Fully Automatic Lexicon Expansion
for Domain-oriented Sentiment Analysis

Hiroshi Kanayama   Tetsuya Nasukawa

Tokyo Research Laboratory, IBM Japan, Ltd.
1623-14 Shimotsuruma, Yamato-shi, Kanagawa-ken, 242-8502 Japan
{hkana,nasukawa}@jp.ibm.com

Abstract

This paper proposes an unsupervised lexicon building method for the detection of *polar clauses*, which convey positive or negative aspects in a specific domain. The lexical entries to be acquired are called *polar atoms*, the minimum human-understandable syntactic structures that specify the polarity of clauses. As a clue to obtain candidate polar atoms, we use *context coherency*, the tendency for same polarities to appear successively in contexts. Using the overall density and precision of coherency in the corpus, the statistical estimation picks up appropriate polar atoms among candidates, without any manual tuning of the threshold values. The experimental results show that the precision of polarity assignment with the automatically acquired lexicon was 94% on average, and our method is robust for corpora in diverse domains and for the size of the initial lexicon.

1 Introduction

Sentiment Analysis (SA) (Nasukawa and Yi, 2003; Yi et al., 2003) is a task to recognize writers' feelings as expressed in positive or negative comments, by analyzing unreadably large numbers of documents. Extensive syntactic patterns enable us to detect sentiment expressions and to convert them into semantic structures with high precision, as reported by Kanayama et al. (2004). From the example Japanese sentence (1) in the digital camera domain, the SA system extracts a sentiment representation as (2), which consists of a predicate and an argument with positive (+) polarity.

(1) *Kono kamera-ha subarashii-to omou.*
　　　'I think this camera is splendid.'

(2) [+] splendid(camera)

SA in general tends to focus on subjective sentiment expressions, which explicitly describe an author's preference as in the above example (1). Objective (or factual) expressions such as in the following examples (3) and (4) may be out of scope even though they describe desirable aspects in a specific domain. However, when customers or corporate users use SA system for their commercial activities, such domain-specific expressions have a more important role, since they convey strong or weak points of the product more directly, and may influence their choice to purchase a specific product, as an example.

(3) *Kontorasuto-ga kukkiri-suru.*
　　　'The contrast is sharp.'
(4) *Atarashii kishu-ha zuumu-mo tsuite-iru.*
　　　'The new model has a zoom lens, too.'

This paper addresses the Japanese version of *Domain-oriented Sentiment Analysis*, which identifies *polar clauses* conveying goodness and badness in a specific domain, including rather objective expressions. Building domain-dependent lexicons for many domains is much harder work than preparing domain-independent lexicons and syntactic patterns, because the possible lexical entries are too numerous, and they may differ in each domain. To solve this problem, we have devised an unsupervised method to acquire domain-dependent lexical knowledge where a user has only to collect unannotated domain corpora.

The knowledge to be acquired is a domain-dependent set of *polar atoms*. A polar atom is a minimum syntactic structure specifying polarity in a predicative expression. For example, to detect polar clauses in the sentences (3)

and (4)[1], the following polar atoms (5) and (6) should appear in the lexicon:

(5) [+] *kukkiri-suru*
'to be sharp'

(6) [+] *tsuku ← zuumu-ga*
'to have ← zoom lens-NOM'

The polar atom (5) specified the positive polarity of the verb *kukkiri-suru*. This atom can be generally used for this verb regardless of its arguments. In the polar atom (6), on the other hand, the nominative case of the verb *tsuku* ('have') is limited to a specific noun *zuumu* ('zoom lens'), since the verb *tsuku* does not hold the polarity in itself. The automatic decision for the scopes of the atoms is one of the major issues.

For lexical learning from unannotated corpora, our method uses *context coherency* in terms of polarity, an assumption that polar clauses with the same polarity appear successively unless the context is changed with adversative expressions. Exploiting this tendency, we can collect candidate polar atoms with their tentative polarities as those adjacent to the polar clauses which have been identified by their domain-independent polar atoms in the initial lexicon. We use both intra-sentential and inter-sentential contexts to obtain more candidate polar atoms.

