Opinion Word Expansion and Target Extraction through Double Propagation

Guang Qiu*
Zhejiang University, China

Bing Liu**
University of Illinois at Chicago

Jiajun Bu†
Zhejiang University, China

Chun Chen‡
Zhejiang University, China

Analysis of opinions, known as opinion mining or sentiment analysis, has attracted a great deal of attention recently due to many practical applications and challenging research problems. In this article, we study two important problems, namely, opinion lexicon expansion and opinion target extraction. Opinion targets (targets, for short) are entities and their attributes on which opinions have been expressed. To perform the tasks, we found that there are several syntactic relations that link opinion words and targets. These relations can be identified using a dependency parser and then utilized to expand the initial opinion lexicon and to extract targets. This proposed method is based on bootstrapping. We call it double propagation as it propagates information between opinion words and targets. A key advantage of the proposed method is that it only needs an initial opinion lexicon to start the bootstrapping process. Thus, the method is semi-supervised due to the use of opinion word seeds. In evaluation, we compare the proposed method with several state-of-the-art methods using a standard product review test collection. The results show that our approach outperforms these existing methods significantly.

1. Introduction

Opinion mining (or sentiment analysis) has attracted a great deal of attention from researchers of natural language processing and data mining in the past few years due

* College of Computer Science, Zhejiang University, 38 Zheda Rd., Hangzhou 310027, Zhejiang, China. E-mail: qiuguang@zju.edu.cn.
** Department of Computer Science, University of Illinois, 851 South Morgan Street Chicago, IL 60607-7053. E-mail: liub@cs.uic.edu.
† College of Computer Science, Zhejiang University, 38 Zheda Rd., Hangzhou 310027, Zhejiang, China. E-mail: bjj@zju.edu.cn.
‡ College of Computer Science, Zhejiang University, Corresponding author, 38 Zheda Rd., Hangzhou 310027, Zhejiang, China. E-mail: chenc@zju.edu.cn.

Submission received: 2 September 2009; revised submission received: 20 January 2010; accepted for publication: 20 July 2010.

© 2011 Association for Computational Linguistics

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

to many challenging research problems and practical applications. Two fundamental problems in opinion mining are opinion lexicon expansion and opinion target extraction (Liu 2006; Pang and Lee 2008). An **opinion lexicon** is a list of opinion words such as *good*, *excellent*, *poor*, and *bad* which are used to indicate positive or negative sentiments. It forms the foundation of many opinion mining tasks, for example, sentence (Yu and Hatzivassiloglou 2003; Kim and Hovy 2004) and document (Pang, Lee, and Vaithyanathan 2002; Turney 2002) sentiment classification, and feature-based opinion summarization (Hu and Liu 2004). Although there are several opinion lexicons publicly available, it is hard, if not impossible, to maintain a universal opinion lexicon to cover all domains as opinion expressions vary significantly from domain to domain. A word can be positive in one domain but has no opinion or even negative opinion in another domain. Therefore, it is necessary to expand a known opinion lexicon for applications in different domains using text corpora from the corresponding domains.

Opinion targets are topics on which opinions are expressed. They are important because without knowing the targets, the opinions expressed in a sentence or document are of limited use. For example, in the opinion sentence *I am not happy with the battery life of this phone*, *battery life* is the target of the opinion. If we do not know that, this opinion is of little value. Although several researchers have studied the opinion lexicon expansion and opinion target extraction (also known as topic, feature, or aspect extraction) problems, their algorithms either need additional and external resources or impose strong constraints and are of limited success. Detailed discussions of existing works will be given in Section 2.

In this article, we propose a novel propagation based method to solve the opinion lexicon expansion and target extraction problems simultaneously. Our approach differs from existing approaches in that it requires no additional resources except an initial seed opinion lexicon, which is readily available. Thus, it can be seen as a semi-supervised method due to the use of the seeds. It is based on the observation that there are natural relations between opinion words and targets due to the fact that opinion words are used to modify targets. Furthermore, we find that opinion words and targets themselves have relations in opinionated expressions too. These relations can be identified via a dependency parser based on the dependency grammar (Tesniere 1959), and then exploited to perform the extraction tasks.

The basic idea of our approach is to extract opinion words (or targets) iteratively using known and extracted (in previous iterations) opinion words and targets through the identification of syntactic relations. The identification of the relations is the key to the extractions. As our approach propagates information back and forth between opinion words and targets, we call it **double propagation**. Opinion word sentiment or polarity assignment (positive, negative, or neutral) and noisy target pruning methods are also designed to refine the initially extracted results. In evaluation, we compare our approach with several state-of-the-art existing approaches in opinion lexicon expansion (or opinion word extraction) and target (or feature/topic) extraction. The results show that our approach outperforms these existing approaches significantly.

2. Related Work

Our work is related to opinion word extraction and target (or topic) extraction in opinion mining.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

2.1 Opinion Word Extraction

Extensive work has been done on sentiment analysis at word, expression (Breck, Choi, and Cardie 2007; Takamura, Inui, and Okumura 2007), sentence (Yu and Hatzivassiloglou 2003; Kim and Hovy 2004) and document (Pang, Lee, and Vaithyanathan 2002; Turney 2002) levels. We only describe work at word level as it is most relevant to our work. In general, the existing work can be categorized as corpora-based (Hatzivassiloglou and McKeown 1997; Wiebe 2000; Wiebe et al. 2004; Turney and Littman 2003; Kanayama and Nasukawa 2006; Kaji and Kitsuregawa 2007) and dictionary-based (Hu and Liu 2004; Kim and Hovy 2004; Kamps et al. 2004; Esuli and Sebastiani 2005; Takamura, Inui, and Okumura 2005) approaches. Our work falls into the corpora-based category.

Hatzivassiloglou and McKeown (1997) proposed the first method for determining adjective polarities or orientations (positive, negative, and neutral). The method predicts orientations of adjectives by detecting pairs of such words conjoined by conjunctions such as *and* and *or* in a large document set. The underlying intuition is that the orientations of conjoined adjectives are subject to some linguistic constraints. For example, in the sentence *This car is beautiful and spacious*, if we know that *beautiful* is positive, we can infer that *spacious* is positive too. The weakness of this method is that as it relies on the conjunction relations it is unable to extract adjectives that are not conjoined.

Wiebe (2000) and Wiebe et al. (2004) proposed an approach to finding subjective adjectives using the results of word clustering according to their distributional similarity. However, they did not tackle the prediction of sentiment polarities of the found subjective adjectives. Turney and Littman (2003) compute the point wise mutual information (PMI) of the target term with each seed positive and negative term as a measure of their semantic association. Their work requires additional access to the Web (or any other corpus similar to the Web to ensure sufficient coverage), which is time consuming.

