Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents

Nobuhiro Kaji and **Masaru Kitsuregawa**
Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505 Japan
{kaji,kitsure}@tkl.iis.u-tokyo.ac.jp

Abstract

Recognizing polarity requires a list of polar words and phrases. For the purpose of building such lexicon automatically, a lot of studies have investigated (semi-) unsupervised method of learning polarity of words and phrases. In this paper, we explore to use structural clues that can extract polar sentences from Japanese HTML documents, and build lexicon from the extracted polar sentences. The key idea is to develop the structural clues so that it achieves extremely high precision at the cost of recall. In order to compensate for the low recall, we used massive collection of HTML documents. Thus, we could prepare enough polar sentence corpus.

1 Introduction

Sentiment analysis is a recent attempt to deal with evaluative aspects of text. In sentiment analysis, one fundamental problem is to recognize whether given text expresses positive or negative evaluation. Such property of text is called polarity. Recognizing polarity requires a list of polar words and phrases such as 'good', 'bad' and 'high performance' etc. For the purpose of building such lexicon automatically, a lot of studies have investigated (semi-) unsupervised approach.

So far, two kinds of approaches have been proposed to this problem. One is based on a thesaurus. This method utilizes synonyms or glosses of a thesaurus in order to determine polarity of words (Kamps et al., 2004; Hu and Liu, 2004; Kim and Hovy, 2004; Esuli and Sebastiani, 2005). The second approach exploits raw corpus. Polarity is decided by using co-occurrence in a corpus. This is based on a hypothesis that polar phrases conveying the same polarity co-occur with each other. Typically, a small set of seed polar phrases are prepared, and new polar phrases are detected based on the strength of co-occurrence with the seeds (Hatzivassiloglous and McKeown, 1997; Turney, 2002; Kanayama and Nasukawa, 2006).

As for the second approach, it depends on the definition of co-occurrence whether the hypothesis is appropriate or not. In Turney's work, the co-occurrence is considered as the appearance in the same window (Turney, 2002). Although this idea is simple and feasible, there is a room for improvement. According to Kanayama's investigation, the hypothesis is appropriate in only 60% of cases if co-occurrence is defined as the appearance in the same window[1]. In Kanayama's method, the co-occurrence is considered as the appearance in intra- or inter-sentential context (Kanayama and Nasukawa, 2006). They reported that the precision was boosted to 72.2%, but it is still not enough. Therefore, we think that the above hypothesis is often inappropriate in practice, and this fact is the biggest obstacle to learning lexicon from corpus.

In this paper, we explore to use structural clues that can extract polar sentences from Japanese HTML documents, and build lexicon from the ex-

[1]To be exact, the precision depends on window size and ranges from 59.7 to 64.1%. See Table 4 in (Kanayama and Nasukawa, 2006) for the detail.

Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1075–1083, Prague, June 2007. ©2007 Association for Computational Linguistics

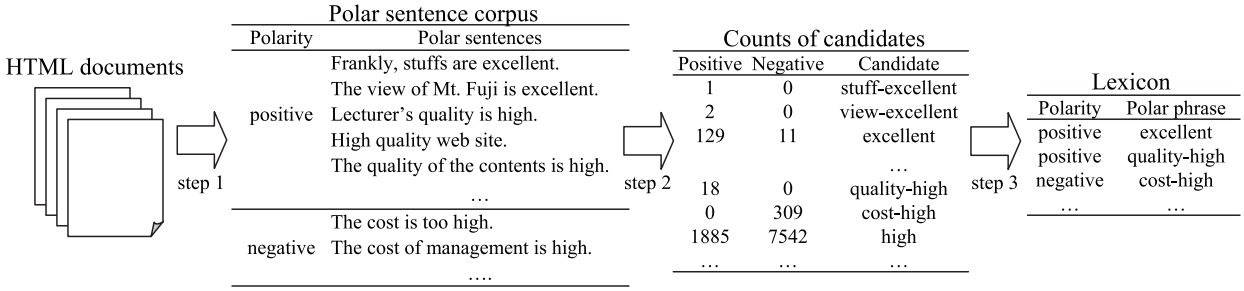


Figure 1: Overview of the proposed method.

