

# Syntactic Patterns versus Word Alignment: Extracting Opinion Targets from Online Reviews

Kang Liu, Liheng Xu and Jun Zhao

National Laboratory of Pattern Recognition  
Institute of Automation, Chinese Academy of Sciences  
{kliu, lhxu, jzhao}@nlpr.ia.ac.cn

## Abstract

Mining opinion targets is a fundamental and important task for opinion mining from online reviews. To this end, there are usually two kinds of methods: syntax based and alignment based methods. Syntax based methods usually exploited syntactic patterns to extract opinion targets, which were however prone to suffer from parsing errors when dealing with online informal texts. In contrast, alignment based methods used word alignment model to fulfill this task, which could avoid parsing errors without using parsing. However, there is no research focusing on which kind of method is more better when given a certain amount of reviews. To fill this gap, this paper empirically studies how the performance of these two kinds of methods vary when changing the size, domain and language of the corpus. We further combine syntactic patterns with alignment model by using a partially supervised framework and investigate whether this combination is useful or not. In our experiments, we verify that our combination is effective on the corpus with small and medium size.

## 1 Introduction

With the rapid development of Web 2.0, huge amount of user reviews are springing up on the Web. Mining opinions from these reviews become more and more urgent since that customers expect to obtain fine-grained information of products and manufacturers need to obtain immediate feedbacks from customers. In opinion mining, extracting opinion targets is a basic subtask. It is to extract a list of the objects which users express their opinions on and can provide the prior information of targets for opinion mining. So this task

has attracted many attentions. To extract opinion targets, previous approaches usually relied on opinion words which are the words used to express the opinions (Hu and Liu, 2004a; Popescu and Etzioni, 2005; Liu et al., 2005; Wang and Wang, 2008; Qiu et al., 2011; Liu et al., 2012). Intuitively, opinion words often appear around and modify opinion targets, and there are opinion relations and associations between them. If we have known some words to be opinion words, the words which those opinion words modify will have high probability to be opinion targets.

Therefore, identifying the aforementioned opinion relations between words is important for extracting opinion targets from reviews. To fulfill this aim, previous methods exploited the words co-occurrence information to indicate them (Hu and Liu, 2004a; Hu and Liu, 2004b). Obviously, these methods cannot obtain precise extraction because of the diverse expressions by reviewers, like long-span modified relations between words, etc. To handle this problem, several methods exploited syntactic information, where several heuristic patterns based on syntactic parsing were designed (Popescu and Etzioni, 2005; Qiu et al., 2009; Qiu et al., 2011). However, the sentences in online reviews usually have informal writing styles including grammar mistakes, typos, improper punctuation etc., which make parsing prone to generate mistakes. As a result, the syntax-based methods which heavily depended on the parsing performance would suffer from parsing errors (Zhang et al., 2010). To improve the extraction performance, we can only employ some exquisite high-precision patterns. But this strategy is likely to miss many opinion targets and has lower recall with the increase of corpus size. To resolve these problems, Liu et al. (2012) formulated identifying opinion relations between words as a monolingual alignment process. A word can find its corresponding modifiers by using a word alignment

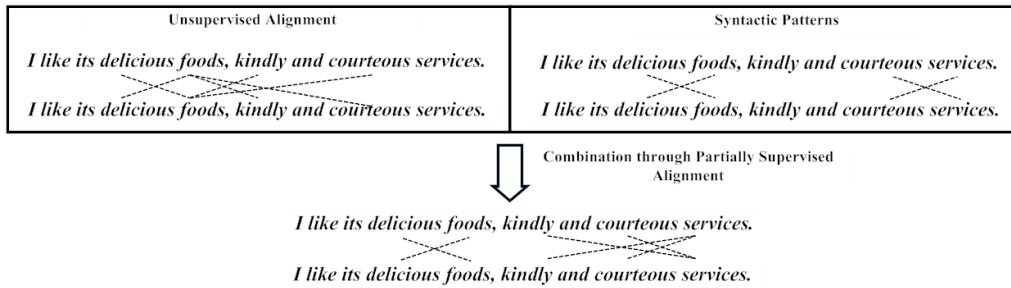


Figure 1: Mining Opinion Relations between Words using Partially Supervised Alignment Model

model (WAM). Without using syntactic parsing, the noises from parsing errors can be effectively avoided. Nevertheless, we notice that the alignment model is a statistical model which needs sufficient data to estimate parameters. When the data is insufficient, it would suffer from data sparseness and may make the performance decline.