Our assumption is intuitively reasonable, but there are many non-polar (neutral) clauses adjacent to polar clauses. Errors in sentence delimitation or syntactic parsing also result in false candidate atoms. Thus, to adopt a candidate polar atom for the new lexicon, some threshold values for the frequencies or ratios are required, but they depend on the type of the corpus, the size of the initial lexicon, etc.

Our algorithm is *fully automatic* in the sense that the criteria for the adoption of polar atoms are set automatically by statistical estimation based on the distributions of coherency: *coherent precision* and *coherent density*. No manual tuning process is required, so the algorithm only needs unannotated domain corpora and the initial lexicon. Thus our learning method can be used not only by the developers of the system, but also by end-users. This feature is very helpful for users to

---

[1]The English translations are included only for convenience.

analyze documents in new domains.

In the next section, we review related work, and Section 3 describes our runtime SA system. In Section 4, our assumption for unsupervised learning, *context coherency* and its key metrics, *coherent precision* and *coherent density* are discussed. Section 5 describes our unsupervised learning method. Experimental results are shown in Section 6, and we conclude in Section 7.

2 Related Work

Sentiment analysis has been extensively studied in recent years. The target of SA in this paper is wider than in previous work. For example, Yu and Hatzivassiloglou (2003) separated facts from opinions and assigned polarities only to opinions. In contrast, our system detects factual polar clauses as well as sentiments.

Unsupervised learning for sentiment analysis is also being studied. For example, Hatzivassiloglou and McKeown (1997) labeled adjectives as positive or negative, relying on *semantic orientation*. Turney (2002) used collocation with "excellent" or "poor" to obtain positive and negative clues for document classification. In this paper, we use contextual information which is wider than for the contexts they used, and address the problem of acquiring lexical entries from the noisy clues.

Inter-sentential contexts as in our approach were used as a clue also for subjectivity analysis (Riloff and Wiebe, 2003; Pang and Lee, 2004), which is two-fold classification into subjective and objective sentences. Compared to it, this paper solves a more difficult problem: three-fold classification into positive, negative and non-polar expressions using imperfect coherency in terms of sentiment polarity.

Learning methods for phrase-level sentiment analysis closely share an objective of our approach. Popescu and Etzioni (2005) achieved high-precision opinion phrases extraction by using *relaxation labeling*. Their method iteratively assigns a polarity to a phrase, relying on semantic orientation of co-occurring words in specific relations in a sentence, but the scope of semantic orientation is limited to within a sentence. Wilson et al. (2005) proposed supervised learning, dividing the resources into
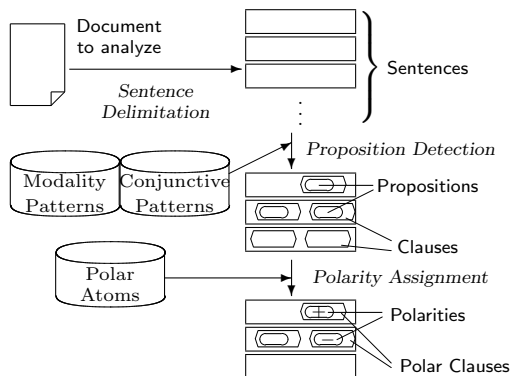
Figure 1: The flow of the clause-level SA.

prior polarity and context polarity, which are similar to polar atoms and syntactic patterns in this paper, respectively. Wilson et al. prepared prior polarities from existing resources, and learned the context polarities by using prior polarities and annotated corpora. Therefore the prerequisite data and learned data are opposite from those in our approach. We took the approach used in this paper because we want to acquire more domain-dependent knowledge, and context polarity is easier to access in Japanese[2]. Our approach and their work can complement each other.

3 Methodology of Clause-level SA

As Figure 1 illustrates, the flow of our sentiment analysis system involves three steps. The first step is *sentence delimitation*: the input document is divided into sentences. The second step is *proposition detection*: propositions which can form polar clauses are identified in each sentence. The third step is *polarity assignment*: the polarity of each proposition is examined by considering the polar atoms. This section describes the last two processes, which are based on a deep sentiment analysis method analogous to machine translation (Kanayama et al., 2004) (hereafter "the MT method").

