Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Opinion Target Extraction Using Partially-Supervised Word Alignment Model*

Kang Liu, Liheng Xu, Yang Liu and Jun Zhao
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
{kliu, lhxu, liuyang09, jzhao}@nlpr.ia.ac.cn

Abstract

Mining opinion targets from online reviews is an important and challenging task in opinion mining. This paper proposes a novel approach to extract opinion targets by using partially-supervised word alignment model (PSWAM). At first, we apply PSWAM in a monolingual scenario to mine opinion relations in sentences and estimate the associations between words. Then, a graph-based algorithm is exploited to estimate the confidence of each candidate, and the candidates with higher confidence will be extracted as the opinion targets. Compared with existing syntax-based methods, PSWAM can effectively avoid parsing errors when dealing with informal sentences in online reviews. Compared with the methods using alignment model, PSWAM can capture opinion relations more precisely through partial supervision from partial alignment links. Moreover, when estimating candidate confidence, we make penalties on higher-degree vertices in our graph-based algorithm in order to decrease the probability of the random walk running into the unrelated regions in the graph. As a result, some errors can be avoided. The experimental results on three data sets with different sizes and languages show that our approach outperforms state-of-the-art methods.

1 Introduction

In recent years, mining opinions and analyzing sentiments in online reviews becomes useful and has attracted a lot of attentions from many researchers [Hu and Liu, 2004b; Li *et al.*, 2012; Zhang *et al.*, 2010; Liu *et al.*, 2012]. In opinion mining, one fundamental problem is to extract opinion targets, which are defined as the objects on which users have expressed their opinions, typically as nouns or noun phrases. This task is very important because customers are usually not satisfied with just the overall sentiment polarity of a product, but expect to find the fine-grained sentiments about an aspect or a product feature mentioned in reviews.

To fulfill this task, existing studies usually regarded opinion words as strong indicators [Popescu and Etzioni, 2005; Hu and Liu, 2004a; Bing *et al.*, 2005; Qiu *et al.*, 2011]. This strategy is based on the observation that opinion words are usually used to modify opinion targets, and there are opinion relations and associations between them. For example, "*wonderful*" and "*fantastic*" are usually used to modify "*design*" about cell-phone in reviews, so there are strong associations between them. If "*wonderful*" and "*fantastic*" had been known to be opinion words, "*design*" is likely to be an opinion target in this domain. Meanwhile the extracted opinion targets can be used to expand more opinion words. It is a mutual reinforcement procedure.

Therefore, mining opinion relations in sentences and estimating associations between opinion words and opinion targets are keys for opinion target extraction. To this end, several methods designed some heuristic patterns based on syntactic parsing [Popescu and Etzioni, 2005; Qiu *et al.*, 2009; 2011]. However, online reviews usually have informal writing styles including grammar mistakes, typos, improper punctuation etc., which make parsing be prone to generate mistakes. As a result, these syntax-based methods which heavily depended on the parsing performance would suffer from parsing errors and often don't work well. To resolve this problem, [Liu *et al.*, 2012] formulate identifying opinion relations between words as an alignment process. An opinion target can find its corresponding modifier through monolingual word alignment model (WAM) without using parsing, so that the noises from parsing errors can be effectively avoided. Experimental results have reported that their method have better performance than syntax-based methods, especially for large corpora.

Nevertheless, we notice that WAM used in Liu's method are trained in a completely unsupervised manner, which makes the alignment quality still unsatisfactory. Although we can improve the alignment performance by using the supervised framework [Moore, 2005], manually labeling full alignment for sentences is still time-consuming and impractical. However, in many situations, we can easily obtain a portion of links of the full alignment in a sentence. They can be used to constrain the alignment process, which is a partially-supervised alignment problem. We argue that it would benefit to improving the alignment performance. For example in Figure 1, "*colorful*" is incorrectly regarded as the modifier for "*phone*" if the WAM is performed in an whole unsuper-

*Kang Liu and Liheng Xu have equal contribution to this work vised manner ((a) in Figure 1). However, we can assert "*colorful*" should be aligned to "*screen*" ((b) in Figure 1). Then the errors in (a) can be corrected by using this partial links to supervise the statistical model, as shown in (c) in Figure 1.