Thus, from the above analysis, we can observe that the size of the corpus has impacts on these two kinds of methods, which arises some important questions: how can we make selection between syntax based methods and alignment based method for opinion target extraction when given a certain amount of reviews? And which kind of methods can obtain better extraction performance with the variation of the size of the dataset? Although (Liu et al., 2012) had proved the effectiveness of WAM, they mainly performed experiments on the dataset with medium size. We are still curious about that when the size of dataset is larger or smaller, can we obtain the same conclusion? To our best knowledge, these problems have not been studied before. Moreover, opinions may be expressed in different ways with the variation of the domain and language of the corpus. When the domain or language of the corpus is changed, what conclusions can we obtain? To answer these questions, in this paper, we adopt a unified framework to extract opinion targets from reviews, in the key component of which we vary the methods between syntactic patterns and alignment model. Then we run the whole framework on the corpus with different size (from #500 to #1,000,000), domain (three domains) and language (Chinese and English) to empirically assess the performance variations and discuss which method is more effective.

Furthermore, this paper naturally addresses another question: is it useful for opinion targets extraction when we combine syntactic patterns and word alignment model into a unified model? To

this end, we employ a partially supervised alignment model (PSWAM) like (Gao et al., 2010; Liu et al., 2013). Based on the exquisitely designed high-precision syntactic patterns, we can obtain some precisely modified relations between words in sentences, which provide a portion of links of the full alignments. Then, these partial alignment links can be regarded as the constrains for a standard unsupervised word alignment model. And each target candidate would find its modifier under the partial supervision. In this way, the errors generated in standard unsupervised WAM can be corrected. For example in Figure 1, “*kindly*” and “*courteous*” are incorrectly regarded as the modifiers for “*foods*” if the WAM is performed in an whole unsupervised framework. However, by using some high-precision syntactic patterns, we can assert “*courteous*” should be aligned to “*services*”, and “*delicious*” should be aligned to “*foods*”. Through combination under partial supervision, we can see “*kindly*” and “*courteous*” are correctly linked to “*services*”. Thus, it’s reasonable to expect to yield better performance than traditional methods. As mentioned in (Liu et al., 2013), using PSWAM can not only inherit the advantages of WAM: effectively avoiding noises from syntactic parsing errors when dealing with informal texts, but also can improve the mining performance by using partial supervision. However, is this kind of combination always useful for opinion target extraction? To access this problem, we also make comparison between PSWAM based method and the aforementioned methods in the same corpora with different size, language and domain. The experimental results show the combination by using PSWAM can be effective on dataset with small and medium size.

## 2 Related Work

Opinion target extraction isn't a new task for opinion mining. There are much work focusing on this task, such as (Hu and Liu, 2004b; Ding et al., 2008; Li et al., 2010; Popescu and Etzioni, 2005; Wu et al., 2009). Totally, previous studies can be divided into two main categories: supervised and unsupervised methods.

In supervised approaches, the opinion target extraction task was usually regarded as a sequence labeling problem (Jin and Huang, 2009; Li et al., 2010; Ma and Wan, 2010; Wu et al., 2009; Zhang et al., 2009). It's not only to extract a lexicon or list of opinion targets, but also to find out each opinion target mentions in reviews. Thus, the contextual words are usually selected as the features to indicate opinion targets in sentences. And classical sequence labeling models are used to train the extractor, such as CRFs (Li et al., 2010), HMM (Jin and Huang, 2009) etc.. Jin et al. (2009) proposed a lexicalized HMM model to perform opinion mining. Both Li et al. (2010) and Ma et al. (2010) used CRFs model to extract opinion targets in reviews. Specially, Li et al. proposed a Skip-Tree CRF model for opinion target extraction, which exploited three structures including linear-chain structure, syntactic structure, and conjunction structure. However, the main limitation of these supervised methods is the need of labeled training data. If the labeled training data is insufficient, the trained model would have unsatisfied extraction performance. Labeling sufficient training data is time and labor consuming. And for different domains, we need label data independently, which is obviously impracticable.

Thus, many researches focused on unsupervised methods, which are mainly to extract a list of opinion targets from reviews. Similar to ours, most approaches regarded opinion words as the indicator for opinion targets. (Hu and Liu, 2004a) regarded the nearest adjective to an noun/noun phrase as its modifier. Then it exploited an association rule mining algorithm to mine the associations between them. Finally, the frequent explicit product features can be extracted in a bootstrapping process by further combining item's frequency in dataset. Only using nearest neighbor rule to mine the modifier for each candidate cannot obtain precise results. Thus, (Popescu and Etzioni, 2005) used syntax information to extract opinion targets, which designed some syntactic patterns to capture