3.1 Proposition Detection

Our basic tactic for clause-level SA is the high-precision detection of polar clauses based on deep syntactic analysis. 'Clause-level' means that only predicative verbs and adjectives such

as in (7) are detected, and adnominal (attributive) usages of verbs and adjectives as in (8) are ignored, because *utsukushii* ('beautiful') in (8) does not convey a positive polarity.

(7) *E-ga utsukushii.*
   'The picture is beautiful.'

(8) *Utsukushii hito-ni aitai.*
   'I want to meet a beautiful person.'

Here we use the notion of a *proposition* as a clause without modality, led by a predicative verb or a predicative adjective. The propositions detected from a sentence are subject to the assignment of polarities.

Basically, we detect a proposition only at the head of a syntactic tree[3]. However, this limitation reduces the recall of sentiment analysis to a very low level. In the example (7) above, *utsukushii* is the head of the tree, while those initial clauses in (9) to (11) below are not. In order to achieve higher recall while maintaining high precision, we apply two types of syntactic patterns, *modality patterns* and *conjunctive patterns*[4], to the tree structures from the full-parsing.

(9) *Sore-ha utsukushii-to omou.*
   'I think it is beautiful.'

(10) *Sore-ha utsukushiku-nai.*
   'It is not beautiful.'

(11) *Sore-ga utsukushii-to yoi.*
   'I hope it is beautiful.'

Modality patterns match some auxiliary verbs or corresponding sentence-final expressions, to allow for specific kinds of modality and negation. One of the typical patterns is [v] *to omou* ('I think [v]')[5], which allows *utsukushii* in (9) to be a proposition. Also negation is handled with a modality pattern, such as [v] *nai* ('not [v]'). In this case a neg feature is attached to the proposition to identify *utsukushii* in (10) as a negated proposition. On the other hand, no proposition is identified in (11) due to the deliberate absence of a pattern [v] *to yoi* ('I hope [v]'). We used a total of 103 domain-independent modality patterns, most of which are derived from the

---

[2]For example, indirect negation such as caused by a subject "nobody" or a modifier "seldom" is rare in Japanese.

[3]This is same as the rightmost part of the sentence since all Japanese modification is directed left to right.

[4]These two types of patterns correspond to *auxiliary patterns* in the MT method, and can be applied independent of domains.

[5][v] denotes a verb or an adjective.

coordinative (roughly 'and')
-te, -shi, -ueni, -dakedenaku, -nominarazu
causal (roughly 'because')
-tame, -kara, -node
adversative (roughly 'but')
-ga, -kedo, -keredo, - monono, -nodaga

Table 1: Japanese conjunctions used for conjunctive patterns.

MT method, and some patterns are manually added for this work to achieve higher recall.

Another type of pattern is conjunctive patterns, which allow multiple propositions in a sentence. We used a total of 22 conjunctive patterns also derived from the MT method, as exemplified in Table 1. In such cases of coordinative clauses and causal clauses, both clauses can be polar clauses. On the other hand, no proposition is identified in a conditional clause due to the absence of corresponding conjunctive patterns.

3.2 Polarity Assignment Using Polar Atoms

To assign a polarity to each proposition, *polar atoms* in the lexicon are compared to the proposition. A polar atom consists of polarity, verb or adjective, and optionally, its arguments. Example (12) is a *simple polar atom*, where no argument is specified. This atom matches any proposition whose head is *utsukushii*. Example (13) is a *complex polar atom*, which assigns a negative polarity to any proposition whose head is the verb *kaku* and where the accusative case is *miryoku*.

(12) [+] *utsukushii*
         'to be beautiful'

(13) [−] *kaku ← miryoku-wo*
         'to lack ← attraction-ACC'

A polarity is assigned if there exists a polar atom for which verb/adjective and the arguments coincide with the proposition, and otherwise no polarity is assigned. The opposite polarity of the polar atom is assigned to a proposition which has the `neg` feature.