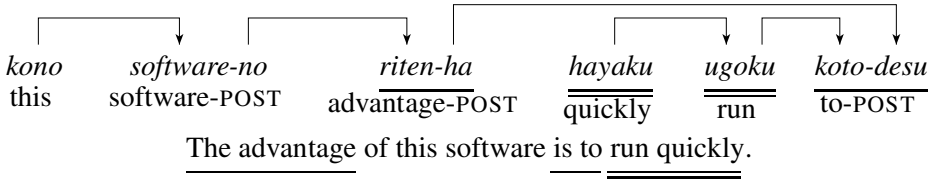


Figure 2: Language structure.

tracted polar sentences. An overview of the proposed method is represented in Figure 1. First, polar sentences are extracted from HTML documents by using structural clues (step 1). The set of polar sentences is called polar sentence corpus. Next, from the polar sentence corpus, candidates of polar phrases are extracted together with their counts in positive and negative sentences (step 2). Finally, polar phrases are selected from the candidates and added to our lexicon (step 3).

The key idea is to develop the structural clues so that it achieves extremely high precision at the cost of recall. As we will see in Section 2.3, the precision was extremely high. It was around 92% even if ambiguous cases were considered as incorrect. In order to compensate for the low recall, we used massive collection of HTML documents. Thus, we could build enough polar sentence corpus. To be specific, we extracted 500,000 polar sentences from one billion HTML documents.

The contribution of this paper is to empirically show the effectiveness of an approach that makes use of the strength of massive data. Nowadays, terabyte is not surprisingly large, and larger corpus would be obtained in the future. Therefore, we think this kind of research direction is important.

2 Extracting Polar Sentences

Our method begins by automatically constructing polar sentence corpus with structural clues (step 1).

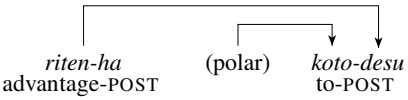
The basic idea is exploiting certain language and layout structures as clues to extract polar sentences. The clues were carefully chosen so that it achieves high precision. The original idea was represented in our previous paper (Kaji and Kitsuregawa, 2006).

2.1 Language structure

Some polar sentences are described by using characteristic language structures. Figure 2 illustrates such Japanese polar sentence attached with English translations. Japanese are written in italics and '-' denotes that the word is followed by postpositional particles. For example, '*software-no*' means that '*software*' is followed by postpositional particle '*no*'. The arrow represents dependency relationship. Translations are shown below the Japanese sentence. '-POST' means postpositional particle.

What characterizes this sentence is the singly underlined phrase. In this phrase, '*riten* (advantage)' is followed by postpositional particle '*-ha*', which is Japanese topic marker. And hence, we can recognize that something positive is the topic of the sentence. This kind of linguistic structure can be recognized by lexico-syntactic pattern. Hereafter, such words like '*riten* (advantage)' are called cue words.

In order to handle the language structures, we utilized lexico-syntactic patterns as illustrated below.



A sub-tree that matches (polar) is extracted as polar sentence. It is obvious whether the polar sentence is positive or negative one. In case of Figure 2, the doubly underlined part is extracted as polar sentence[2].

Besides '*riten* (advantage)', other cue words were also used. A list of cue words (and phrases) were manually created. For example, we used 'pros' or 'good point' for positive sentences, and 'cons', 'bad point' or 'disadvantage' for negative ones. This list is also used when dealing with layout structures.

2.2 Layout structure

Two kinds of layout structures are utilized as clues. The first clue is the itemization. In Figure 3, the itemizations have headers and they are cue words ('pros' and 'cons'). Note that we illustrated translations for the sake of readability. By using the cue words, we can recognize that polar sentences are described in these itemizations.

The other clue is table structure. In Figure 4, a car review is summarized in the table format. The left column acts as a header and there are cue words ('plus' and 'minus') in that column.

Pros:

- The sound is natural.
- Music is easy to find.
- Can enjoy creating my favorite play-lists.

Cons:

- The remote controller does not have an LCD display.
- The body gets scratched and fingerprinted easily.
- The battery drains quickly when using the backlight.

Figure 3: Itemization structure.

Mileage(urban)	7.0km/litter
Mileage(highway)	9.0km/litter
Plus	This is a four door car, but it's so cool.
Minus	The seat is ragged and the light is dark.

Figure 4: Table structure.

It is easy to extract polar sentences from the itemization. Such itemizations as illustrated in Figure 3 can be detected by using the list of cue words and HTML tags such as <h1> and <ul> etc. Three positive and negative sentences are extracted respectively from Figure 3.