the modified relations between words. The experimental results showed that their method had better performance than (Hu and Liu, 2004a). Moreover, (Qiu et al., 2011) proposed a Double Propagation method to expand sentiment words and opinion targets iteratively, where they also exploited syntactic relations between words. Specially, (Qiu et al., 2011) didn't only design syntactic patterns for capturing modified relations, but also designed patterns for capturing relations among opinion targets and relations among opinion words. However, the main limitation of Qiu's method is that the patterns based on dependency parsing tree may miss many targets for the large corpora. Therefore, Zhang et al. (2010) extended Qiu's method. Besides the patterns used in Qiu's method, they adopted some other special designed patterns to increase recall. In addition they used the HITS (Kleinberg, 1999) algorithm to compute opinion target confidences to improve the precision. (Liu et al., 2012) formulated identifying opinion relations between words as an alignment process. They used a completely unsupervised WAM to capture opinion relations in sentences. Then the opinion targets were extracted in a standard random walk framework where two factors were considered: opinion relevance and target importance. Their experimental results have shown that WAM was more effective than traditional syntax-based methods for this task. (Liu et al., 2013) extend Liu's method, which is similar to our method and also used a partially supervised alignment model to extract opinion targets from reviews. We notice these two methods ((Liu et al., 2012) and (Liu et al., 2013)) only performed experiments on the corpora with a medium size. Although both of them proved that WAM model is better than the methods based on syntactic patterns, they didn't discuss the performance variation when dealing with the corpora with different sizes, especially when the size of the corpus is less than 1,000 and more than 10,000. Based on their conclusions, we still don't know which kind of methods should be selected for opinion target extraction when given a certain amount of reviews.

## 3 Opinion Target Extraction Methodology

To extract opinion targets from reviews, we adopt the framework proposed by (Liu et al., 2012), which is a graph-based extraction framework and

has two main components as follows.

1) **The first component** is to capture opinion relations in sentences and estimate associations between opinion target candidates and potential opinion words. In this paper, we assume opinion targets to be nouns or noun phrases, and opinion words may be adjectives or verbs, which are usually adopted by (Hu and Liu, 2004a; Qiu et al., 2011; Wang and Wang, 2008; Liu et al., 2012). And a potential opinion relation is comprised of an opinion target candidate and its corresponding modified word.

2) **The second component** is to estimate the confidence of each candidate. The candidates with higher confidence scores than a threshold will be extracted as opinion targets. In this procedure, we formulate the associations between opinion target candidates and potential opinion words in a bipartite graph. A random walk based algorithm is employed on this graph to estimate the confidence of each target candidate.

In this paper, we fix the method in the second component and vary the algorithms in the first component. In the first component, we respectively use syntactic patterns and unsupervised word alignment model (WAM) to capture opinion relations. In addition, we employ a partially supervised word alignment model (PSWAM) to incorporate syntactic information into WAM. In experiments, we run the whole framework on the different corpora to discuss which method is more effective. In the following subsections, we will present them in detail.

### 3.1 The First Component: Capturing Opinion Relations and Estimating Associations between Words

#### 3.1.1 Syntactic Patterns

To capture opinion relations in sentences by using syntactic patterns, we employ the manual designed syntactic patterns proposed by (Qiu et al., 2011). Similar to Qiu, only the syntactic patterns based on the direct dependency are employed to guarantee the extraction qualities. The direct dependency has two types. The first type indicates that one word depends on the other word without any additional words in their dependency path. The second type denotes that two words both depend on a third word directly. Specifically, we employ Minipar<sup>1</sup> to parse sentences. To further make syn-

tactic patterns precisely, we only use a few dependency relation labels outputted by Minipar, such as mod, pnm, subj, desc etc. To make a clear explanation, we give out some syntactic pattern examples in Table 1. In these patterns, *OC* is a potential opinion word which is an adjective or a verb. *TC* is an opinion target candidate which is a noun or noun phrase. The item on the arrows means the dependency relation type. The item in parenthesis denotes the part-of-speech of the other word. In these examples, the first three patterns are based on the first direct dependency type and the last two patterns are based on the second direct dependency type.

Pattern#1: <OC> $\xrightarrow{mod}$ <TC>
Example: This phone has an <i>amazing design</i>
Pattern#2: <TC> $\xrightarrow{obj}$ <OC>
Example: I <i>like</i> this <i>phone</i> very much
Pattern#3: <OC> $\xrightarrow{pnm}$ <TC>
Example: the <i>buttons</i> <i>easier</i> to use
Pattern#4: <OC> $\xrightarrow{mod}$ (NN) $\xleftarrow{subj}$ <TC>
Example: <i>IPhone</i> is a <i>revolutionary</i> smart phone
Pattern#5: <OC> $\xrightarrow{pred}$ (VBE) $\xleftarrow{subj}$ <TC>
Example: The <i>quality</i> of LCD is <i>good</i>