Another recent corpora-based approach is proposed by Kanayama and Nasukawa (2006). Their work first uses clause level context coherency to find candidates, then uses a statistical estimation method to determine whether the candidates are appropriate opinion words. Their method for finding candidates would have low recall if the occurrences of seed words in the data are infrequent or an unknown opinion word has no known opinion words in its context, however. Besides, the statistical estimation can be unreliable if the corpus is small, which is a common problem for statistical approaches. We will compare our approach with this approach in our experiments.

In dictionary-based approaches, Kamps et al. (2004) take advantage of WordNet to construct a synonymy network by connecting pairs of synonymous words. The semantic orientation of a word is decided by its shortest paths to two seed words *good* and *bad* which are chosen as representatives of positive and negative orientations. Esuli and Sebastiani (2005) use text classification techniques to classify orientations. Their method is based on the glosses (textual definitions) in an on-line "glossary" or dictionary. The work of Takamura, Inui, and Okumura (2005) also exploits the gloss information from dictionaries. The method constructs a lexical network by linking two words if one appears in the gloss of the other. The weights of links reflect if these two connected words are of the same orientation. The works of Hu and Liu (2004) and Kim and Hovy (2004) are simpler as they simply used synonyms and antonyms. However, all dictionary-based methods are unable to find domain dependent sentiment words because most entries in dictionaries are domain-independent. For example, *unpredictable* is often a positive opinion word in movie reviews, as in *unpredictable plot*, but in car reviews *unpredictable* is likely to be negative, as in *unpredictable steering*. Our approach

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

extracts opinion words using domain dependent corpora; thus we are able to find domain-dependent opinion words.

2.2 Opinion Target Extraction

Opinion target (or topic) extraction is a difficult task in opinion mining. Several methods have been proposed, mainly in the context of product review mining (Hu and Liu 2004; Popescu and Etzioni 2005; Kobayashi, Inui, and Matsumoto 2007; Mei et al. 2007; Scaffidi et al. 2007; Wong, Lam, and Wong 2008; Stoyanov and Cardie 2008). In this mining task, opinion targets usually refer to product features, which are defined as product components or attributes, as in Liu (2006).

In the work of Hu and Liu (2004), frequent nouns and noun phrases are treated as product feature candidates. In our work, we also extract only noun targets. Different pruning methods are proposed to remove the noise. To cover infrequent features that are missed, they regard the nearest nouns/noun phrases of the opinion words identified by frequent features as infrequent features. In Popescu and Etzioni (2005), the authors investigated the same problem. Their extraction method, however, requires that the product class is known in advance. The algorithm determines whether a noun/noun phrase is a feature by computing the PMI score between the phrase and class-specific discriminators through a Web search. Querying the Web is a problem, as discussed earlier. We will compare these two representative methods with our approach in the experiments.

In Scaffidi et al. (2007), the authors proposed a language model approach to product feature extraction with the assumption that product features are mentioned more often in a product review than they are mentioned in general English text. However, statistics may not be reliable when the corpus is small, as pointed out earlier.

The recent work by Kobayashi, Inui, and Matsumoto (2007) focused on the aspect-evaluation (aspect and evaluation mean the opinion target and opinion word respectively in our context) and aspect-of extraction problems in blogs. Their aspect-evaluation extraction uses syntactic patterns learned via pattern mining to extract $\langle aspect, evaluation \rangle$ pairs. Our work differs from theirs in that we make use of syntactic relations from dependency trees. Additionally, we consider not only the relations of opinion targets and opinion words, but also many other types of relations, as we will see in Section 3.

In Stoyanov and Cardie (2008), the authors treated target extraction as a topic coreference resolution problem. The key to their approach is to cluster opinions sharing the same target together. They proposed to train a classifier to judge if two opinions are on the same target, which indicates that their approach is supervised. Our work differs from theirs in that our approach is semi-supervised.

Other related work on target extraction mainly uses the idea of topic modeling to capture targets in reviews (Mei et al. 2007). Topic modeling is to model the generation of a document set and mine the implied topics in the documents. However, our experiments with topic modeling show that it is only able to find some general or coarse topics in texts and represent them as clusters of words. Their aim is thus different from our fine-grained opinion target extraction task.

3. Relation Identification

As stated previously, identification of the relations between opinion words/targets and other opinion words/targets is the key to our opinion lexicon expansion and target

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

extraction methods. In this section, we will describe the relation identification in detail. Hereafter, for convenience, we refer to the relations between opinion words and targets as OT-Rel, between opinion words themselves as OO-Rel, and between targets as TT-Rel.

In this work, we employ a dependency grammar to describe the relations syntactically. In the dependency grammar, a syntactic relation between two words *A* and *B* can be described as *A* (or *B*) depends on *B* (or *A*). We define two categories to summarize all possible dependencies between two words in sentences.

Definition 1 (Direct Dependency (DD))
A **direct dependency** indicates that one word depends on the other word without any additional words in their dependency path (i.e., directly) or they both depend on a third word directly.

Some examples are given in Figures 1 (a) and (b). In (a), *A* depends on *B* directly and they both depend on *H* directly in (b).

Definition 2 (Indirect Dependency (IDD))
An **indirect dependency** indicates that one word depends on the other word through some additional words (i.e., indirectly) or they both depend on a third word through additional words.

Some examples are shown in Figures 1 (c) and (d). In (c), *A* depends on *B* through *H1*; in (d), *A* depends on *H* through *H1* and *B* depends on *H* through *H2*. Actually, IDDs denote all possible relations apart from DDs.

Note that DDs and IDDs describe only the topology of all possible dependencies. We then impose some constraints of the Part-of-speech (POS) tags on the opinion words

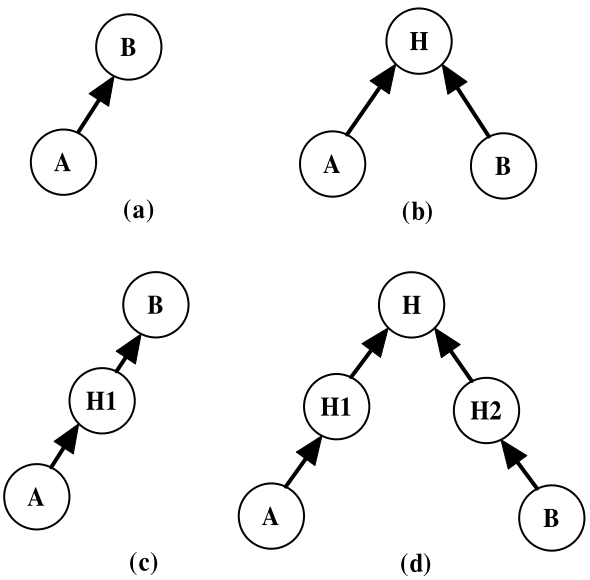


Figure 1
Different dependencies between words *A* and *B*.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

