOs, as discussed for the active solution above, we need to classify U-CFOs into semantic categories so that users can selectively read relevant opinions. In other words, the identification for U-CFOs facilitates predicting the review helpfulness (O’Mahony and Smyth, 2010; Moghaddam et al., 2012). Candidate categories

include demographic and psychographic attributes for target users (e.g., age and hobby) and situations of target users (e.g., purpose, time, and place). However, we leave the classification for U-CFOs future work.

2 Related work

As described in Section 1, the fundamental methods for opinion mining include opinion extraction, which identifies elements for opinion units (i.e., target, aspect, evaluation, holder, and time) (He et al., 2011; Jin et al., 2009; Liu et al., 2013; Seki et al., 2009; Yang and Cardie, 2013; Zhao et al., 2010), and opinion classification, which determines the non-literal evaluation of each opinion unit based on bipolar categories (i.e., positive and negative) (He et al., 2011; Meng et al., 2012) or multipoint scale categories (Fu and Wang, 2010; Moghaddam and Ester, 2013). However, none of these methods intends to determine whether or not an opinion is conditional and to extract their condition.

Narayanan et al. (2009) proposed a method for sentiment classification targeting conditional sentences. Although a conditional opinion is a kind of conditional sentence, their research is fundamentally different from our research. Narayanan et al. (2009) targeted such a conditional sentence that comprises a single opinion as a whole, and intended to categorize its polarity into any of positive, negative, or neutral. Examples (7) and (8) are such conditional sentences associated with neutral and positive categories, respectively.

- (7) Hotel A would not have survived if the price was not reasonable.
- (8) If you are looking for a hotel with a reasonable price, stay at hotel A.

In example (7), although the subordinate clause includes the opinion word “reasonable”, none of the subordinate clause, main clause, or entire sentence is an opinion. In example (8), the entire sentence is an unconditional opinion about the price for hotel A, but the main and subordinate clauses are not opinions independently. In contrast, the purpose of our research is to identify conditional opinions, in which the main and subordinate clauses are an opinion and its condition, respectively.

Kim and Hovy (2006) proposed a method to identify a reason for the evaluation in an opinion,

such as “the service was terrible because the staff was rude”. Although as discussed in Section 1 reasons can be CFOs, their purpose is to identify grounds that justify the evaluation and thus is different from our purpose.

As discussed in Section 1, our research is related to predicting the review helpfulness (O’Mahony and Smyth, 2010; Moghaddam et al., 2012). The method proposed by O’Mahony and Smyth (2010) determines the helpfulness of a product review independent of the user profile and thus cannot recommend reviews based on user-related attributes.

Moghaddam et al. (2012) used collaborative filtering to predict the review helpfulness. The evaluation by a target user for past reviews is used to model the user and predict the helpfulness for unread reviews, which results in different predictions depending on the user. An advantage of collaborative filtering is its applicability to items whose content is usually difficult to analyze, such as videos. However, this advantage is diluted in recommending review text, from which effective features for user modeling, such as U-CFOs, can be obtained by opinion mining.

3 Proposed method

The task in this paper is to extract condition-opinion relations from reviews in Japanese. Currently, we assume that an opinion unit and its corresponding CFO are in the same sentence, and thus perform the extraction on a sentence-by-sentence basis. Given a sentence in reviews, we first search for an opinion unit, and if found, we also search for its corresponding CFO. Because in the first process we rely on an existing method for the opinion extraction, in this paper we focus only on the extraction for CFOs.

As discussed in Section 1, because CFOs can be different grammatical units, their length and structure are not standardized. We model the extraction for CFOs as the BIO chunking, which labels each token in a sentence as being the beginning (B), inside (I), or outside (O) of a span of interest. We use “Other” to refer to “O” to avoid confusion between “O” and “0” (zero). To subdivide “B” and “I” into U-CFOs and other CFOs, we use suffixes “U” and “C”, respectively, such as “BU” denoting the beginning of a U-CFO. We use “Cond” to refer to any of BU, IU, BC, or IC.