Table 1: Some Examples of Used Syntactic Patterns

#### 3.1.2 Unsupervised Word Alignment Model

In this subsection, we present our method for capturing opinion relations using unsupervised word alignment model. Similar to (Liu et al., 2012), every sentence in reviews is replicated to generate a parallel sentence pair, and the word alignment algorithm is applied to the monolingual scenario to align a noun/noun phrase with its modifiers. We select IBM-3 model (Brown et al., 1993) as the alignment model. Formally, given a sentence  $S = \{w_1, w_2, \dots, w_n\}$ , we have

$$P_{ibm3}(A|S) \propto \prod_{i=1}^N n(\phi_i|w_i) \prod_{j=1}^N t(w_j|w_{a_j}) d(j|a_j, N) \quad (1)$$

where  $t(w_j|w_{a_j})$  models the co-occurrence information of two words in dataset.  $d(j|a_j, N)$  models word position information, which describes the probability of a word in position  $a_j$  aligned with a word in position  $j$ . And  $n(\phi_i|w_i)$  describes the ability of a word for modifying (being modified by) several words.  $\phi_i$  denotes the number of words

<sup>1</sup><http://webdocs.cs.ualberta.ca/lindek/minipar.htm>

that are aligned with  $w_i$ . In our experiments, we set  $\phi_i = 2$ .

Since we only have interests on capturing opinion relations between words, we only pay attentions on the alignments between opinion target candidates (nouns/noun phrases) and potential opinion words (adjectives/verbs). If we directly use the alignment model, a noun (noun phrase) may align with other unrelated words, like prepositions or conjunctions and so on. Thus, we set constrains on the model: 1) Alignment links must be assigned among nouns/noun phrases, adjectives/verbs and null words. Aligning to null words means that this word has no modifier or modifies nothing; 2) Other unrelated words can only align with themselves.

### 3.1.3 Combining Syntax-based Method with Alignment-based Method

In this subsection, we try to combine syntactic information with word alignment model. As mentioned in the first section, we adopt a partially supervised alignment model to make this combination. Here, the opinion relations obtained through the high-precision syntactic patterns (Section 3.1.1) are regarded as the ground truth and can only provide a part of full alignments in sentences. They are treated as the constrains for the word alignment model. Given some partial alignment links  $\hat{A} = \{(k, a_k) | k \in [1, n], a_k \in [1, n]\}$ , the optimal word alignment  $A^* = \{(i, a_i) | i \in [1, n], a_i \in [1, n]\}$  can be obtained as  $A^* = \underset{A}{\operatorname{argmax}} P(A|S, \hat{A})$ , where  $(i, a_i)$  means that a noun (noun phrase) at position  $i$  is aligned with its modifier at position  $a_i$ .

Since the labeled data provided by syntactic patterns is not a full alignment, we adopt a EM-based algorithm, named as constrained hill-climbing algorithm (Gao et al., 2010), to estimate the parameters in the model. In the training process, the constrained hill-climbing algorithm can ensure that the final model is marginalized on the partial alignment links. Particularly, in the E step, their method aims to find out the alignments which are consistent to the alignment links provided by syntactic patterns, where there are main two steps involved.

1) *Optimize towards the constraints.* This step aims to generate an initial alignments for alignment model (IBM-3 model in our method), which can be close to the constraints. First, a simple alignment model (IBM-1, IBM-2, HMM etc.) is

trained. Then, the evidence being inconsistent to the partial alignment links will be got rid of by using the move operator  $m_{i,j}$  which changes  $a_j = i$  and the swap operator  $s_{j_1, j_2}$  which exchanges  $a_{j_1}$  and  $a_{j_2}$ . The alignment is updated iteratively until no additional inconsistent links can be removed.

2) *Towards the optimal alignment under the constraints.* This step aims to optimize towards the optimal alignment under the constraints which starts from the aforementioned initial alignments. Gao et.al. (2010) set the corresponding cost value of the invalid move or swap operation in  $M$  and  $S$  to be negative, where  $M$  and  $S$  are respectively called Moving Matrix and Swapping Matrix, which record all possible move and swap costs between two different alignments. In this way, the invalid operators will never be picked which can guarantee that the final alignment links to have high probability to be consistent with the partial alignment links provided by high-precision syntactic patterns.

Then in M-step, evidences from the neighbor of final alignments are collected so that we can produce the estimation of parameters for the next iteration. In the process, those statistics which come from inconsistent alignment links aren't be picked up. Thus, we have

$$P(w_i | w_{a_i}, \hat{A}) = \begin{cases} \lambda, & \text{otherwise} \\ P(w_i | w_{a_i}) + \lambda, & \text{inconsistent with } \hat{A} \end{cases} \quad (2)$$

where  $\lambda$  means that we make soft constraints on the alignment model. As a result, we expect some errors generated through high-precision patterns (Section 3.1.1) may be revised in the alignment process.