As for table structures, two kinds of tables are considered (Figure 5). In the Figure, $+$ and $-$ represent positive and negative polar sentences, and C_+ and C_- represent cue words. Type A is a table in which the leftmost column acts as a header. Figure 4 is categorized into this type. Type B is a table in which the first row acts as a header.

Type A

C_+	$+$
C_-	$-$

Type B

	C_+	C_-	
	$+$	$-$	

Figure 5: Two types of table structures.

In order to extract polar sentences, first of all, it is necessary to determine the type of the table. The table is categorized into type A if there are cue words in the leftmost column. The table is categorized into type B if it is not type A and there are cue words in the first row. After the type of the table is decided, we can extract polar sentences from the cells that correspond to $+$ and $-$ in the Figure 5.

2.3 Result of corpus construction

The method was applied to one billion HTML documents. In order to get dependency tree, we used KNP[3]. As the result, 509,471 unique polar sentences were obtained. 220,716 are positive and the others are negative[4]. Table 1 illustrates some translations of the polar sentences.

[2]To be exact, the doubly underlined part is polar clause. However, it is called polar sentence because of the consistency with polar sentences extracted by using layout structures.

[3]http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html

[4]The polar sentence corpus is available from http://www.tkl.iis.u-tokyo.ac.jp/~kaji/acp/.

Table 1: Examples of polar sentences.

Polarity	Polar sentence
positive	It becomes easy to compute cost.
	It's easy and can save time.
	The soup is rich and flavorful.
negative	Cannot use mails in HTML format.
	The lecture is really boring.
	There is no impressive music.

In order to investigate the quality of the corpus, two human judges (judge A/B) assessed 500 polar sentences in the corpus. According to the judge A, the precision was 91.4%. 459 out of 500 polar sentences were regarded as valid ones. According to the judge B, the precision was 92.0% (460/500). The agreement between the two judges was 93.5% (Kappa value was 0.90), and thus we can conclude that the polar sentence corpus has enough quality (Kaji and Kitsuregawa, 2006).

After error analysis, we found that most of the errors are caused by the lack of context. The following is a typical example.

There is much information.

This sentence is categorized into positive one in the corpus, and it was regarded as invalid by both judges because the polarity of this sentence is ambiguous without context.

As we described in Section 1, the hypothesis of co-occurrence based method is often inappropriate. (Kanayama and Nasukawa, 2006) reported that it was appropriate in 72.2% of cases. On the other hand, by using extremely precise clues, we could build polar sentence corpus that have high precision (around 92%). Although the recall of structural clues is low, we could build large corpus by using massive collection of HTML documents. Of course, we cannot directly compare these two percentages. We think, however, the high precision of 92% implies the strength of our approach.

3 Acquisition of Polar Phrases

The next step is to acquire polar phrases from the polar sentence corpus (step 2 and 3 in Figure 1).

3.1 Counting candidates

From the corpus, candidates of polar phrases are extracted together with their counts (step 2).

As is often pointed out, adjectives are often used to express evaluative content. Considering that polarity of isolate adjective is sometimes ambiguous (e.g. high), not only adjectives but also adjective phrases (noun + postpositional particle + adjective) are treated as candidates. Adjective phrases are extracted by the dependency parser. To handle negation, an adjective with negation words such as 'not' is annotated by <NEGATION> tag. For the sake of readability, we simply represent adjective phrases in the form of 'noun-adjective' by omiting postpositional particle, as in the Figure 1.

For each candidate, we count the frequency in positive and negative sentences separately. Intuitively, we can expect that positive phrases often appear in positive sentences, and vice versa. However, there are exceptional cases as follows.

Although the price is high, its shape is beautiful.

Although this sentence as a whole expresses positive evaluation and it is positive sentence, negative phrase 'price is high' appears in it. To handle this, we hypothesized that positive/negative phrases tend to appear in main clause of positive/negative sentences, and we exploited only main clauses to count the frequency.

3.2 Selecting polar phrases

For each candidate, we determine numerical value indicating the strength of polarity, which is referred as polarity value. On the basis of this value, we select polar phrases from the candidates and add them to our lexicon (step 3).

For each candidate c, we can create a contingency table as follows.