Because we use the same method for both U-

CFOs and other CFOs, the above distinction only increases the number of categories to which each token is classified. If the distinction of U-CFOs is not important, the above suffixes can be omitted.

We regard Japanese *bunsetsu* phrases, which consist of a content word and one or more postpositional particles, as tokens, and extract a sequence of a BU-phrase and one or more IU-phrases, or an independent BU-phrase as a condition. The same method is used for BC/IC-phrases. However, words and phrases in an opinion unit are classified into its corresponding element. For example, an aspect phrase is classified into the aspect category.

Given an input sequence of *bunsetsu* phrases, $\mathbf{x} = x_1 \dots x_n$, our task is to predict a sequence of labels, $\mathbf{y} = y_1 \dots y_n$, where $y_i \in \{BU, IU, BC, IC, Other, Target, Aspect, OpinionWord\}$. However, because an opinion unit in an input sentence has been identified in advance, the task is a quinary classification with respect to $y_i \in \{BU, IU, BC, IC, Other\}$. We use Conditional Random Fields (CRF) (Lafferty et al., 2001) to train a classifier for categorizing each *bunsetsu* phrase into any of the aforementioned five categories. We use a combination of unigram and bigram models and calculate the conditional probability, $p(\mathbf{y}|\mathbf{x})$, for linear-chain CRF by Equation (1).

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z_{\mathbf{x}}} \exp\left(\sum_{i,k} \lambda_k \cdot f_k(y_i, \mathbf{x}) + \sum_{i,k} \mu_k \cdot g_k(y_{i-1}, y_i, \mathbf{x})\right) \quad (1)$$

Here, $Z_{\mathbf{x}}$ denotes a normalization factor, and f_k and g_k denote feature functions for unigram and bigram models, respectively. Let $x_{i,v}$ denote a feature value for x_i . While in the unigram model y_i depends on either $x_{i-1,v}$ or $x_{i,v}$, in the bigram model y_i depends on either a combination of $x_{i,v}$ and y_{i-1} or that of $x_{i-1,v}$ and y_{i-1} . Feature functions are produced for any possible combinations of the values for the variables used ($x_{i,v}$, y_{i-1} , and y_i in f_k), and take 1 if the corresponding combination appears and 0 otherwise. We use the four combinations “unigram $x_{i,v}$ ”, “unigram $x_{i-1,v}$ ”, “bigram $y_{i-1} x_{i,v}$ ”, and “bigram $y_{i-1} x_{i-1,v}$ ” for feature functions.

The question here is how CFOs and U-CFOs can be modeled and what kind of features are needed. We assume characteristics of CFOs, and U-CFOs and partially exemplify their validity us-

ing Figure 1, which depicts an example input sentence and information related to its constituent *bunsetsu* phrases. In the upper part of Figure 1, a rectangle and an arrow denote a *bunsetsu* phrase and a syntactic dependency between two phrases, respectively, and in each phrase we show Japanese words based on the Hepburn system and their English translations in parentheses.

CFOs are associated with the following characteristics.

- (a) By definition, CFOs determine the validity of the evaluation in an opinion unit, and thus syntactically modify an opinion word. Consequently, CFOs usually do not modify other elements in an opinion unit, such as an aspect.
 - (b) Like a conjunction in a conditional clause in English, such as “if”, a CFO in Japanese also includes a clue expression, which is usually a functional expression (Matsuyoshi et al., 2006) in the tail phrase, such as “ni wa (“for” in English)”.
 - (c) The distribution for parts of speech as the head of CFOs is skewed and heads of CFOs are usually a noun or verb.
- Additionally, U-CFOs are associated with the following characteristics.
- (d) If a CFO is an opinion holder as in example (6) in Section 1, it is usually a U-CFO, which is the subject appearing at the beginning of a target sentence.
 - (e) By definition, U-CFOs include expressions related to user attributes, such as “nervosity” in Figure 1.