## 3.2 Estimating Associations between Words

After capturing opinion relations in sentences, we can obtain a lot of word pairs, each of which is comprised of an opinion target candidate and its corresponding modified word. Then the conditional probabilities between potential opinion target  $w_t$  and potential opinion word  $w_o$  can be estimated by using maximum likelihood estimation. Thus, we have  $P(w_t | w_o) = \frac{\operatorname{Count}(w_t, w_o)}{\operatorname{Count}(w_o)}$ , where  $\operatorname{Count}(\cdot)$  means the item's frequency information.  $P(w_t | w_o)$  means the conditional probabilities between two words. At the same time, we can obtain conditional probability  $P(w_o | w_t)$ . Then,

similar to (Liu et al., 2012), the association between an opinion target candidate and its modifier is estimated as follows.  $Association(w_t, w_o) = (\alpha \times P(w_t|w_o) + (1 - \alpha) \times P(w_o|w_t))^{-1}$ , where  $\alpha$  is the harmonic factor. We set  $\alpha = 0.5$  in our experiments.

### 3.3 The Second Component: Estimating Candidate Confidence

In the second component, we adopt a graph-based algorithm used in (Liu et al., 2012) to compute the confidence of each opinion target candidate, and the candidates with higher confidence than the threshold will be extracted as the opinion targets. Here, opinion words are regarded as the important indicators. We assume that two target candidates are likely to belong to the similar category, if they are modified by similar opinion words. Thus, we can propagate the opinion target confidences through opinion words.

To model the mined associations between words, a bipartite graph is constructed, which is defined as a weighted undirected graph  $G = (V, E, W)$ . It contains two kinds of vertex: opinion target candidates and potential opinion words, respectively denoted as  $v_t \in V$  and  $v_o \in V$ . As shown in Figure 2, the white vertices represent opinion target candidates and the gray vertices represent potential opinion words. An edge  $e_{v_t, v_o} \in E$  between vertices represents that there is an opinion relation, and the weight  $w$  on the edge represents the association between two words.

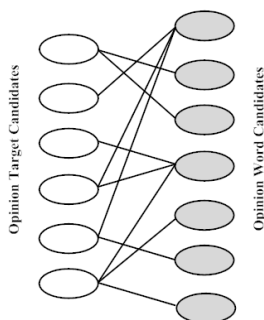


Figure 2: Modeling Opinion Relations between Words in a Bipartite Graph

To estimate the confidence of each opinion target candidate, we employ a random walk algorithm on our graph, which iteratively computes the weighted average of opinion target confidences from neighboring vertices. Thus we have

$$C^{i+1} = (1 - \beta) \times M \times M^T \times C^i + \beta \times I \quad (3)$$

where  $C^{i+1}$  and  $C^i$  respectively represent the opinion target confidence vector in the  $(i + 1)^{th}$  and  $i^{th}$  iteration.  $M$  is the matrix of word associations, where  $M_{i,j}$  denotes the association between the opinion target candidate  $i$  and the potential opinion word  $j$ . And  $I$  is defined as the prior confidence of each candidate for opinion target. Similar to (Liu et al., 2012), we set each item in  $I_v = \frac{tf(v)idf(v)}{\sum_v tf(v)idf(v)}$ , where  $tf(v)$  is the term frequency of  $v$  in the corpus, and  $idf(v)$  is computed by using the Google n-gram corpus<sup>2</sup>.  $\beta \in [0, 1]$  represents the impact of candidate prior knowledge on the final estimation results. In experiments, we set  $\beta = 0.4$ . The algorithm run until convergence which is achieved when the confidence on each node ceases to change in a tolerance value.

## 4 Experiments

### 4.1 Datasets and Evaluation Metrics

In this section, to answer the questions mentioned in the first section, we collect a large collection named as *LARGE*, which includes reviews from three different domains and different languages. This collection was also used in (Liu et al., 2012). In the experiments, reviews are first segmented into sentences according to punctuation. The detailed statistical information of the used collection is shown in Table 2, where Restaurant is crawled from the Chinese Web site: [www.dianping.com](http://www.dianping.com). The Hotel and MP3 are used in (Wang et al., 2011), which are respectively crawled from [www.tripadvisor.com](http://www.tripadvisor.com) and [www.amazon.com](http://www.amazon.com). For each dataset, we perform random sampling to generate testing set with different sizes, where we use sampled subsets with  $\#sentences = 5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4, 10^5$  and  $10^6$  sentences respectively. Each

Domain	Language	Sentence	Reviews
<b>Restaurant</b>	Chinese	1,683,129	395,124
<b>Hotel</b>	English	1,855,351	185,829
<b>MP3</b>	English	289,931	30,837

Table 2: Experimental Dataset

sentence is tokenized, part-of-speech tagged by using Stanford NLP tool<sup>3</sup>, and parsed by using Minipar toolkit. And the method of (Zhu et al., 2009) is used to identify noun phrases.

<sup>2</sup><http://books.google.com/ngrams/datasets>

<sup>3</sup><http://nlp.stanford.edu/software/tagger.shtml>

We select precision and recall as the metrics. Specifically, to obtain the ground truth, we manually label all opinion targets for each subset. In this process, three annotators are involved. First, every noun/noun phrase and its contexts in review sentences are extracted. Then two annotators were required to judge whether every noun/noun phrase is opinion target or not. If a conflict happens, a third annotator will make judgment for final results. The average inter-agreements is 0.74. We also perform a significant test, i.e., a t-test with a default significant level of 0.05.