We used a total of 3,275 polar atoms, most of which are derived from an English sentiment lexicon (Yi et al., 2003).

According to the evaluation of the MT method (Kanayama et al., 2004), high-precision sentiment analysis had been achieved using the polar atoms and patterns, where the
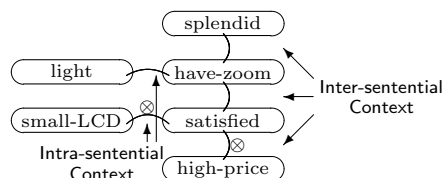

Figure 2: The concept of the intra- and inter-sentential contexts, where the polarities are perfectly coherent. The symbol '⊗' denotes the existence of an adversative conjunction.

system never took positive sentiment for negative and vice versa, and judged positive or negative to neutral expressions in only about 10% cases. However, the recall is too low, and most of the lexicon is for domain-independent expressions, and thus we need more lexical entries to grasp the positive and negative aspects in a specific domain.

4 Context Coherency

This section introduces the intra- and inter-sentential contexts in which we assume *context coherency* for polarity, and describes some preliminary analysis of the assumption.

4.1 Intra-sentential and Inter-sentential Context

The identification of propositions described in Section 3.1 clarifies our viewpoint of the contexts. Here we consider two types of contexts: *intra-sentential context* and *inter-sentential context*. Figure 2 illustrates the context coherency in a sample discourse (14), where the polarities are perfectly coherent.

(14) *Kono kamera-ha subarashii-to omou.*
    'I think this camera is splendid.'
    Karui-shi, zuumu-mo tsuite-iru.
    'It's light and has a zoom lens.'
    Ekishou-ga chiisai-kedo, manzoku-da.
    'Though the LCD is small, I'm satisfied.'
    Tada, nedan-ga chotto takai.
    'But, the price is a little high.'

The intra-sentential context is the link between propositions in a sentence, which are detected as coordinative or causal clauses. If there is an adversative conjunction such as *-kedo* ('but') in the third sentence in (14), a flag is attached to the relation, as denoted with '⊗' in Figure 2. Though there are differences in syntactic phenomena, this is sim-

shikashi ('however'), demo ('but'), sorenanoni ('even though'), tadashi ('on condition that'), dakedo ('but'), gyakuni ('on the contrary'), tohaie ('although'), keredomo ('however'), ippou ('on the other hand')

Table 2: Inter-sentential adversative expressions.

Domain	Post.	Sent.	Len.
digital cameras	263,934	1,757,917	28.3
movies	163,993	637,054	31.5
mobile phones	155,130	609,072	25.3
cars	159,135	959,831	30.9

Table 3: The corpora from four domains used in this paper. The "Post." and "Sent." columns denote the numbers of postings and sentences, respectively. "Len." is the average length of sentences (in Japanese characters).

ilar to the semantic orientation proposed by Hatzivassiloglou and McKeown (1997).

The inter-sentential context is the link between propositions in the main clauses of pairs of adjacent sentences in a discourse. The polarities are assumed to be the same in the inter-sentential context, unless there is an adversative expression as those listed in Table 2. If no proposition is detected as in a nominal sentence, the context is split. That is, there is no link between the proposition of the previous sentence and that of the next sentence.

4.2 Preliminary Study on Context Coherency

We claim these two types of context can be used for unsupervised learning as clues to assign a tentative polarity to unknown expressions. To validate our assumption, we conducted preliminary observations using various corpora.

4.2.1 Corpora

Throughout this paper we used Japanese corpora from discussion boards in four different domains, whose features are shown in Table 3. All of the corpora have clues to the boundaries of postings, so they were suitable to identify the discourses.