In this paper, we propose thirteen features to model CFOs and U-CFOs. In the bottom part of Figure 1, for each phrase we show the values of the thirteen features F1–F13 described below. These features were developed for the above five characteristics. F1–F5, F7–F10 and F13 are associated with (a), (b) and (c), respectively, while F6 and F11–F12 are associated with (d) and (e).

F1: Dependency distance to opinion word
CFOs, which affect the evaluation in that opinion, usually syntactically modifies the opinion word. Thus, there should be a pass of dependencies between a Cond-phrase and the opinion word, and a

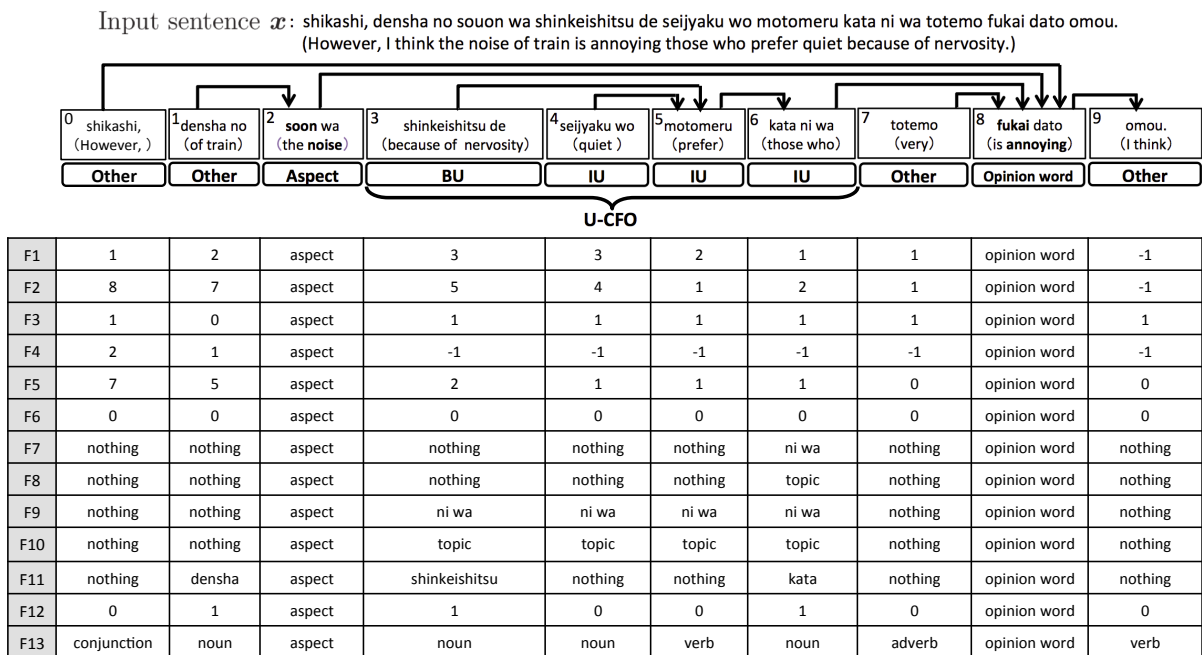


Figure 1: Example of Japanese sentence and the feature value for each constituent *bunsetsu* phrase

phrase that leads to the opinion word via a smaller number of dependency arrows is more likely to be a Cond-phrase. We use the dependency distance (i.e., the number of dependencies) between a phrase in question and the opinion word as the value for feature F1. The value for a phrase is -1 if there is no pass between that phrase and the opinion word. We use “CaboCha” (Kudo and Matsumoto, 2002) for dependency analysis purposes.

F2: Phrase distance to opinion word F1 is not robust against errors of the dependency analysis. To alleviate this problem, we approximate the dependency distance by a phrase distance. In practice, we subtract the ID for a phrase in question from that for the opinion word as the value for feature F2. If the opinion word consists of more than one phrase, we take the minimum difference. Because in Japanese a modifier is usually followed by its modifying object, a phrase with a negative value for feature F2 is usually an Other-phrase. For example, in the last phrase in Figure 1, which cannot be a modifier for the opinion word, is an Other-phrase.