## 4.2 Compared Methods

We select three methods for comparison as follows.

- **Syntax**: It uses syntactic patterns mentioned in Section 3.1.1 in the first component to capture opinion relations in reviews. Then the associations between words are estimated and the graph based algorithm proposed in the second component (Section 3.3) is performed to extract opinion targets.
- **WAM**: It is similar to **Syntax**, where the only difference is that **WAM** uses unsupervised WAM (Section 3.1.2) to capture opinion relations.
- **PSWAM** is similar to **Syntax** and **WAM**, where the difference is that **PSWAM** uses the method mentioned in Section 3.1.3 to capture opinion relations, which incorporates syntactic information into word alignment model by using partially supervised framework.

The experimental results on different domains are respectively shown in Figure 3, 4 and 5.

## 4.3 Syntax based Methods vs. Alignment based Methods

Comparing **Syntax** with **WAM** and **PSWAM**, we can obtain the following observations:

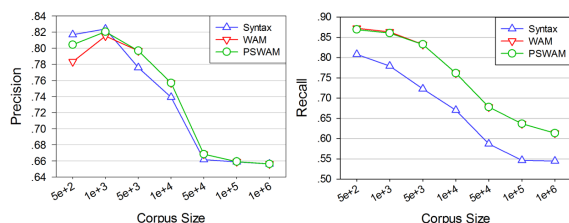


Figure 3: Experimental results on Restaurant

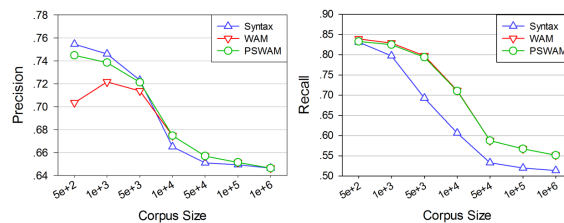


Figure 4: Experimental results on Hotel

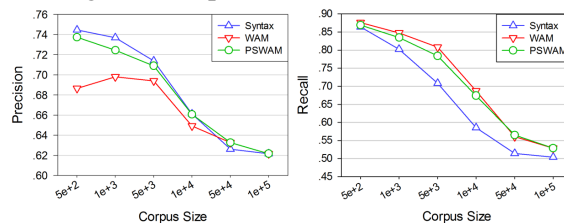


Figure 5: Experimental results on MP3

1) When the size of the corpus is small, **Syntax** has better precision than alignment based methods (**WAM** and **PSWAM**). We believe the reason is that the high-precision syntactic patterns employed in **Syntax** can effectively capture opinion relations in a small amount of texts. In contrast, the methods based on word alignment model may suffer from data sparseness for parameter estimation, so the precision is lower.

2) However, when the size of the corpus increases, the precision of **Syntax** decreases, even worse than alignment based methods. We believe it's because more noises were introduced from parsing errors with the increase of the size of the corpus, which will have more negative impacts on extraction results. In contrast, for estimating the parameters of alignment based methods, the data is more sufficient, so the precision is better compared with syntax based method.

3) We also observe that recall of **Syntax** is worse than other two methods. It's because the human expressions of opinions are diverse and the manual designed syntactic patterns are limited to capture all opinion relations in sentences, which may miss an amount of correct opinion targets.

4) It's interesting that the performance gap between these three methods is smaller with the increase of the size of the corpus (more than 50,000). We guess the reason is that when the data is sufficient enough, we can obtain sufficient statistics for each opinion target. In such situation, the graph-based ranking algorithm in the second component will be apt to be affected by the frequency information, so the final performance could not be sensitive to the performance of opinion relations iden-

tification in the first component. Thus, in this situation, we can get conclusion that there is no obviously difference on performance between syntax-based approach and alignment-based approach.

5) From the results on dataset with different languages and different domains, we can obtain the similar observations. It indicates that choosing either syntactic patterns or word alignment model for extracting opinion targets can take a few consideration on the language and domain of the corpus.

Thus, based on the above observations, we can draw the following conclusions: making chooses between different methods is only related to the size of the corpus. The method based on syntactic patterns is more suitable for small corpus ( $\#sentences < 5 \times 10^3$  shown in our experiments). And word alignment model is more suitable for medium corpus ( $5 \times 10^3 < \#sentences < 5 \times 10^4$ ). Moreover, when the size of the corpus is big enough, the performance of two kinds of methods tend to become the same ( $\#sentences \geq 10^5$  shown in our experiments).