4.2.2 Coherent Precision

How strong is the coherency in the context proposed in Section 4.1? Using the polar clauses detected by the SA system with the

initial lexicon, we observed the *coherent precision* of domain d with lexicon L, defined as:

$$cp(d, L) = \frac{\#(\text{Coherent})}{\#(\text{Coherent}) + \#(\text{Conflict})} \quad (15)$$

where #(Coherent) and #(Conflict) are occurrence counts of the same and opposite polarities observed between two polar clauses as observed in the discourse. As the two polar clauses, we consider the following types:

Window. A polar clause and the nearest polar clause which is found in the preceding n sentences in the discourse.

Context. Two polar clauses in the intra-sentential and/or inter-sentential context described in Section 4.1. This is the viewpoint of context in our method.

Table 4 shows the frequencies of coherent pairs, conflicting pairs, and the coherent precision for half of the digital camera domain corpus. "Baseline" is the percentage of positive clauses among the polar clauses[6].

For the "Window" method, we tested for $n=0, 1, 2$, and ∞. "0" means two propositions within a sentence. Apparently, the larger the window size, the smaller the cp value. When the window size is "∞", implying anywhere within a discourse, the ratio is larger than the baseline by only 2.7%, and thus these types of coherency are not reliable even though the number of clues is relatively large.

"Context" shows the coherency of the two types of context that we considered. The cp values are much higher than those in the "Window" methods, because the relationships between adjacent pairs of clauses are handled more appropriately by considering syntactic trees, adversative conjunctions, etc. The cp values for inter-sentential and intra-sentential contexts are almost the same, and thus both contexts can be used to obtain 2.5 times more clues for the intra-sentential context. In the rest of this paper we will use both contexts.

We also observed the coherent precision for each domain corpus. The results in the center column of Table 5 indicate the number is slightly different among corpora, but all of them are far from perfect coherency.

---

[6]If there is a polar clause whose polarity is unknown, the polarity is correctly predicted with at least 57.0% precision by assuming "positive".

Model		Coherent	Conflict	$cp(d, L)$
Baseline				57.0%
Window	$n = 0$	3,428	1,916	64.1%
	$n = 1$	11,448	6,865	62.5%
	$n = 2$	16,231	10,126	61.6%
	$n = \infty$	26,365	17,831	59.7%
Context	intra.	2,583	996	72.2%
	inter.	3,987	1,533	72.2%
	both	6,570	2,529	72.2%

Table 4: Coherent precision with various viewpoints of contexts.

Domain	$cp(d, L)$	$cd(d, L)$
digital cameras	72.2%	7.23%
movies	76.7%	18.71%
mobile phones	72.9%	7.31%
cars	73.4%	7.36%

Table 5: Coherent precision and coherent density for each domain.

4.2.3 Coherent Density

Besides the conflicting cases, there are many more cases where a polar clause does not appear in the polar context. We also observed the *coherent density* of the domain d with the lexicon L defined as:

$$cd(d, L) = \frac{\#(\text{Coherent})}{\#(\text{Polar})} \qquad (16)$$

This indicates the ratio of polar clauses that appear in the coherent context, among all of the polar clauses detected by the system.

The right column of Table 5 shows the coherent density in each domain. The movie domain has notably higher coherent density than the others. This indicates the sentiment expressions are more frequently used in the movie domain.

The next section describes the method of our unsupervised learning using this imperfect context coherency.

5 Unsupervised Learning for Acquisition of Polar Atoms

Figure 3 shows the flow of our unsupervised learning method. First, the runtime SA system identifies the polar clauses, and the candidate polar atoms are collected. Then, each candidate atom is validated using the two metrics in the previous section, cp and cd, which are calculated from all of the polar clauses found in the domain corpus.
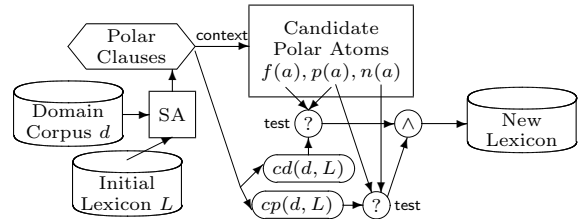

Figure 3: The flow of the learning process.