F3: Dependency pass to aspect Because a CFO rarely modifies an aspect, for the value of feature F3 we take 0 if there is a pass of dependencies between a phrase in question and an aspect and 1 otherwise.

F4: Phrase distance to aspect Similar to F1, F3 is not robust against errors of the dependency analysis. As in F2, we approximate the value of F4 by a phrase distance between a phrase including an aspect and a phrase in question.

F5: Difference between values for F2 and F1 A CFO usually consists of a sequence of Cond-phrases where each phrase modifies the next phrase, as in Figure 1. There is a tendency that as the difference of values of F1 and F2 for a phrase becomes smaller, that phrase is more likely to be a Cond-phrase. In Figure 1, the values for Cond-phrases #3–#6 are smaller than those for Other-phrases #0–#1.

F6: Beginning of sentence The subject of an opinion sentence is often its U-CFO because the evaluation is valid only from the perspective of that specific subject. For example, in “my daughter was pleased with toys in the room” the positive evaluation is restricted by the daughter’s perspective. Thus, the value of feature F6 takes 1 for the first phrase in a sentence excluding a conjunction, and 0 otherwise.

F7: Clue expression Because a CFO often ends with one or more specific particles or auxiliary verbs, we use the existence of those clue expressions in a phrase as the value for feature F7. We use words in a dictionary of Japanese functional expressions “Tsutsuji” (Matsuyoshi et al., 2006)

as the clue expressions. Table 1 shows examples of entries for Tsutsuji. Each entry is represented in a hierarchy structure with nine abstraction levels. We firstly collected “Head words” in the nineteen categories (e.g., resultative condition and purpose in L2) associated with our purpose, consulting “Meaning categories”. Then we collected “Surface forms” corresponding to the collected head words and identified their corresponding surface forms to standardize different forms. For example, for ID 1 and ID 3 in Table 1, “to sure ba” and “nde” are regarded as identical to “to suru to” and “node”, respectively. As a result, we collected 388 words, such as “ba (if)” and “ni (for)” and used their existence in a phrase in question as value for F7.

F8: Semantic categories for clue expression

Because the data sparseness is a crucial problem for F7, we use the existence of semantic categories in Tsutsuji as the values of F8 for smoothing purposes. For example, in Table 1, “to suru to” and “ba” have the same feature values “resultative condition”. If a clue expression belongs to more than one semantic category as in “ni” of Table 1, the feature value is a set of these categories.

F9: Dependency pass to phrase including clue expression (Surface form)

As described in F7 above, the last phrase in a CFO often includes one or more clue expressions. In addition, a CFO often consists of more than one phrase. Given those conditions, a phrase that modifies a phrase containing a clue expression is also likely to be a Cond-phrase. We use the existence of a dependency pass between a phrase in question and a phrase containing a clue expression as the values of feature F9.

F10: Dependency pass to phrase including clue expression (Category)

As with F8, we use the existence of semantic categories of Tsutsuji as the values of feature F10.

F11: Restrictive words

We use the existence of words that are strongly associated with U-CFO as the value for F11. We call such words restrictive words. We automatically produced a dictionary of restrictive words from advertising slogans for hotels, which often include descriptions for target users, such as “★joshikai ya kappuru ni osusume!!★ (Recommended to girls get-together and couples)”. First, we extracted words in the advertising slogan based on the following steps.

Entry ID	Abstraction levels			
	L1: Head word	L2: Meaning categories	...	L9: Surface forms
1	to suru to	resultative condition	...	to sure ba
2	ba		...	ba
3	node	reason	...	nde
4	ni	purpose	...	ni
5		target	...	