Table 2: Contingency table

	pos	neg
c	$f(c, pos)$	$f(c, neg)$
$\neg c$	$f(\neg c, pos)$	$f(\neg c, neg)$

$f(c, pos)$ is the frequency of c in positive sentences. $f(\neg c, pos)$ is that of all candidates but c. $f(c, neg)$ and $f(\neg c, neg)$ are similarly decided.

From this contingency table, c's polarity value is determined. Two ideas are examined for compari-

son. One is based on chi-square value and the other is based on Pointwise Mutual Information (PMI).

Chi-square based polarity value The chi-square value is a statistical measure used to test the null hypothesis that, in our case, the probability of a candidate in positive sentences is equal to the probability in negative sentences. Given Table 2, the chi-square value is calculated as follows.

$$\chi^2(c) = \sum_{x \in (c, \neg c)} \sum_{y \in (pos, neg)} \frac{\{f(x,y) - \hat{f}(x,y)\}^2}{\hat{f}(x,y)}$$

Here, $\hat{f}(x,y)$ is the expected value of $f(x,y)$ under the null hypothesis.

Although $\chi^2(c) (\geq 0)$ indicates the strength of bias toward positive or negative sentences, its direction is not clear. We determined polarity value so that it is greater than zero if c appears in positive sentences more frequently than in negative sentences and otherwise it is less than zero.

$$PV_{\chi^2}(c) = \begin{cases} \chi^2(c) & \text{if} P(c|neg) < P(c|pos) \\ -\chi^2(c) & \text{otherwise} \end{cases}$$

$P(c|pos)$ is c's probability in positive sentences, and $P(c|neg)$ is that in negative sentences. They are estimated by using Table 2.

$$P(c|pos) = \frac{f(c, pos)}{f(c, pos) + f(\neg c, pos)}$$
$$P(c|neg) = \frac{f(c, neg)}{f(c, neg) + f(\neg c, neg)}$$

PMI based polarity value Using PMI, the strength of association between c and positive sentences (and negative sentences) is defined as follows (Church and Hanks, 1989).

$$PMI(c, pos) = \log_2 \frac{P(c, pos)}{P(c)P(pos)}$$
$$PMI(c, neg) = \log_2 \frac{P(c, neg)}{P(c)P(neg)}$$

PMI based polarity value is defined as their difference. This idea is the same as (Turney, 2002).

$$\begin{aligned} PV_{PMI}(c) &= PMI(c, pos) - PMI(c, neg) \\ &= \log_2 \frac{P(c, pos)/P(pos)}{P(c, neg)/P(neg)} \\ &= \log_2 \frac{P(c|pos)}{P(c|neg)} \end{aligned}$$

$P(c|pos)$ and $P(c|neg)$ are estimated in the same way as shown above. $PV_{PMI}(c)$ is (log of) the ratio of c's probability in positive sentences to that in negative sentences. This formalization follows our intuition. Similar to $PV_{\chi^2}(c)$, $PV_{PMI}(c)$ is greater than zero if $P(c|neg) < P(c|pos)$, otherwise it is less than zero.

Selecting polar phrases By using polarity value and threshold $\theta (> 0)$, it is decided whether a candidate c is polar phrase or not. If $\theta < PV(c)$, the candidate is regarded as positive phrase. Similarly, if $PV(c) < -\theta$, it is regarded as negative phrase. Otherwise, it is regarded as neutral. Only positive and negative phrases are added to our lexicon. By changing θ, the trade-off between precision and recall can be adjusted. In order to avoid data sparseness problem, if both $f(c, pos)$ and $f(c, neg)$ are less than three, such candidates were ignored.

4 Related Work

As described in Section 1, there have been two approaches to (semi-) unsupervised learning of polarity. This Section introduces the two approaches and other related work.

4.1 Thesaurus based approach

Kamps et al. built lexical network by linking synonyms provided by a thesaurus, and polarity was defined by the distance from seed words ('good' and 'bad') in the network (Kamps et al., 2004). This method relies on a hypothesis that synonyms have the same polarity. Hu and Liu used similar lexical network, but they considered not only synonyms but antonyms (Hu and Liu, 2004). Kim and Hovy proposed two probabilistic models to estimate the strength of polarity (Kim and Hovy, 2004). In their models, synonyms are used as features. Esuli et al. utilized glosses of words to determine polarity (Esuli and Sebastiani, 2005; Esuli and Sebastiani, 2006).