and targets, and also the potential syntactic relations on the dependency path. In this work, we employ the Stanford POS tagging tool[1] to do the POS tagging and Minipar[2] as the sentence parser. We consider opinion words to be adjectives and targets to be nouns/noun phrases, which has been widely adopted in previous work (Hu and Liu 2004; Popescu and Etzioni 2005; Mei et al. 2007). Thus the potential POS tags for opinion words are *JJ* (adjectives), *JJR* (comparative adjectives), and *JJS* (superlative adjectives), whereas those for targets are *NN* (singular nouns) and *NNS* (plural nouns). The dependency relations describing relations between opinion words and targets include *mod*, *pnmod*, *subj*, *s*, *obj*, *obj2* and *desc*; and the relations for opinion words and targets themselves contain only the conjunction relation *conj*. Therefore, we formulate OT-Rel, OO-Rel, or TT-Rel as a quadruple $\langle$ POS(w_i), DT, R, POS(w_j) $\rangle$, in which POS(w_i) is the POS tag of word w_i, DT is the dependency type (i.e., DD or IDD), and R is the syntactic relation. The values of POS(w_i) and R are listed as described here.

4. Opinion Lexicon Expansion and Target Extraction

We perform the opinion lexicon expansion and target extraction tasks iteratively based on propagation using the relations defined herein. To bootstrap the propagation, we only require a seed opinion lexicon. Currently, we focus on one major type of opinionated content, namely, product reviews, in which targets refer to product features. Hereafter, we use target and product feature (or feature for short) interchangeably for convenience.

Our extraction approach adopts the rule-based strategy which is quite natural given the well-defined relations. For example, in an opinion sentence *Canon G3 takes great pictures*, the adjective *great* is parsed as directly depending on the noun *pictures* through *mod*, formulated as a OT-Rel quadruple $\langle JJ, DD, mod, NN \rangle$. If we know *great* is an opinion word and are given a rule like "a noun on which an opinion word directly depends through *mod* is taken as the target," we can easily extract *pictures* as the target. Similarly, if we know *pictures* is a target, we could extract the adjective *great* as an opinion word using a similar rule. Based on such observations, the idea of the whole propagation approach is first to extract opinion words and targets using the seed opinion lexicon and then use the newly extracted opinion words and targets for further target and opinion word extraction. The propagation ends until no more new opinion words or targets can be identified. In this way, even if the seed opinion lexicon is small, targets can still be extracted with high recall (as we will see in the experiments) and at the same time the opinion lexicon is also expanded.

In the following sections, we first describe the extraction rules in detail and then demonstrate the whole propagation algorithm with a walk-through example to show how the propagation works. For the opinion lexicon expansion, one important issue is to assign sentiment polarities to the newly found opinion words. We propose a novel polarity assignment method to perform this task. In target extraction, we also propose several pruning methods to remove different types of noise introduced during the propagation process. We will describe these methods in Sections 4.3 and 4.4.

1 http://nlp.stanford.edu/software/tagger.shtml.
2 http://webdocs.cs.ualberta.ca/lindek/minipar.htm.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

4.1 Propagation Rules Defined Based on Relations

In our propagation, there are four subtasks: (1) extracting targets using opinion words; (2) extracting targets using the extracted targets; (3) extracting opinion words using the extracted targets; (4) extracting opinion words using both the given and the extracted opinion words. OT-Rels are used for tasks (1) and (3), TT-Rels are used for task (2), and OO-Rels are used for task (4). Four types of rules are defined respectively for these four subtasks and the details are shown in Table 1. As parsing is considerably more difficult and error prone with informal expressions used in the Web environment, we only utilize DD dependencies in our current approach. IDDs are more suitable for formal texts. Note that the rules here are all domain-independent.

In the table, o (or t) stands for the output (extracted) opinion word (or target). $\{O\}$ (or $\{T\}$) is the set of known opinion words (or the set of targets) either given or extracted. H means any word. POS(O(or T)) and O(or T)-Dep stand for the POS information and dependency relation of the word O (or T) respectively. $\{JJ\}$ and $\{NN\}$ are sets of POS tags of potential opinion words and targets, respectively. As discussed previously, $\{JJ\}$ contains JJ, JJR, and JJS; $\{NN\}$ contains NN and NNS. $\{MR\}$ consists of dependency

Table 1
Rules for target and opinion word extraction. Column 1 is the rule ID, column 2 is the observed dependency and the constraint that it must satisfy (after s.t.), column 3 is the output, and column 4 is an example. In each example, the underlined word is the known word and the word with double quotes is the extracted word. We also show the corresponding instantiated dependency in the parentheses.

RuleID	Observations	output	Examples
$R1_1$	$O \rightarrow$ *O-Dep* $\rightarrow T$ *s.t.* $O \in \{O\}$, *O-Dep* $\in \{MR\}$, $POS(T) \in \{NN\}$	$t = T$	The phone has a good "screen". (*good* $\rightarrow$ *mod* $\rightarrow$ *screen*)
$R1_2$	$O \rightarrow$ *O-Dep* $\rightarrow H \leftarrow$ *T-Dep* $\leftarrow T$ *s.t.* $O \in \{O\}$, *O/T-Dep* $\in \{MR\}$, $POS(T) \in \{NN\}$	$t = T$	"iPod" is the best mp3 player. (*best* $\rightarrow$ *mod* $\rightarrow$ *player* $\leftarrow$ *subj* $\leftarrow$ *iPod*)
$R2_1$	$O \rightarrow$ *O-Dep* $\rightarrow T$ *s.t.* $T \in \{T\}$, *O-Dep* $\in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as $R1_1$ with screen as the known word and good as the extracted word
$R2_2$	$O \rightarrow$ *O-Dep* $\rightarrow H \leftarrow$ *T-Dep* $\leftarrow T$ *s.t.* $T \in \{T\}$, *O/T-Dep* $\in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as $R1_2$ with iPod as the known word and best as the extract word
$R3_1$	$T_{i(j)} \rightarrow T_{i(j)}$-$Dep \rightarrow T_{j(i)}$ *s.t.* $T_{j(i)} \in \{T\}$, $T_{i(j)}$-$Dep \in \{CONJ\}$, $POS(T_{i(j)}) \in \{NN\}$	$t = T_{i(j)}$	Does the player play dvd with audio and "video"? (*video* $\rightarrow$ *conj* $\rightarrow$ *audio*)
$R3_2$	$T_i \rightarrow T_i$-$Dep \rightarrow H \leftarrow T_j$-$Dep \leftarrow T_j$ *s.t.* $T_i \in \{T\}$, T_i-$Dep == T_j$-Dep, $POS(T_j) \in \{NN\}$	$t = T_j$	Canon "G3" has a great lens. (*lens* $\rightarrow$ *obj* $\rightarrow$ *has* $\leftarrow$ *subj* $\leftarrow$ *G3*)
$R4_1$	$O_{i(j)} \rightarrow O_{i(j)}$-$Dep \rightarrow O_{j(i)}$ *s.t.* $O_{j(i)} \in \{O\}$, $O_{i(j)}$-$Dep \in \{CONJ\}$, $POS(O_{i(j)}) \in \{JJ\}$	$o = O_{i(j)}$	The camera is amazing and "easy" to use. (*easy* $\rightarrow$ *conj* $\rightarrow$ *amazing*)
$R4_2$	$O_i \rightarrow O_i$-$Dep \rightarrow H \leftarrow O_j$-$Dep \leftarrow O_j$ *s.t.* $O_i \in \{O\}$, O_i-$Dep == O_j$-Dep, $POS(O_j) \in \{JJ\}$	$o = O_j$	If you want to buy a sexy, "cool", accessory-available mp3 player, you can choose iPod. (*sexy* $\rightarrow$ *mod* $\rightarrow$ *player* $\leftarrow$ *mod* $\leftarrow$ *cool*)