Table 1: Example entries for Tsutsuji

Step 1: Extracting sentences that match to a regular expression “(| hito | mono | kata) ni (| wa | mo) osusume” (i.e., “recommended to” or “recommend to those who”).

Step 2: Collecting a sequence of content words for each bunsetsu-phrase in the extracted sentences.

For the above advertising slogan, we can collect two restrictive words “joshikai (girls get-together)” and “kappuru (couple)” by performing those 2 steps.

Second, we collected a sequence of independent words for bunsetsu phrases which comprises U-CFO in an annotated corpus. We combined the extracted words from the advertising slogans an annotated corpora, discarded redundancy, and standardized similar words, such as “kanko suru (do sightseeing)” and “kanko (sightseeing)”. As a result, we collected 934 words.

Finally, we calculated a mutual information like score, $Score(r, u)$, between a restrictive word r and labels u , Cond-phrases for U-CFOs (i.e., phrases labeled with either of BU or IU), by Equation 2.

$$Score(r, u) = P(r, u) \log \frac{P(r, u)}{P(r)P(u)} \quad (2)$$

$P(r, u)$ denotes the probability that a phrase including r is labeled with BU or IU in the annotated corpus. $P(r)$ denotes the probability that a phrase including r appears in the annotated corpus while $P(u)$ denotes the probability that a phrase labeled with BU or IU in the annotated corpus. If a phrase includes a restrictive word r and $Score(r, u)$ is greater than threshold θ , the feature value is r , and “nothing” otherwise.

F12: Existence of restrictive word Because the data sparseness is a crucial problem for F11, we integrate all the restrictive words for F11 into a single category for smoothing purposes. The value for F12 is the existence (1/0) of restrictive words.

F13: Part of speech for head The likelihood that a phrase in question is a Cond-phrases partially depends on the part of speech for the head in that phrase. For example, in Figure 1, a phrase whose head is a noun or verb tends to be a Cond-phrase

4 Experiments

To evaluate the effectiveness of our method, we used the Rakuten Travel data¹, which consists of 348 564 Japanese reviews for hotels in Japan. From this dataset, we selected 580 reviews and manually identified elements for opinion units. We removed sentences consisting only of opinion unit such as “The location is good” from the evaluation. As a result, 3 155 sentences remained, which comprise our corpus. To evaluate the effectiveness for identifying CFOs, we used the manually annotated opinion elements as output of a pseudo automatic method.

Given the above corpus, two annotators independently identified U-CFOs or CFOs, if any, for each opinion unit. For both annotations of CFOs and U-CFOs, the Kappa value for the inter-annotator agreement was 0.87, indicating strong agreement. We show the details of our corpus in Table 2. Using this corpus, we performed 10-fold cross-validation and compared different methods from different perspectives. Also, we determined the threshold for Score (see Eq 2) by a development set for each fold.

To evaluate the effectiveness of extracting U-CFOs and CFOs independently, we first classified *bunsetsu* phrases into any of BU, IU, BC, IC, or Other. Then, for the U-CFO extraction we regarded phrases for BU and IU as the Cond-phrases while for the CFO extraction we regarded phrases for BU, IU, BC, and IC as the Cond-phrases.

We used “Partial match” and “Exact match”, which denote different criteria for the correctness of methods under evaluation. While in the partial match each method was requested to only detect whether or not a test sentence includes CFO, in the exact match each method was also requested to identify the span of each CFO. Also, we used different evaluation measures, namely precision (P), recall (R), F-measure (F), and accuracy (A).

Rule-based method and SVM-based method are used for comparison purposes. Rule-based

method first identifies a *bunsetsu* phrase whose dependency distance to the opinion word is 1 and including a clue expression (see Section 3), and also identifies a sequence of the phrases from which there is a dependency path to the above phrase as a CFO. For example, in Figure 1 because phrase #6 includes a clue expression, a sequence of phrases #3–#6 is extracted as a CFO. These rules are based on features F1, F7 and F9. For the U-CFO extraction task, we regarded a sequence of Cond-phrases extracted by the above method as U-CFO if that sequence includes a restrictive word. For SVM, the thirteen features F1–F13 proposed in Section 3 was used. We used LIBSVM (Chang and Lin, 2011) to train a classifier. Our method used CRF to train a classifier with the thirteen features and four patterns for feature functions. We used CRF++² to train a classifier for each phrase and regularized the parameters using L2-norm.