ID	Candidate Polar Atom	$f(a)$	$p(a)$	$n(a)$
1*	*chiisai* 'to be small'	3,014	226	227
2	*shikkari-suru* 'to be firm'	246	54	10
3	*chiisai ← bodii-ga* 'to be small ← body-NOM'	11	4	0
4*	*todoku ← mokuyou-ni* 'to be delivered←on Thursday'	2	0	2

Table 6: Examples of candidate polar atoms and their frequencies. '*' denotes that it should not be added to the lexicon. $f(a)$, $p(a)$, and $n(a)$ denote the frequency of the atom and in positive and negative contexts, respectively.

5.1 Counts of Candidate Polar Atoms

From each proposition which does not have a polarity, *candidate polar atoms* in the form of simple atoms (just a verb or adjective) or complex atoms (a verb or adjective and its rightmost argument consisting of a pair of a noun and a postpositional) are extracted. For each candidate polar atom a, the total appearances $f(a)$, and the occurrences in positive contexts $p(a)$ and negative contexts $n(a)$ are counted, based on the context of the adjacent clauses (using the method described in Section 4.1). If the proposition has the neg feature, the polarity is inverted. Table 6 shows examples of candidate polar atoms with their frequencies.

5.2 Determination for Adding to Lexicon

Among the located candidate polar atoms, how can we distinguish *true polar atoms*, which should be added to the lexicon, from *fake polar atoms*, which should be discarded?

As shown in Section 4, both the coherent precision (72-77%) and the coherent density (7-19%) are so small that we cannot rely on each single appearance of the atom in the polar context. One possible approach is to set the threshold values for frequency in a polar context, $\max(p(a), n(a))$ and for the ratio of appearances in polar contexts among the to-

tal appearances, $\frac{\max(p(a),n(a))}{f(a)}$. However, the optimum threshold values should depend on the corpus and the initial lexicon.

In order to set general criteria, here we assume that a true positive polar atom a should have higher $\frac{p(a)}{f(a)}$ than its average *i.e.* coherent density, $cd(d, L_{+a})$, and also have higher $\frac{p(a)}{p(a)+n(a)}$ than its average *i.e.* coherent precision, $cp(d, L_{+a})$ and these criteria should be met with 90% confidence, where L_{+a} is the initial lexicon with a added. Assuming the binomial distribution, a candidate polar atom is adopted as a positive polar atom[7] if both (17) and (18) are satisfied[8].

$$q > cd(d, L),$$
where
$$\sum_{k=0}^{p(a)} {}_{f(a)}C_k q^k (1-q)^{f(a)-k} = 0.9 \qquad (17)$$

$$r > cp(d, L) \text{ or } n(a) = 0,$$
where
$$\sum_{k=0}^{p(a)} {}_{p(a)+n(a)}C_k r^k (1-r)^{p(a)+n(a)-k} = 0.9 \qquad (18)$$

We can assume $cd(d, L_{+a}) \simeq cd(d, L)$, and $cp(d, L_{+a}) \simeq cp(d, L)$ when L is large. We compute the confidence interval using approximation with the F-distribution (Blyth, 1986).

These criteria solve the problems in minimum frequency and scope of the polar atoms simultaneously. In the example of Table 6, the simple atom *chiisai* (ID=1) is discarded because it does not meet (18), while the complex atom *chiisai ← bodii-ga* (ID=3) is adopted as a positive atom. *shikkari-suru* (ID=2) is adopted as a positive simple atom, even though 10 cases out of 64 were observed in the negative context. On the other hand, *todoku ← mokuyou-ni* (ID=4) is discarded because it does not meet (17), even though $\frac{n(a)}{f(a)} = 1.0$, *i.e.* always observed in negative contexts.

6 Evaluation

6.1 Evaluation by Polar Atoms

First we propose a method of evaluation of the lexical learning.

		Annotator B		
		Positive	Neutral	Negative
Anno-	Positive	65	11	3
tator	Neutral	3	72	0
A	Negative	1	4	41

Table 7: Agreement of two annotators' judgments of 200 polar atoms. κ=0.83.