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

relations describing relations between opinion words and targets (*mod*, *pnmod*, *subj*, *s*, *obj*, *obj2* and *desc*). {*CONJ*} contains *conj* only. The arrows represent dependency. For example, $O \rightarrow O\text{-}Dep \rightarrow T$ means O depends on T through a syntactic relation *O-Dep*. "==" represents the same or equivalent (Here equivalent specifically means *mod* is the same as *pnmod*, and *s* or *subj* is the same as *obj*). For example, $T_i\text{-}Dep == T_j\text{-}Dep$ means T_i-*Dep* being the same as T_j-*Dep* or equivalent (e.g., *subj* and *obj* in $R3_2$).

Specifically, we employ $R1_i$ to extract targets (t) using opinion words (O), $R2_i$ to extract opinion words (o) using targets (T), $R3_i$ to extract targets (t) using extracted targets (T_i) and $R4_i$ to extract opinion words (o) using known opinion words (O_i). Take $R1_1$ as an example. Given the opinion word O, the word with *NN* as its POS tag and satisfying the relation *O-Dep* is extracted as the target. For example, we have the sentence *The phone has a good screen* whose corresponding dependency tree is shown in Figure 2. If we know that *good* is an opinion word, and it depends on *screen* through *mod* which is contained in {*MR*} and *screen* is tagged as *NN*, $R1_1$ can be applied to extract *screen* as a target.

4.2 The Propagation Algorithm

Figure 3 shows the detailed algorithm. In the algorithm, the opinion word lexicon O and review data R about a product are provided as the input. The steps are set following the propagation order. It stops when no more new opinion words or targets can be added. The algorithm has been explained in the preceding text. We will not repeat it here. We now use an example to illustrate the working of the algorithm. Assume we have the following four sentences in a review: *Canon G3 takes great pictures*, *The picture is amazing*, *You may have to get more storage to store high quality pictures and recorded movies*, and *The software is amazing*. We only have one input opinion word *great*. Using lines 4 to 6 in the algorithm, we can extract *picture* as a product feature (or target) based on $R1_1$. Given this extracted feature, we can then determine that *amazing* is also an opinion word using lines 16 to 18 based on $R2_2$, and *movies* also as a feature using lines 13 to 15 based on $R3_1$. In the second iteration, as *amazing* is recognized as an opinion word, *software* can be extracted as a feature using lines 4 to 6 based on $R1_2$. The propagation then stops as no more features or opinion words can be extracted. As we can see, through the propagation, the three product features (i.e., the targets) and new opinion words in the review are discovered using only a single opinion word.

4.3 Opinion Word Polarity Assignment

Polarities of opinion words are important for many opinion mining tasks. Thus, the newly extracted opinion words should be assigned with polarities. We now propose a polarity assignment method based on the contextual evidence. The basic intuition is

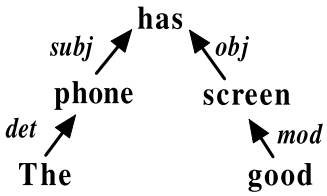


Figure 2
The dependency tree for the sentence *The phone has a good screen*.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

```
Input: Opinion Word Dictionary {O}, Review Data R
Output: All Possible Features {F}, The Expanded Opinion Lexicon {O-Expanded}
Function:
1.  {O-Expanded} = {O}
2.  {Fi} = Ø, {Oi} = Ø
3.  for each parsed sentence in R
4.      if(Extracted features not in {F})
5.          Extract features {Fi} using R11 and R12 based on opinion words in {O-Expanded}
6.      endif
7.      if(Extracted opinion words not in {O-Expanded})
8           Extract new opinion words {Oi} using R41 and R42 based on opinion words in {O-Expanded}
9.      endif
10. endfor
11. Set {F} = {F} + {Fi}, {O-Expanded} = {O-Expanded} + {Oi}
12. for each parsed sentence in R
13.     if(Extracted features not in {F})
14.         Extract features {F'} using R31 and R32 based on features in {Fi}
15.     endif
16.     if(Extracted opinion words not in {O-Expanded})
17.         Extract opinion words {O'} using R21 and R22 based on features in {Fi}
18.     endif
19. end for
20. Set {Fi} = {Fi} + {F'}, {Oi} = {Oi} + {O'}
21. Set {F} = {F} + {F'}, {O-Expanded} = {O-Expanded} + {O'}
22. Repeat 2 till size({Fi}) = 0, size({Oi}) = 0
```

Figure 3
The propagation algorithm.

that people often express their opinions in a consistent manner unless there are explicit *contrary words* such as *but* and *however*. In practice, the assignment is done during the opinion word extraction process. Before we describe our method, let us make some observations about opinion words and targets:

Observation 1 (same polarity for same target in a review): A *review* is a document written by a single reviewer. It is usually the case that the reviewer has the same sentiment or polarity on the same target, although the target may appear more than once in the review.

Observation 2 (same polarity for same opinion word in a domain corpus): It is usually the case that the same opinion word has the same polarity in one domain corpus.

Based on these observations, we assign polarities to both newly extracted targets and opinion words. The polarity of a target in a review is the identified sentiment polarity on the target given in the review. The following rules are exploited to infer polarities for extracted opinion words and targets:

1. Heterogeneous rule: For opinion words extracted by known targets, and targets extracted by known opinion words, we assign them the same polarities as the known ones. For example, if word *A* is an opinion word (or target) and *B* is a target (or opinion word) and *A* is extracted through *B*, *A* will be assigned with the same polarity as *B*. Note that targets themselves convey no polarities and opinion words are the only expressions that people use to show their attitudes towards targets. Therefore, the polarities of targets inherit those of associated opinion words. We also consider whether there are negations/contrary words associated with the opinion words (by examining each word in the surrounding 5-word window). (In our current work, the negation/contrary word set consists of *not*, *n't*, *'t*, *however*, *but*, *despite*, *though*, *except*, *although*, *oddly*, and *aside*. We compiled these words based on our experiences.)