Figure 2 shows the relationship between values of regularization parameter and F-measure for exact match. In Figure 2, “Rule”, “SVM”, and “CRF” denote a rule-based method, SVM-based method, and our method, respectively. The F-measure for Rule, independent of the regularization parameter, is a constant. While the F-measure for SVM substantially varied depending on the parameter value, that for CRF did not vary that much. Additionally, the F-measure for CRF was larger than that for SVM irrespective of the parameter value and matching criterion.

Table 3 shows results obtained with the optimal value for the regularization parameter. Looking at Table 3, one can see that CRF outperformed the other methods in terms of F-measure and accuracy for both partial and exact matches. We used the two-tailed paired t-test for statistical testing and found that the differences of CRF and each of the other methods in F-measure and accuracy were statistically significant at the 1% level irrespective of the configuration.

Figure 3 shows the effectiveness of the proposed features for exact match. The horizontal axis “w/o X” denotes a method without feature X. The vertical axis denotes a ratio of each method to our method. If a method without feature X takes less than 1 for value of vertical axis, the feature X is effective for extracting CFOs. Looking at Figure 3, one can see that our complete method outperformed any variation of our method in terms of

¹<https://alaginrc.nict.go.jp/resources/rakuten-dataset/rakuten-outline.html>

²<http://crfpp.googlecode.com/svn/trunk/doc/index.html>

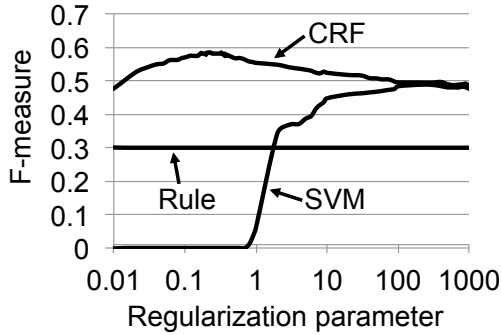
Opinion sentence	# w/ CFO	799
	# w/ U-CFO	526
	# w/o CFO	1 257
# Non opinion sentence		1 099
# Total		3 155

(a) Sentence unit

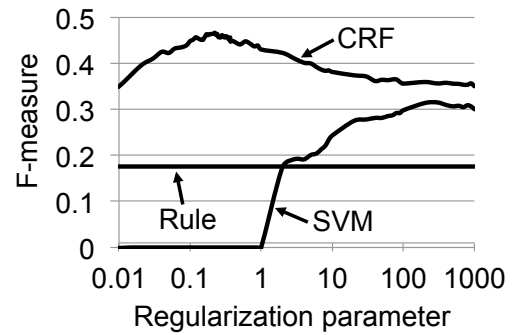
# BU-phrase		571
# IU-phrase		741
# BC-phrase		307
# IC-phrase		632
# Other-phrase		16 584
Opinion unit	# Opinion word	3 764
	# Aspect	3 406
	# Target	132
# Total		26 137

(b) Phrase unit

Table 2: Details of our corpus



(a) CFO extraction



(b) U-CFO extraction

Figure 2: Relationship between values for regularization parameter and F-measure in exact match