It is costly to make consistent and large 'gold standards' in multiple domains, especially in identification tasks such as clause-level SA (*cf.* classification tasks). Therefore we evaluated the learning results by asking human annotators to classify the acquired polar atoms as positive, negative, and neutral, instead of the instances of polar clauses detected with the new lexicon. This can be done because the polar atoms themselves are informative enough to imply to humans whether the expressions hold positive or negative meanings in the domain.

To justify the reliability of this evaluation method, two annotators[9] evaluated 200 randomly selected candidate polar atoms in the digital camera domain. The agreement results are shown in Table 7. The manual classification was agreed upon in 89% of the cases and the Kappa value was 0.83, which is high enough to be considered consistent.

Using manual judgment of the polar atoms, we evaluated the performance with the following three metrics.

Type Precision. The coincidence rate of the polarity between the acquired polar atom and the human evaluators' judgments. It is always false if the evaluators judged it as 'neutral.'

Token Precision. The coincidence rate of the polarity, weighted by its frequency in the corpus. This metric emulates the precision of the detection of polar clauses with newly acquired poler atoms, in the runtime SA system.

Relative Recall. The estimated ratio of the number of detected polar clauses with the expanded lexicon to the number of detected polar clauses with the initial lex-

---

[7] The criteria for the negative atoms are analogous.
[8] $_nC_r$ notation is used here for combination (n choose k).

[9] For each domain, we asked different annotators who are familiar with the domain. They are not the authors of this paper.

Domain	#	Type Prec.	Token Prec.	Relative Recall
digital cameras	708	65%	96.5%	1.28
movies	462	75%	94.4%	1.19
mobile phones	228	54%	92.1%	1.13
cars	487	68%	91.5%	1.18

Table 8: Evaluation results with our method. The column '#' denotes the number of polar atoms acquired in each domain.

icon. Relative recall will be 1 when no new polar atom is acquired. Since the precision was high enough, this metric can be used for approximation of the recall, which is hard to evaluate in extraction tasks such as clause-/phrase-level SA.

6.2 Robustness for Different Conditions

6.2.1 Diversity of Corpora

For each of the four domain corpora, the annotators evaluated 100 randomly selected polar atoms which were newly acquired by our method, to measure the precisions. Relative recall is estimated by comparing the numbers of detected polar clauses from randomly selected 2,000 sentences, with and without the acquired polar atoms. Table 8 shows the results. The token precision is higher than 90% in all of the corpora, including the movie domain, which is considered to be difficult for SA (Turney, 2002). This is extremely high precision for this task, because the correctness of both the extraction and polarity assignment was evaluated simultaneously. The relative recall 1.28 in the digital camera domain means the recall is increased from 43%[10] to 55%. The difference was smaller in other domains, but the domain-dependent polar clauses are much informative than general ones, thus the high-precision detection significantly enhances the system.

To see the effects of our method, we conducted a control experiment which used preset criteria. To adopt the candidate atom a, the frequency of polarity, $\max(p(a), n(a))$ was required to be 3 or more, and the ratio of polarity, $\frac{\max(p(a),n(a))}{f(a)}$ was required to be higher than the threshold θ. Varying θ from 0.05 to

---

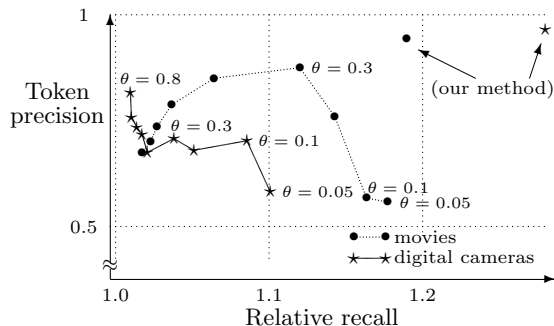[10] The human evaluation result for digital camera domain (Kanayama et al., 2004).

Figure 4: Relative recall vs. token precision with various preset threshold values θ for the digital camera and movie domains. The rightmost star and circle denote the performance of our method.

0.8, we evaluated the token precision and the relative recall in the domains of digital cameras and movies. Figure 4 shows the results.