2. Homogeneous rule: For opinion words extracted by known opinion words and targets extracted by known targets, we assign them the same polarities as the known

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

ones unless there are contrary words between them. For example, considering words A and B both targets (or opinion words) and A being extracted through B, if there are no contrary words between A and B, A will be assigned with the same polarity as B, otherwise the opposite polarity. We also observe that these words can cancel the polarity change when they are used together or associated with negations. Therefore, we consider that the polarity changes only when there is an odd number of such contrary words and negations between the two opinion words or targets.

3. Intra-review rule: There are new opinion words that are extracted by some targets from other reviews. These targets should convey no polarities in the current review because they do not conform to **Observation 1**. Hence, no polarities will be assigned to the opinion words. **Observation 2** cannot be applied either if these opinion words are found only in the current review. To assign polarities to such opinion words, we make use of the overall review polarity to infer. We assume that the opinion word takes the polarity of the review, that is, if the review is positive, the opinion word is assigned with positive polarity, otherwise negative polarity. The review polarity value is computed as the sum of polarity values of the contained known opinion words (+1 for positive polarity and -1 for negative polarity). If the final sum is greater than 0, the review is positive, and negative otherwise.

Note that, due to the two observations, multiple polarities may be assigned to an opinion word or target. To resolve conflict, we sum the polarity values. A positive polarity is +1 and a negative polarity is -1. If the sum is larger than 0, the final polarity is positive, otherwise negative.

4.4 Opinion Target Pruning

During the propagation, noise (incorrect targets and opinion words) may be introduced besides genuine targets and opinion words. We now describe some methods for removing noisy target words. We do not perform pruning of extracted opinion words as our current results give balanced precision and recall, which is desirable. We still do not have an effective method for pruning opinion words to achieve better results. We will study it in our future work.

One major type of target noise is the ordinary nouns that are not targets but are extracted as targets due to parsing errors or their associations with opinion words or targets. Another major kind of error in product reviews is the name of other competing products or dealers on which the reviewers also expressed opinions. We propose two corresponding pruning methods to identify these two types of noise. So far, all extracted targets are individual words (such as *weight*, *size*). However, because many targets are phrases (such as *battery life*), we need to identify them from the extracted individual words. A third pruning technique is to filter the remaining non-targets after the target phrase identification. Note that the first and third pruning techniques can be used for other kinds of opinion texts as well as product reviews.

4.4.1 Pruning Based on Clauses. We correct the first type of errors by using the following observation: *A sentence clause usually contains only one target unless there are conjunctions such as "and" and "or."* For example, in the sentence *I bought apex a month ago* in a review for *Apex DVD Player*, both *apex* and *month* were extracted as potential targets in some other sentences based on the rules. As these two potential targets are in the same clause (we identify the boundary of a clause using Minipar) and are not connected by a conjunction, one of them has to be removed. We call this method **clause pruning**.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

In this work, we filter non-targets based on frequency. That is, the one which is less frequent in the data set is removed. The reason for using the frequency-based pruning is that although reviewers usually have different things to say, when they comment on the same product features, they tend to use similar words (Hu and Liu 2004).

4.4.2 Pruning of Other Products and Dealers. We use a heuristic method to prune such non-targets. The basic idea is that when people compare the product under review with other products, there are indications such as *compare to*, *better than*; when they mention dealers/merchants, the indications are patterns such as *shop with*. We denote the indications for products as {*ProductINDI*} (currently including *compare to*, *compare with*, *better than*, *worse than*) and indications for dealers as {*DealerINDI*} (currently including *shop with* and *buy from*). In this heuristic method, we take the nearest nouns following any indication in {*ProductINDI*} as product names and those behind any indication in {*DealerINDI*} as dealer names. The distance is measured by the number of words in between. This simple heuristic method works quite well in practice for product reviews. However, we should note that this domain-specific pruning method does not always need to be employed as it trades recall for precision, but does not improve F-score. Our target extraction method is already quite strong. This pruning, however, does present some useful options in applications.

4.4.3 Identifying Target Phrases and Global Pruning. So far all the extracted targets are individual words. After removal of non-target individual words, we identify target phrases. As we consider targets to be nouns/noun phrases, we identify target phrases by combining each target word with Q consecutive nouns right before and after the target word, and K adjectives before the target word. We set $Q = 2$, $K = 1$ in our experiments. After obtaining the target phrases, we conduct another frequency-based pruning to remove targets that appear only once in the whole opinion data. This *global pruning* compensates for the *clause pruning* in case that a clause contains only one target which is kept no matter what its frequency is. Note that in *clause pruning*, it is possible to prune some words in a target phrase (with two or more words), but they will be recovered in this step by the combination as long as one of the words in the phrase is not pruned.

5. Experiments and Discussions

We now present the experimental results on opinion lexicon expansion and target extraction. We use the customer review collection[3] from Hu and Liu (2004) as the testing data. The collection contains five review data sets: two on two digital cameras, one on a DVD player, one on an mp3 player, and one on a cell phone. The detailed information of each review data set is shown in Table 2. The targets (i.e., product features) in these reviews are already labeled. Although the opinion expressed on each target in each sentence is also labeled, the polarities (or orientations) of opinion words are not labeled. In our experiments, we manually labeled the opinion words and their polarities. The seed opinion lexicon is also provided by Hu and Liu (2004), which contains 654 positive and 1,098 negative opinion words.

3 `http://www.cs.uic.edu/liub/FBS/sentiment-analysis.html`.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

Table 2
Detailed information of the five review data sets.

Data set	Number of reviews	Number of sentences
D1	45	597
D2	34	346
D3	41	546
D4	95	1,716
D5	99	740
Avg	62.8	789

5.1 Experiments on Opinion Lexicon Expansion

For the comparison of our approach in opinion lexicon expansion, we implemented the approach in Kanayama and Nasukawa (2006; referred to as KN06 hereafter). Details about this approach were given in Section 2. We only considered adjectives as the candidates in our experiments because our method is only concerned with adjective opinion words. As propagation is not performed in KN06, we also implemented a non-propagation version of our approach, in which opinion words are only extracted by the seed words and targets which are extracted by both the seeds and extracted opinion words. Furthermore, as our tasks can be regarded as a sequential labeling problem (to label if a word is an opinion word, a target, or an ordinary word), we experimented with the conditional random fields (CRF) technique (Lafferty, McCallum, and Pereira 2001) for extraction, which is a popular information extraction method and has been successfully used in labeling tasks such as POS tagging (Lafferty, McCallum, and Pereira 2001) and Named Entity Recognition (Finkel, Grenager, and Manning 2005). The well-known toolkit CRF++[4] is employed. We consider two kinds of processing windows, one using the whole sentence (CRF); the other using words between any pair of adjective and noun (CRF-D). In the first case, we designed seven labels for training, *product features*, *non-feature nouns*, *opinion adjectives*, *non-opinion adjectives*, *verbs*, *prepositions/conjunctions*, and *others*. In the second case, we took advantage of the relations on the shortest dependency path between the two words and used them as labels. In this way, CRF is made to capture long range dependencies between words. For both cases, we use the default parameter settings in CRF++.