#### 4.4 Is It Useful Combining Syntactic Patterns with Word Alignment Model

In this subsection, we try to see whether combining syntactic information with alignment model by using PSWAM is effective or not for opinion target extraction. From the results in Figure 3, 4 and 5, we can see that PSWAM has the similar recall compared with WAM in all datasets. PSWAM outperforms WAM on precision in all dataset. But the precision gap between PSWAM and WAM decreases when the size of the corpus increases. When the size is larger than  $5 \times 10^4$ , the performance of these two methods is almost the same. We guess the reason is that more noises from parsing errors will be introduced by syntactic patterns with the increase of the size of corpus, which have negative impacts on alignment performance. At the same time, as mentioned above, a great deal of reviews will bring sufficient statistics for estimating parameters in alignment model, so the roles of partial supervision from syntactic information will be covered by frequency information used in our graph based ranking algorithm.

**Compared with State-of-the-art Methods.** However, it's not say that this combination is not useful. From the results, we still see that PSWAM outperforms WAM in all datasets on

precision when size of corpus is smaller than  $5 \times 10^4$ . To further prove the effectiveness of our combination, we compare PSWAM with some state-of-the-art methods, including Hu (Hu and Liu, 2004a), which extracted frequent opinion target words based on association mining rules, DP (Qiu et al., 2011), which extracted opinion targets through syntactic patterns, and LIU (Liu et al., 2012), which fulfilled this task by using unsupervised WAM. The parameter settings in these baselines are the same as the settings in the original papers. Because of the space limitation, we only show the results on Restaurant and Hotel, as shown in Figure 6 and 7.

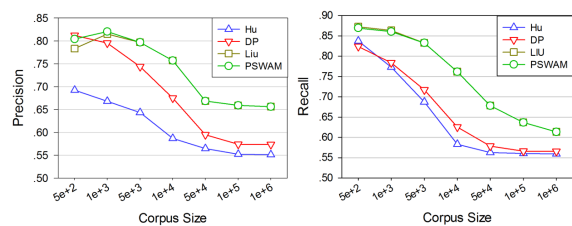


Figure 6: Compared with the State-of-the-art Methods on Restaurant

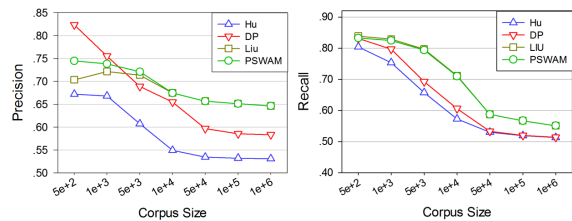


Figure 7: Compared with the State-of-the-art Methods on Hotel

From the experimental results, we can obtain the following observations. PSWAM outperforms other methods in most datasets. This indicates that our method based on PSWAM is effective for opinion target extraction. Especially compared PSWAM with LIU, both of which are based on word alignment model, we can see PSWAM identifies opinion relations by performing WAM under partial supervision, which can effectively improve the precision when dealing with small and medium corpus. However, these improvements are limited when the size of the corpus increases, which has the similar observations obtained above.

**The Impact of Syntactic Information on Word Alignment Model.** Although we have prove the effectiveness of PSWAM in the corpus with small and medium size, we are still curious about how the performance varies when we incor-



porate different amount of syntactic information into WAM. In this experiment, we rank the used syntactic patterns mentioned in Section 3.1.1 according to the quantities of the extracted alignment links by these patterns. Then, to capture opinion relations, we respectively use top N syntactic patterns according to frequency mentioned above to generate partial alignment links for PSWAM in section 3.1.3. We respectively define  $N=[1,7]$ . The larger is N, the more syntactic information is incorporated. Because of the space limitation, only the average performance of all dataset is shown in Figure 8.

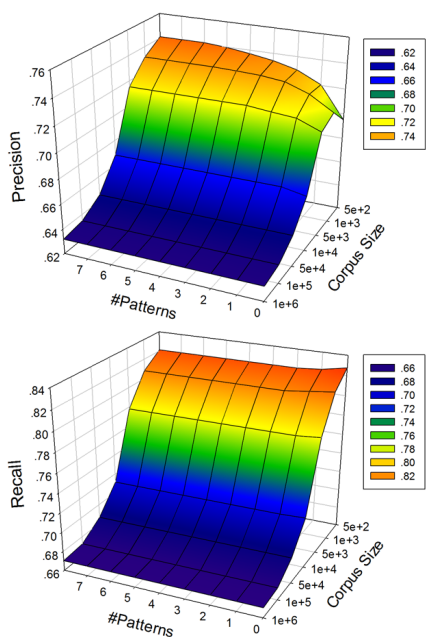


Figure 8: The Impacts of Different Syntactic Information on Word Alignment Model

In Figure 8, we can observe that the syntactic information mainly have effect on precision. When the size of the corpus is small, the opinion relations mined by high-precision syntactic patterns are usually correct, so incorporating more syntactic information can improve the precision of word alignment model more. However, when the size of the corpus increases, incorporating more syntactic information has little impact on precision.