The results showed both relative recall and token precision were lower than in our method for every θ, in both corpora. The optimum θ was 0.3 in the movie domain and 0.1 in the digital camera domain. Therefore, in this preset approach, a tuning process is necessary for each domain. Our method does not require this tuning, and thus fully automatic learning was possible.

Unlike the normal precision-recall tradeoff, the token precision in the movie domain got lower when the θ is strict. This is due to the frequent polar atoms which can be acquired at the low ratios of the polarity. Our method does not discard these important polar atoms.

6.2.2 Size of the Initial Lexicon

We also tested the performance while varying the size of the initial lexicon L. We prepared three subsets of the initial lexicon, $L_{0.8}$, $L_{0.5}$, and $L_{0.2}$, removing polar atoms randomly. These lexicons had 0.8, 0.5, 0.2 times the polar atoms, respectively, compared to L. Table 9 shows the precisions and recalls using these lexicons for the learning process. Though the cd values vary, the precision was stable, which means that our method was robust even for different sizes of the lexicon. The smaller the initial lexicon, the higher the relative recall, because the polar atoms which were removed from L were recovered in the learning process. This result suggests the possibility of

lexicon	cd	Token Prec.	Relative Rec.
L	7.2%	96.5%	1.28
$L_{0.8}$	6.1%	97.5%	1.41
$L_{0.5}$	3.9%	94.2%	2.10
$L_{0.2}$	3.6%	84.8%	3.55

Table 9: Evaluation results for various sizes of the initial lexicon (the digital camera domain).

the bootstrapping method from a small initial lexicon.

6.3 Qualitative Evaluation

As seen in the agreement study, the polar atoms used in our study were intrinsically meaningful to humans. This is because the atoms are predicate-argument structures derived from predicative clauses, and thus humans could imagine the meaning of a polar atom by generating the corresponding sentence in its predicative form.

In the evaluation process, some interesting results were observed. For example, a negative atom $nai \leftarrow kerare\text{-}ga$ ('to be free from vignetting') was acquired in the digital camera domain. Even the evaluator who was familiar with digital cameras did not know the term $kerare$ ('vignetting'), but after looking up the dictionary she labeled it as negative. Our learning method could pick up such technical terms and labeled them appropriately.

Also, there were discoveries in the error analysis. An evaluator assigned positive to $aru \leftarrow kamera\text{-}ga$ ('to have camera') in the mobile phone domain, but the acquired polar atom had the negative polarity. This was actually an insight from the recent opinions that many users want phones without camera functions[11].

7 Conclusion

We proposed an unsupervised method to acquire polar atoms for domain-oriented SA, and demonstrated its high performance. The lexicon can be expanded automatically by using unannotated corpora, and tuning of the threshold values is not required. Therefore even end-users can use this approach to improve the sentiment analysis. These features allow them to do on-demand analysis of more narrow domains, such as the domain of digital

cameras of a specific manufacturer, or the domain of mobile phones from the female users' point of view.

References

C. R. Blyth. 1986. Approximate binomial confidence limits. *Journal of the American Statistical Asscoiation*, 81(395):843–855.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th ACL and the 8th EACL*, pages 174–181.

Hiroshi Kanayama, Tetsuya Nasukawa, and Hideo Watanabe. 2004. Deeper sentiment analysis using machine translation technology. In *Proceedings of the 20th COLING*, pages 494–500.

Tetsuya Nasukawa and Jeonghee Yi. 2003. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the Second K-CAP*, pages 70–77.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd ACL*, pages 271–278.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of HLT/EMNLP-05*, pages 339–346.

Ellen Riloff and Janyee Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of EMNLP-03*, pages 105–112.

Peter D. Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th 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-05*, pages 347–354.

Jeonghee Yi, Tetsuya Nasukawa, Razvan Bunescu, and Wayne Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the Third IEEE International Conference on Data Mining*, pages 427–434.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of EMNLP-2003*, pages 129–136.

---

[11]Perhaps because cameras tend to consume battery power and some users don't need them.