Compared with our approach, the drawback of using thesaurus is the lack of scalability. It is difficult to handle such words that are not contained in a thesaurus (e.g. newly-coined words or colloquial words). In addition, phrases cannot be handled because the entry of usual thesaurus is not phrase but word.

4.2 Corpus based approach

Another approach is based on an idea that polar phrases conveying the same polarity co-occur with each other in corpus.

(Turney, 2002) is one of the most famous work that discussed learning polarity from corpus. Turney determined polarity value[5] based on co-occurrence with seed words ('excellent' and 'poor'). The co-occurrence is measured by the number of hits returned by a search engine. The polarity value proposed by (Turney, 2002) is as follows.

$$\log_2 \frac{\mathrm{hits}(c\ \mathbf{NEAR}\ \mathrm{excellent})\mathrm{hits}(\mathrm{poor})}{\mathrm{hits}(c\ \mathbf{NEAR}\ \mathrm{poor})\mathrm{hits}(\mathrm{excellent})}$$

$\mathrm{hits}(q)$ means the number of hits returned by a search engine when query q is issued. **NEAR** means NEAR operator, which enables to retrieve only such documents that contain two queries within ten words.

Hatzivassiloglou and McKeown constructed lexical network and determine polarity of adjectives (Hatzivassiloglous and McKeown, 1997). Although this is similar to thesaurus based approach, they built the network from intra-sentential co-occurrence. Takamura et al. built lexical network from not only such co-occurrence but other resources including thesaurus (Takamura et al., 2005). They used spin model to predict polarity of words.

Popescu and Etzioni applied relaxation labeling to polarity identification (Popescu and Etzioni, 2005). This method iteratively assigns polarity to words by using various features including intra-sentential co-occurrence and synonyms of a thesaurus.

Kanayama and Nasukawa used both intra- and inter-sentential co-occurrence to learn polarity of words and phrases (Kanayama and Nasukawa, 2006). Their method covers wider range of co-occurrence than other work such as (Hatzivassiloglous and McKeown, 1997). An interesting point of this work is that they discussed building domain oriented lexicon. This is contrastive to other work including ours that addresses to build domain independent lexicon.

In summary, the strength of our approach is to exploit extremely precise structural clues, and to use massive collection of HTML documents to compensate for the low recall. Although Turney's method also uses massive collection of HTML documents, his method does not make much of precision compared with our method. As we will see in Section 5, our experimental result revealed that our method overwhelms Turney's method.

[5]Semantic Orientation in (Turney, 2002).

4.3 Other related work

In some review sites, pros and cons are stated using such layout that we introduced in Section 2. Some work examined the importance of such layout (Liu et al., 2005; Kim and Hovy, 2006). However, they regarded layout structures as clues specific to a certain review site. They did not propose to use layout structure to extract polar sentences from arbitrary HTML documents.

Some studies addressed supervised approach to learning polarity of phrases (Wilson et al., 2005; Takamura et al., 2006). These are different from ours in a sense that they require manually tagged data.

Kobayashi et al. proposed a framework to reduce the cost of manually building lexicon (Kobayashi et al., 2004). In the experiment, they compared the framework with fully manual method and investigated the effectiveness.

5 Experiment

A test set consisting of 405 adjective phrases were created. From the test set, we extract polar phrases by looking up our lexicon. The result was evaluated through precision and recall[6].

5.1 Setting

The test set was created in the following manner. 500 adjective phrases were randomly extracted from the Web text. Note that there is no overlap between our polar sentence corpus and this text. After removing parsing error and duplicates, 405 unique adjective phrases were obtained. Each phase was manually annotated with polarity tag (positive, negative and neutral), and we obtained 158 positive phrases, 150 negative phrases and 97 neutral phrases. In order to check the reliability of annotation, another

[6]The lexicon is available from http://www.tkl.iis.u-tokyo.ac.jp/~kaji/polardic/.

Table 3: The experimental result (chi-square).

θ		0	10	20	30	40	50	60
Precision/Recall	Positive	76.4/92.4	84.0/86.7	84.1/83.5	86.2/79.1	88.7/74.7	86.7/65.8	86.7/65.8
	Negative	68.5/84.0	65.5/63.3	64.3/60.0	62.7/57.3	81.1/51.3	80.0/48.0	80.0/48.0
# of polar words and phrases		9,670	2,056	1,047	698	533	423	335

Table 4: The experimental result (PMI).