To train CRF for the extraction task, we use one data set for training and the remaining four sets for testing. Consequently, we have five runs. The average results are reported here. In the set-up of our approaches and KN06, to examine the accuracy in extracting opinion words with different numbers of seeds we divide the initial opinion lexicon into 10 subsets, each with roughly the same number of words. We call these lists of opinion words the **10p lists**. These ten 10p lists are combined to produce 20p, 50p, and 80p lists which mean 20%, 50%, and 80% of the original set (1,752 opinion words), respectively. The experiments using four kinds of seed lists are performed separately.

Note that all metrics (precision, recall, and F-score) are computed on the newly extracted opinion words. This is an important point because only the new extractions are meaningful. Using all the extracted words to compute precision and recall is not

4 http://crfpp.sourceforge.net/.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

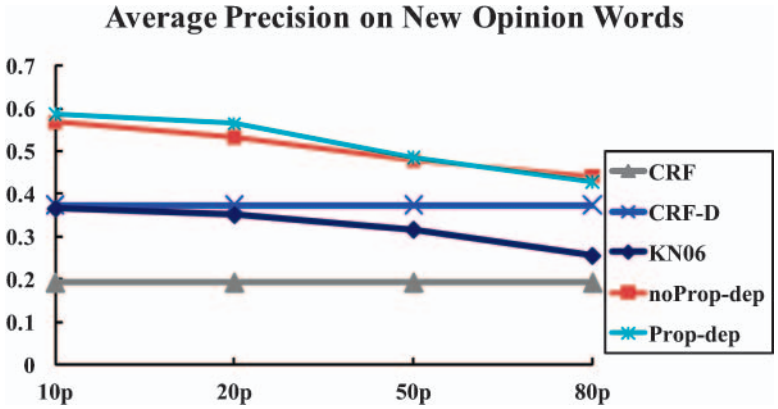


Figure 4
Precisions of CRF, CRF-D, KN06, noProp-dep, and Prop-dep.

appropriate as they can include many words that are already in the seed list or the labeled training set in the case of CRF.

5.1.1 Comparison Results and Discussions. Figures 4, 5, and 6 show the average results of precision, recall, and F-score of different approaches using different numbers of seed opinion words. CRF and CRF-D are not evaluated against the number of seed opinion words because that results in too few training data for CRF and CRF-D and consequently poorer results. That is also why their results stay the same for all cases. Prop-dep is our propagation approach and noProp-dep is the non-propagation version of our technique.

Observing from Figure 4, we can see that our approaches, both propagation and non-propagation versions, outperform others in all the four cases in precision. This indicates that our rules based on the dependency relations are effective in extracting correct opinion words. The precision of CRF is low, which means CRF has difficulty in distinguishing ordinary adjectives from opinion ones. The better performance of CRF-D over CRF indicates that long-range dependency relations are helpful. KN06 is reported to have around 60% precision in the Japanese test data, but it does not perform as well in

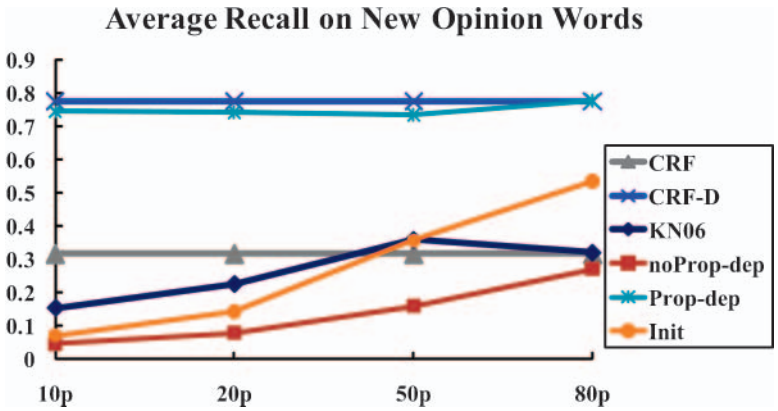


Figure 5
Recalls of Init, CRF, CRF-D, KN06, noProp-dep, and Prop-dep.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

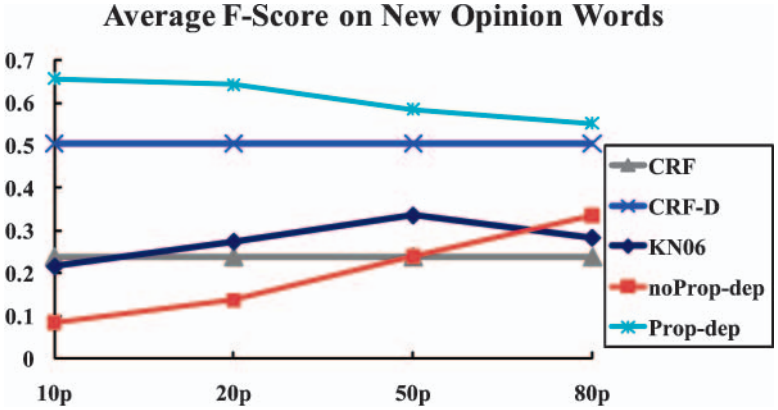


Figure 6
F-scores of CRF, CRF-D, KN06, noProp-dep, and Prop-dep.

our experiments. One reason could be that the statistical estimation of KN06 measures a word's positive or negative occurrences compared to its total occurrences, which can introduce unreliability if words are infrequent when the corpus is small. Considering the size of the testing data in our experiments (which is much smaller than theirs), the estimation can be quite unreliable. Many infrequent non-opinion words are identified as opinion words, which lowers the precision. In our technique, rules are applied in terms of single sentences. Thus it is not sensitive to the size of the testing data. Another observation is that in our approaches, the best performance is gained at 10p rather than 80p. This is because at 80p most of the opinion words are already known (in the seed list) and the number of remaining ones to be extracted is small and they are usually harder to identify.