## 5 Conclusions and Future Work

This paper discusses the performance variation of syntax based methods and alignment based methods on opinion target extraction task for the dataset with different sizes, different languages and different domains. Through experimental results, we can see that choosing which method is not related

with corpus domain and language, but strongly associated with the size of the corpus. We can conclude that syntax-based method is likely to be more effective when the size of the corpus is small, and alignment-based methods are more useful for the medium size corpus. We further verify that incorporating syntactic information into word alignment model by using PSWAM is effective when dealing with the corpora with small or medium size. When the size of the corpus is larger and larger, the performance gap between syntax based, WAM and PSWAM will decrease.

In future work, we will extract opinion targets based on not only opinion relations. Other semantic relations, such as the topical associations between opinion targets (or opinion words) should also be employed. We believe that considering multiple semantic associations will help to improve the performance. In this way, how to model heterogenous relations in a unified model for opinion targets extraction is worthy to be studied.

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 61070106, No. 61272332 and No. 61202329), the National High Technology Development 863 Program of China (No. 2012AA011102), the National Basic Research Program of China (No. 2012CB316300), Tsinghua National Laboratory for Information Science and Technology (TNList) Cross-discipline Foundation and the Opening Project of Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (ICDD201201).

## References

- Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. *Comput. Linguist.*, 19(2):263–311, June.
- Xiaowen Ding, Bing Liu, and Philip S. Yu. 2008. A holistic lexicon-based approach to opinion mining. In *Proceedings of the Conference on Web Search and Web Data Mining (WSDM)*.
- Qin Gao, Nguyen Bach, and Stephan Vogel. 2010. A semi-supervised word alignment algorithm with partial manual alignments. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 1–10, Uppsala, Sweden, July. Association for Computational Linguistics.

- Mingqin Hu and Bing Liu. 2004a. Mining opinion features in customer reviews. In *Proceedings of Conference on Artificial Intelligence (AAAI)*.
- Mingqing Hu and Bing Liu. 2004b. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '04*, pages 168–177, New York, NY, USA. ACM.
- Wei Jin and Hay Ho Huang. 2009. A novel lexicalized hmm-based learning framework for web opinion mining. In *Proceedings of International Conference on Machine Learning (ICML)*.
- Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, September.
- Fangtao Li, Chao Han, Minlie Huang, Xiaoyan Zhu, Yingju Xia, Shu Zhang, and Hao Yu. 2010. Structure-aware review mining and summarization. In Chu-Ren Huang and Dan Jurafsky, editors, *COLING*, pages 653–661. Tsinghua University Press.
- Bing Liu, Mingqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In Allan Ellis and Tatsuya Hagino, editors, *WWW*, pages 342–351. ACM.
- Kang Liu, Liheng Xu, and Jun Zhao. 2012. Opinion target extraction using word-based translation model. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1346–1356, Jeju Island, Korea, July. Association for Computational Linguistics.
- Kang Liu, Liheng Xu, Yang Liu, and Jun Zhao. 2013. Opinion target extraction using partially supervised word alignment model.
- Tengfei Ma and Xiaojun Wan. 2010. Opinion target extraction in chinese news comments. In Chu-Ren Huang and Dan Jurafsky, editors, *COLING (Posters)*, pages 782–790. Chinese Information Processing Society of China.
- Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, pages 339–346, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Che. 2009. Expanding domain sentiment lexicon through double propagation.
- Guang Qiu, Bing Liu 0001, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27.
- Bo Wang and Houfeng Wang. 2008. Bootstrapping both product features and opinion words from chinese customer reviews with cross-inducing.
- Hongning Wang, Yue Lu, and ChengXiang Zhai. 2011. Latent aspect rating analysis without aspect keyword supervision. In Chid Apt, Joydeep Ghosh, and Padhraic Smyth, editors, *KDD*, pages 618–626. ACM.
- Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. 2009. Phrase dependency parsing for opinion mining. In *EMNLP*, pages 1533–1541. ACL.
- Qi Zhang, Yuanbin Wu, Tao Li, Mitsunori Ogihara, Joseph Johnson, and Xuanjing Huang. 2009. Mining product reviews based on shallow dependency parsing. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, SIGIR '09*, pages 726–727, New York, NY, USA. ACM.
- Lei Zhang, Bing Liu, Suk Hwan Lim, and Eamonn O'Brien-Strain. 2010. Extracting and ranking product features in opinion documents. In Chu-Ren Huang and Dan Jurafsky, editors, *COLING (Posters)*, pages 1462–1470. Chinese Information Processing Society of China.
- Jingbo Zhu, Huizhen Wang, Benjamin K. Tsou, and Muhua Zhu. 2009. Multi-aspect opinion polling from textual reviews. In David Wai-Lok Cheung, Il-Yeol Song, Wesley W. Chu, Xiaohua Hu, and Jimmy J. Lin, editors, *CIKM*, pages 1799–1802. ACM.