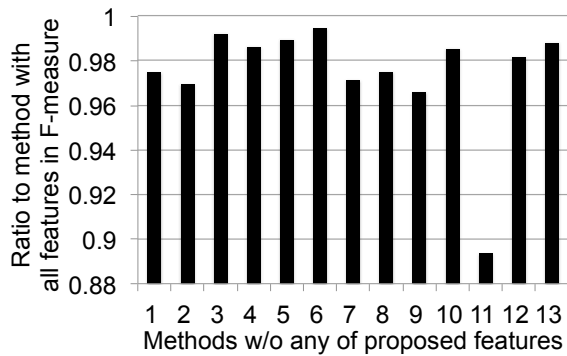
	Partial match				Exact match			
	P	R	F	A	P	R	F	A
Rule	.576	.790	.664	.771	.294	.319	.300	.649
SVM	.779	.842	.809	.885	.490	.500	.488	.789
CRF	.889	.758	.818	.915	.592	.580	.583	.865

(a) CFO extraction

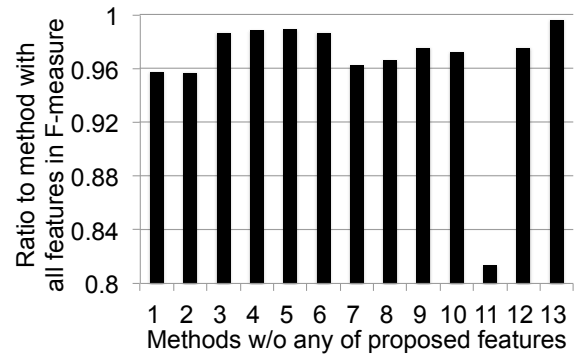
	Partial match				Exact match			
	P	R	F	A	P	R	F	A
Rule	.395	.692	.502	.740	.174	.187	.176	.663
SVM	.622	.703	.659	.863	.319	.323	.315	.800
CRF	.818	.643	.720	.917	.472	.462	.464	.892

(b) U-CFO extraction

Table 3: Results for different configurations (P: precision, R: recall, F: F-measure, A: accuracy)



(a) CFO extraction



(b) U-CFO extraction

Figure 3: Effectiveness of proposed features for exact match

F-measure. Thus, we conclude that each of our thirteen features was independently effective for extracting CFO and U-CFO in review sentences and that when used together the improvement was even greater.

For the U-CFO extraction, we analyzed the errors by our method. The total number of errors was 363 by condition unit. We describe causes of the errors with example sentences, translated into English by the authors. In those examples, double and single underlines denote false positive and false negative, respectively. For each cause, we show the number of errors in parentheses.

E1 (124) Errors were due to F11 and F12 with insufficient dictionary for restrictive words. Typically, low frequency words (e.g., pilgrimage) and words related to miscellaneous activities during a travel (e.g., charging a battery of a mobile phone) were not included in our dictionary. While it is important to increase the vocabulary size of our dictionary, identifying synonymous expressions with partial matching (e.g., go to sleep / go to bed) is also important.

E2 (53) Errors were due to dependency analysis, which often mistakenly recognizes sentence boundaries in an informal writing style and dependency relations in a sentence comprising a phrase, such as “the best location for fully enjoying Asakusa”. In this example, CaboCha mistakenly associated the adnominal modifier “for fully enjoying Asakusa” with “location (aspect)” instead of “best (opinion word)”. As a result, F1 and F3 did not regard this modifier as a U-CFO.

E3 (40) Restrictive modifiers that modify a non-opinion segment were mistakenly extracted as U-CFOs. For example, in “I used this hotel for business and the meal was good”, “for business” includes the clue expression “for” but does not modify the opinion unit.

E4 (39) Similar to E3 but errors were due to restrictive words instead of clue expressions. In the example for E3, the restrict word “business” caused the error.

E5 (26) U-CFOs that consist of a large number of phrases were often not extracted due to F5, such as “This hotel is acceptable for one night to take the train at the Chuo station next morning”.

E6 (25) Errors were due to irrelevant entries in our restrictive word dictionary.

E7 (11) Due to the sparseness problem for restrictive words in the training data, U-CFOs and CFOs were not correctly distinguished.

E8 (9) Errors were due to part-of-speech tagging.