From Figure 5, we can see that our approach makes significant improvement over others in recall except CRF-D. Clearly, the propagation is at work. In the 10p case, the new opinion words extracted by our approach could cover almost 75% of the whole opinion set whereas the corresponding seed words only cover 8% of the opinion words in the data (see the Init line). Thus our propagation method is quite powerful in identifying a large number of new opinion words. We also notice that the results do not change dramatically in different cases (all around 75%), which shows that the propagation performs steadily in extracting new opinion words with a different number of seeds (although it does introduce more noise as the number of seeds increases, as shown in Figure 4.). CRF is found to cover only about 30% of the opinion words. Technically, CRF captures only local patterns rather than long-range patterns. Many dependency relationships are long range (i.e., there are many words between the opinion word and the feature that it modifies), which explains the weak performance of CRF. CRF-D performs the best (about 78%), which confirms the usefulness of long range patterns. However, considering the large size of training data for CRF-D (the training data set already contains most of the correct opinion words) and poorer precision (Figure 4), its result is weaker than our approach. KN06 performs poorly in finding new opinion words, which we believe is due to its strategy in selecting candidates. The strategy only considers adjectives in successive sentences and does not use features or any dependency relationships. Such relationships clearly exist and are useful.

Figure 6 shows the F-score results. In all four cases, our propagation approach (Prop-dep) achieves the highest F-score. We can thus draw the conclusion that our

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

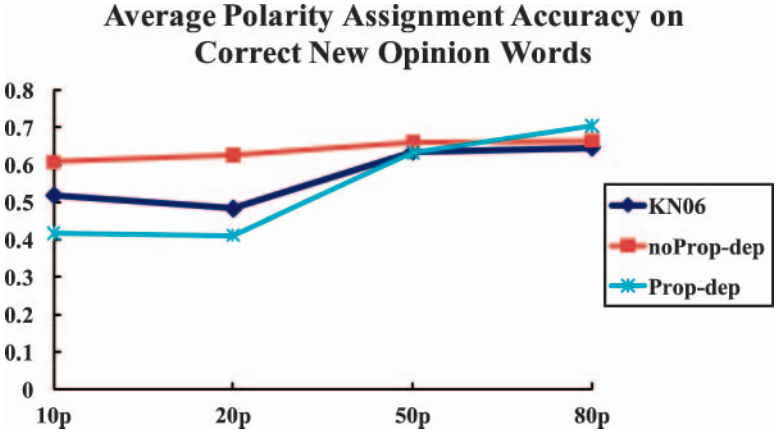


Figure 7
Average polarity assignment accuracy on correct new opinion words.

approach is superior to the existing methods. It indicates that rules defined based on the relations are effective and the propagation idea is powerful.

5.1.2 Results of Polarity Assignment. Figure 7 shows the accuracy of polarity assignment of different approaches computed on the newly discovered correct opinion words by each approach. From the results, we can see that noProp-dep performs the best before around 65p (which means 65% of the complete opinion lexicon). Prop-dep performs worse than KN06 but approaches it when the number of seeds increases, and outperforms KN06 from around 50p and noProp-dep from around 65p. Considering our approach has a much higher recall, about 20% higher at 80p (Figure 5), this result is useful. At 10p, 20p, and 50p, the recall values of our methods are even higher than KN06. In the cases of 50p and 80p, our method becomes more accurate than KN06 (Figure 7). We consider those two cases to be realistic for practical applications because there are already several existing opinion lexicons compiled by researchers. Thus, in practice, one does not need to start with a very small number of seeds. Note that this does not conflict with our earlier statement about our propagation approach's ability in extracting a large number of new opinion words with only a small set of seeds.

5.2 Experiments on Opinion Target Extraction

For performance evaluation on opinion target extraction, we compare our work (and also the non-propagation version, i.e., extracting targets using only the opinion words) with those in Hu and Liu (2004, henceforth Hu) and Popescu and Etzioni (2005, henceforth Popescu), which also considered only explicit noun targets and experimented with the same data sets. Details of both approaches have been described in Section 2. Additionally, we experimented with the popular topic modeling algorithm PLSA (Hofmann 1999), using a public domain program,[5] and CRF (CRF-D) using the toolkit CRF++. The parameters and training set-ups were set the same as in the opinion word extraction experiment. Note in this set of experiments, all our initial opinion words were used.

5 `http://www.kyb.mpg.de/bs/people/pgehler/code/index.html`.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

In PLSA, the maximum number of iterations was set to 500. As PLSA only clusters words of the same rough topic together but does not perform fine-grained target extraction directly, we computed the precision, recall, and F-score results by combining the top M nouns of each cluster together as the extracted targets by PLSA. The value of M and the number of clusters were chosen empirically. We set M as 10, 20, and 30 and number of clusters as 10, 20, 30, 40, and 50. We used the best results based on F-scores as the final results of each data set for PLSA. For the five data sets, the best results for the DVD player were gained at $M = 10$ with the number of clusters being 30, and those for the remaining four data sets were all gained at $M = 20$ with the number of clusters being 10.

5.2.1 Comparison Results and Discussions. Tables 3, 4, and 5 show the precision, recall, and F-score results respectively of our propagation approach (Prop-dep), the non-propagation version (noProp-dep), Hu, Popescu, PLSA, and CRF (CRF-D).

From Table 3, we can see that on average our approach has a 16% improvement in precision over Hu, 34% over PLSA, 28% over CRF, and 30% over CRF-D, and has similar results to Popescu. noProp-dep performs the best, which indicates that the rules are helpful in extracting targets, but the propagation introduces noise. The results of Hu and Popescu are taken from their respective papers as their systems are not available for experimentation. The Popescu approach is much more time consuming as it needs to complete an extensive Web search and also must know the product class in advance. Our proposed approach relies only on the review data itself and

Table 3
Precisions of our propagation approach (Prop-dep), the non-propagation version (noProp-dep), Hu, Popescu, PLSA, CRF, and CRF-D.

	Precision						
	Hu	Popescu	PLSA	CRF	CRF-D	noProp-dep	Prop-dep
D1	0.75	0.89	0.63	0.62	0.59	0.94	0.87
D2	0.71	0.87	0.48	0.64	0.58	0.97	0.90
D3	0.72	0.89	0.56	0.58	0.57	0.97	0.90
D4	0.69	0.86	0.53	0.53	0.54	0.88	0.81
D5	0.74	0.90	0.49	0.64	0.62	0.92	0.92
Avg	0.72	0.88	0.54	0.60	0.58	0.94	0.88

Table 4
Recalls of our propagation approach (Prop-dep), the non-propagation version (noProp-dep), Hu, Popescu, PLSA, CRF, and CRF-D.

	Recall						
	Hu	Popescu	PLSA	CRF	CRF-D	noProp-dep	Prop-dep
D1	0.82	0.80	0.53	0.37	0.52	0.70	0.81
D2	0.79	0.74	0.59	0.41	0.59	0.59	0.81
D3	0.76	0.74	0.56	0.30	0.45	0.67	0.86
D4	0.82	0.80	0.47	0.35	0.50	0.72	0.84
D5	0.80	0.78	0.59	0.27	0.45	0.64	0.86
Avg	0.80	0.78	0.55	0.34	0.50	0.66	0.83

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

Table 5
F-scores of our propagation approach (Prop-dep), the non-propagation version (noProp-dep), Hu, Popescu, PLSA, CRF, and CRF-D.