θ		0	0.5	1.0	1.5	2.0	2.5	3.0
Precision/Recall	Positive	76.4/92.4	79.6/91.1	86.1/89.9	87.2/86.1	90.9/82.3	92.4/76.6	92.9/65.8
	Negative	68.5/84.0	75.8/81.3	82.3/77.3	84.8/74.7	85.8/72.7	86.8/70.0	87.9/62.7
# of polar words and phrases		9,670	9,320	9,039	8,804	8,570	8,398	8,166

Table 5: The effect of data size (PMI, θ=1.0).

size		1/20	1/15	1/10	1/5	1
Precision/Recall	Positive	87.0/63.9	84.6/65.8	85.1/75.9	85.4/84.8	86.1/89.9
	Negative	76.9/55.8	86.2/50.0	82.1/58.0	80.3/62.7	82.3/77.3

human judge annotated the same data. The Kappa value between the two judges was 0.73, and we think the annotation is reliable.

From the test set, we extracted polar phrases by looking up our lexicon. As for adjectives in the lexicon, partial match is allowed. For example, if the lexicon contains an adjective 'excellent', it matches every adjective phrase that includes 'excellent' such as 'view-excellent' etc.

As a baseline, we built lexicon similarly by using polarity value of (Turney, 2002). As seed words, we used '*saikou* (best)' and '*saitei* (worst)'. Some seeds were tested and these words achieved the best result. As a search engine, we tested Google and our local engine, which indexes 150 millions Japanese documents. Its size is compatible to (Turney and Littman, 2002). Since Google does not support NEAR, we used AND. Our local engine supports NEAR.

5.2 Results and discussion

We evaluated the result of polar phrase extraction. By changing the threshold θ, we investigated recall-precision curve (Figure 6 and 7). The detail is represented in Table 3 and 4. The second/third row represents precision and recall of positive/negative phrases. The fourth row is the size of the lexicon.

The Figures show that both of the proposed methods outperform the baselines. The best F-measure was achieved by PMI (θ=1.0). Although Turney's method may be improved with minor configurations (e.g. using other seeds etc.), we think this results indicate the feasibility of the proposed method. Although the size of lexicon is not surprisingly large, it would be possible to make the lexicon larger by using more HTML documents. In addition, notice that we focus on only adjectives and adjective phrases.

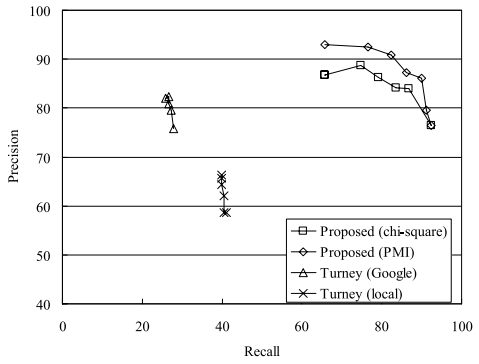


Figure 6: Recall-precision curve (positive phrases)

Comparing the two proposed methods, PMI is always better than chi-square. Especially, chi-square suffers from low recall, because the size of lexicon is extremely small. For example, when the threshold is 60, the precision is 80% and the recall is 48% for negative phrases. On the other hand, PMI would achieve the same precision when recall is around 80% (θ is between 0.5 and 1.0).

Turney's method did not work well although they reported 80% accuracy in (Turney and Littman, 2002). This is probably because our experimental setting is different. Turney examined binary classification of positive and negative words, and we discussed extracting positive and negative phrases from the set of positive, negative and neutral phrases.

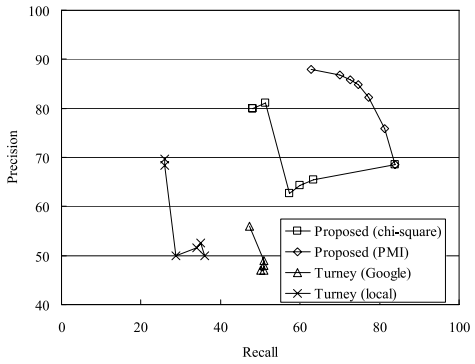


Figure 7: Recall-precision curve (negative phrases)

Error analysis revealed that most of the errors are related to neutral phrases. For example, PMI (θ=1.0) extracted 48 incorrect polar phrases, and 37 of them were neutral phrases. We think one reason is that we did not use neutral corpus. It is one future work to exploit neutral corpus. The importance of neutral category is also discussed in other literatures (Esuli and Sebastiani, 2006).