E9 (6) Errors were due to extracting modifiers consisting of a personal pronoun without additional user-related attributes, such as “enough for me”, as U-CFOs. We need to identify whether an expression for a person is associated with user-related attributes, such as “the bed is small for a person who is tall”, which indicates a physical attribute of a user.

Additionally, there are 65 errors for which we have not found a reason.

5 Conclusion

Although a number of methods have been proposed to search an opinionated corpus for opinion units, few attempts have so far been made at addressing cases where the validity of an evaluation is restricted on a condition in the source text. We proposed a method to identify such conditions from sentences including opinion units. Our method performs sequence labeling to determine whether each phrase is a constituent of an condition for opinion. We proposed thirteen features associated with lexical and syntactic information of Japanese, and showed their effectiveness using reviews for hotels. The contributions of this paper are introducing the notion of conditions for opinions, which is language-independent, proposing a method to extract condition-opinion relations from opinionated corpora, and giving an insight into its potential applications in opinion mining.

Acknowledgments

We would like to thank Professor Takenobu Tokunaga (Tokyo Institute of Technology) for his valuable comments. This research was supported in part by Grant-in-Aid for Scientific Research (Grant No. 15H02747).

References

- Chih-Chung Chang and Chih-Jen Lin. 2011. LIB-SVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27.
- Guohong Fu and Xin Wang. 2010. Chinese sentence-level sentiment classification based on fuzzy sets. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 312–319.
- Yulan He, Chenghua Lin, and Harith Alani. 2011. Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 123–131.
- Wei Jin, Hung Hay Ho, and Rohini K. Srihari. 2009. Opinionminer: A novel machine learning system for web opinion mining and extraction. In *Proceedings of the 15th ACM SIGKDD*, pages 1195–1204.
- Soo-Min Kim and Eduard Hovy. 2006. Automatic identification of pro and con reasons in online reviews. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 483–490.
- Taku Kudo and Yuji Matsumoto. 2002. Japanese dependency analysis using cascaded chunking. In *Proceedings of the 6th Conference on Natural Language Learning*, pages 1–7.
- John Lafferty, Andrew McCallum, and Fernando C.N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, pages 282–289.
- Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In C.C. Aggarwal and C.X.Zhai, editors, *Mining Text Data*, pages 415–463. Springer.
- Kang Liu, Liheng Xu, and Jun Zhao. 2013. Syntactic patterns versus word alignment: Extracting opinion targets from online reviews. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1754–1763.
- Kang Liu, Liheng Xu, and Jun Zhao. 2014. Extracting opinion targets and opinion words from online reviews with graph co-ranking. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 314–324.
- Suguru Matsuyoshi, Satoshi Sato, and Takehito Utsuro. 2006. Compilation of a dictionary of Japanese functional expressions with hierarchical organization. In Yuji Matsumoto, Richard Sproat, Kam-Fai Wong, and Min Zhang, editors, *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead*, volume 4285 of *Lecture Notes in Computer Science*, pages 395–402. Springer Berlin Heidelberg.
- Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu, and Houfeng Wang. 2012. Cross-lingual mixture model for sentiment classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 572–581.
- Samaneh Moghaddam and Martin Ester. 2013. The FLDA model for aspect-based opinion mining: Addressing the cold start problem. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 909–918.
- Samaneh Moghaddam, Mohsen Jamali, and Martin Ester. 2012. ETF: Extended tensor factorization model for personalizing prediction of review helpfulness. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pages 163–172.
- Ramanathan Narayanan, Bing Liu, and Alok Choudhary. 2009. Sentiment analysis of conditional sentences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 180–189.
- M.P. O’Mahony and B. Smyth. 2010. A classification-based review recommender. *Knowledge-Based Systems*, 23(4):323–329.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.
- Yohei Seki, Noriko Kando, and Masaki Aono. 2009. Multilingual opinion holder identification using author and authority viewpoints. *Information Processing and Management*, 45(2):189–199.
- Bishan Yang and Claire Cardie. 2013. Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1640–1649.
- Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li. 2010. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 56–65.