	F-score						
	Hu	Popescu	PLSA	CRF	CRF-D	noProp-dep	Prop-dep
D1	0.78	0.84	0.58	0.46	0.55	0.80	0.84
D2	0.75	0.80	0.53	0.50	0.58	0.73	0.85
D3	0.74	0.81	0.56	0.40	0.50	0.79	0.88
D4	0.75	0.83	0.49	0.42	0.52	0.79	0.82
D5	0.77	0.84	0.54	0.38	0.52	0.75	0.89
Avg	0.76	0.82	0.54	0.43	0.54	0.77	0.86

no external information is needed. The poor results of PLSA show that many of the top-ranked terms in clusters are not genuine opinion targets but other words, such as *telephone*, *message*, *location*, *buyer* for the DVD player data set. The low precisions of CRF and CRF-D show that CRF performs poorly in distinguishing targets and non-targets.

Table 4 shows that our approach outperforms all the other approaches in recall. We have 3% improvement over Hu, 5% over Popescu, 28% over PLSA, 49% over CRF, and 33% over CRF-D. However, if we do not perform propagation, the recall results are much worse than Hu and Popescu, as shown in noProp-dep. Therefore, propagation is necessary in extracting a large number of targets. The reason for the improvement of our approach over Popescu is quite evident. Although some of the product features tend to appear with product discriminators like *product has*, *of product*, there are many other features that do not have high co-occurrences with these discriminators. The better performance of CRF-D over CRF again indicates that dependency relations are helpful in identifying more targets.

F-scores in Table 5 show that our approach outperforms Hu (by 10%), Popescu (4%), PLSA (32%), CRF (43%), and CRF-D (32%). Paired t-tests show that all the improvements are statistically significant at the confidence level of 95%. noProp-dep also has good F-score results (even a little better than Hu) due to its high precision. The poor results seen with PLSA and CRF indicate that these approaches may not be suitable for this task. As analyzed earlier, PLSA is known for its ability to mine rough topics in a large text collection by clustering topic words together based on the probability distribution in the corpus. It is not suitable for the fine-grained extraction task. Therefore, a large number of non-targets are included. CRF and CRF-D perform poorly as explained in the opinion word extraction task.

6. Conclusions

This article focuses on two important tasks in opinion mining, namely, opinion lexicon expansion and target extraction. We propose a propagation approach to extract opinion words and targets iteratively given only a seed opinion lexicon of small size. The extraction is performed using identified relations between opinion words and targets, and also opinion words/targets themselves. The relations are described syntactically based on the dependency grammar. We also propose novel methods for new opinion word polarity assignment and noisy target pruning. In the evaluation, we compared our new

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

approach with others on standard testing data sets. The results show that our approach outperforms other state-of-the-art methods in these two tasks. In the future, we plan to first focus on improving the precision of opinion word extraction by working on opinion word pruning methods. We will then also try to learn syntactic relations automatically from large corpuses using pattern mining techniques to improve the relation coverage.

Acknowledgments

We thank Xiaowen Ding, and Lei Zhang from University of Illinois at Chicago for helpful discussions for our work; Xiyun Gan, Xiao Cheng, Dazhou Wang, and Xuan Zhao from Zhejiang University for their contributions to the annotations. This work was supported by National Key Technology R&D Programs 2008BAH26B00 and 2007BAH11B06.

References

Breck, Eric, Yejin Choi, and Claire Cardie. 2007. Identifying expressions of opinion in context. In *Proceedings of IJCAI'07*, pages 2683–2688. Menlo Park, CA.

Esuli, Andrea and Fabrizio Sebastiani. 2005. Determining the semantic orientation of terms through gloss classification. In *Proceedings of CIKM'05*, pages 617–624. New York, NY.

Finkel, Jenny Rose, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of ACL'05*, pages 363–370. Stroudsburg, PA.

Hatzivassiloglou, Vasileios and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of ACL'97*, pages 174–181. Stroudsburg, PA.

Hofmann, Thomas. 1999. Probabilistic latent semantic analysis. In *Proceedings of UAI'99*, pages 289–296. San Francisco, CA.

Hu, Mingqing and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of SIGKDD'04*, pages 168–177.

Kaji, Nobuhiro and Masaru Kitsuregawa. 2007. Building lexicon for sentiment analysis from massive collection of html documents. In *Proceedings of EMNLP'07*, pages 1075–1083.

Kamps, Jaap, Maarten Marx, Robert J. Mokken, and Maarten de Rijke. 2004. Using Wordnet to measure semantic orientation of adjectives. In *Proceedings of LREC'04*, pages 1115–1118.

Kanayama, Hiroshi and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of EMNLP'06*, pages 355–363.

Kim, Soo-Min and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of COLING'04*, pages 1367–1373.

Kobayashi, Nozomi, Kentaro Inui, and Yuji Matsumoto. 2007. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proceedings of EMNLP'07*, pages 1065–1074.

Lafferty, John, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of ICML'01*, pages 282–289.

Liu, Bing. 2006. *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*. Springer, Berlin.

Mei, Qiaozhu, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. 2007. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *Proceedings of WWW'07*, pages 171–180.

Pang, Bo and Lillian Lee. 2008. *Opinion Mining and Sentiment Analysis*. Now Publishers Inc., Hanover, MA.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP'02*, pages 79–86.

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

Scaffidi, Christopher, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, and Chun Jin. 2007. Red opal: Product-feature scoring from reviews. In *Proceedings of EC'07*, pages 182–191.

Stoyanov, Veselin and Claire Cardie. 2008. Topic identification for fine-grained opinion analysis. In *Proceedings of COLING'08*, pages 817–824.

Takamura, Hiroya, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of ACL'05*, pages 133–140.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

Takamura, Hiroya, Takashi Inui, and Manabu Okumura. 2007. Extracting semantic orientations of phrases from dictionary. In *Proceedings of NAACL HLT'07*, pages 292–299.

Tesniere, Lucien. 1959. *Elements de Syntaxe Structurale*. Librairie C. Klincksieck, Paris.

Turney, Peter D. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL'02*, pages 417–424.

Turney, Peter D. and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information System*, 21(4):315–346.

Wiebe, Janyce. 2000. Learning subjective adjective from corpora. In *Proceedings of AAAI'00*, pages 735–740.

Wiebe, Janyce, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational Linguistics*, 30(3):277–308.

Wong, Tak-Lam, Wai Lam, and Tik-Shun Wong. 2008. An unsupervised framework for extracting and normalizing product attributes from multiple Web sites. In *Proceedings of SIGIR'08*, pages 35–42.

Yu, Hong and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of EMNLP'03*, pages 129–136.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034 by guest on 29 March 2024