To further assess our method, we did two additional experiments. In the first experiment, to investigate the effect of data size, the same experiment was conducted using 1/n (n=1,5,10,15,20) of the entire polar sentence corpus (Table 5). PMI (θ=1.0) was also used. As the size of corpus increases, the performance becomes higher. Especially, the recall is improved dramatically. Therefore, the recall would be further improved using more corpus.

In the other experiment, the lexicon was evaluated directly so that we can examine polar words and phrases that are not in the test set. We think it is difficult to fully assess low frequency words in the previous setting. Two human judges assessed 200 unique polar words and phrases in the lexicon (PMI, θ=1.0). The average precision was 71.3% (Kappa value was 0.66). The precision is lower than the result in Table 4. This result indicates that it is difficult to handle low frequency words.

The Table 6 illustrates examples of polar phrases and their polarity values. We can see that both phrases and colloquial words such as 'uncool' are appropriately learned. They are difficult to handle for thesaurus based approach, because such words are not usually in thesaurus.

It is important to discuss how general our framework is. Although the lexico-syntactic patterns shown in Section 2 are specific to Japanese, we think that the idea of exploiting language structure is applicable to other languages including English. Roughly speaking, the pattern we exploited can be translated into 'the advantage/weakness of something is to ...' in English. It is worth pointing out that lexico-syntactic patterns have been widely used in English lexical acquisition (Hearst, 1992). Obviously, other parts of the proposed method does not depend on Japanese.

Table 6: Examples

polar phrase	$PV_{\chi^2}(c)$	$PV_{PMI}(c)$
kenkyoda (modest)	38.3	12.1
exiting (exiting)	13.5	10.4
more-sukunai (leak-small)	9.2	9.8
dasai (uncool)	-2.9	-3.3
yakkaida (annoying)	-11.9	-3.9
shomo-hayai (consumption-quick)	-17.7	-4.4

6 Conclusion

In this paper, we explore to use structural clues that can extract polar sentences from Japanese HTML documents, and build lexicon from the extracted polar sentences. The key idea is to develop the structural clues so that it achieves extremely high precision at the cost of recall. In order to compensate for the low recall, we used massive collection of HTML documents. Thus, we could prepare enough polar sentence corpus. Experimental result demonstrated the feasibility of our approach.

Acknowledgement This work was supported by the *Comprehensive Development of e-Society Foundation Software* program of the Ministry of Education, Culture, Sports, Science and Technology, Japan. We would like to thank Assistant Researcher Takayuki Tamura for his development of the Web crawler.

References

Kenneth Ward Church and Patric Hanks. 1989. Word association norms, mutual information, and lexicography. In *Proceedings of ACL*, pages 76–83.

Andrea Esuli and Fabrizio Sebastiani. 2005. Determining the semantic orientation of terms throush gloss classification. In *Proceedings of CIKM*, pages 617–624.

Andrea Esuli and Fabrizio Sebastiani. 2006. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of EACL*, pages 193–200.

Vasileios Hatzivassiloglous and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of ACL*, pages 174–181.

Marti Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of COLING*, pages 539–545.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of KDD*, pages 168–177.

Nobuhiro Kaji and Masaru Kitsuregawa. 2006. Automatic construction of polarity-tagged corpus from html documents. In *Proceedings of COLING/ACL, poster sessions*, pages 452–459.

Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten de Rijke. 2004. Using wordnet to measure semantic orientations of adjectives. In *Proceedings of LREC*.

Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of ENMLP*, pages 355–363.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of COLING*, pages 1367–1373.

Soo-Min Kim and Eduard Hovy. 2006. Automatic identification of pro and con reasons in online reviews. In *Proceedings of COLING/ACL Poster Sessions*, pages 483–490.

Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima. 2004. Collecting evaluative expressions for opinion extraction. In *Proceedings of IJCNLP*, pages 584–589.

Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of WWW*.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of HLT/EMNLP*.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientation of words using spin model. In *Proceedings of ACL*, pages 133–140.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2006. Latent variable mdels for semantic orientations of phrases. In *Proceedings of EACL*, pages 201–208.

Peter D. Turney and Michael L. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Technical report, National Research Council Canada.

Peter D. Turney. 2002. Thumbs up or thumbs down ? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL*, pages 417–424.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT/EMNLP*.