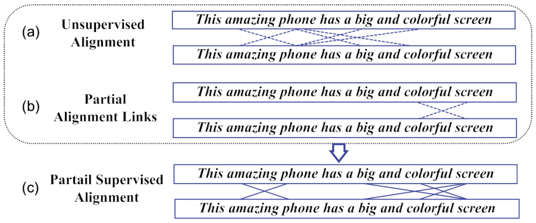

Figure 1: Mining Opinion Relations between Words using Partially-Supervised Alignment Model

Thus, in this paper, we propose a novel approach to extract opinion targets by using partially-supervised word alignment model (PSWAM). We first use some high-precise-low-recall syntactic patterns to capture partial opinion relations (partial alignment links) in sentences. Although existing syntactic parsing algorithms cannot obtain the precise whole syntactic tree of the informal sentences, we believe some short or direct dependency relations between words can be still obtained precisely. Then these extracted partial alignment links would be regarded as ground truths. And an constrained EM algorithm based on hill-climbing is performed to determine all alignments in sentences, where the model will be consistent with these links as far as possible. In this way, more correct opinion relations can be mined. Our model can not only inherit the advantages of word alignment model: considering multiple factors (word co-occurrence frequencies, word positions etc.) in the global process, effectively avoiding noises from syntactic parsing errors when dealing with informal texts, but also can improve the mining performance by using partial supervision. Thus, it's reasonable to expect that PSWAM is likely to yield better performance than traditional methods.

Then we extract opinion targets in a graph-based framework based on the mined associations. All nouns (noun phrases) are regarded as opinion target candidates. A bipartite graph is constructed to model the opinion relations between words. We assume that two candidates are modified by similar opinion words, they are likely to belong to the similar category. If we have known one of them is an opinion target, the other one has high probability to be an opinion target. Thus, the opinion target confidence can propagate among vertices. A random walk algorithm can be applied to estimate the confidence of each candidate, and the candidates with higher confidence will be extracted as the opinion targets. However, in traditional random work algorithm, we observe that the higher-degree vertices are prone to collect more information from other vertices and put more impacts on other vertices. These words usually are *general words* and may introduce noises. For example, the opinion word "*good*", may be used to modify multiple objects like "*good design*", "*good feeling*" and "*good things*". The degree of "*good*" will be high in the graph. If we have know that the "*design*" has higher confidence to be an opinion target, it's confidence will be propagated to "*feeling*" and "*thing*" through "*good*". As a result, "*feeling*" and "*thing*" will probably to be given higher confidence as opinion targets. It's unreasonable. To resolve this problem, we make penalty on the higher-degree vertices to weaken the impacts of them and decrease the probability of the random walk running into the unrelated regions in the graph. In this way, errors can be effectively avoided.

The experimental results on three datasets with different sizes and languages show that our approach can achieve performance improvement over the traditional methods.

2 Related Work

Previous studies focused on opinion target extraction, such as [Hu and Liu, 2004b; Ding *et al.*, 2008; Li *et al.*, 2010; Popescu and Etzioni, 2005; Wu *et al.*, 2009], 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 task [Jin and Huang, 2009; Li *et al.*, 2010; Ma and Wan, 2010; Wu *et al.*, 2009]. The main limitation of these methods is that labeling training data for each domain is time consuming and impracticable. In unsupervised methods, similar to ours, most approaches regarded opinion words as the important indicators for opinion targets. [Hu and Liu, 2004a] exploited an association rule mining algorithm and frequency information to extract frequent explicit product features in a bootstrapping process. [Popescu and Etzioni, 2005] designed some syntactic patterns to extract opinion targets. [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. The main limitation of Qiu's method is that the patterns based on dependency parsing tree may introduce many noises for the large corpora. [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] is similar to our method, they use a completely unsupervised WTM 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.

3 Mining Associations between Opinion Targets and Opinion Words using Partially Supervised Word Alignment Model

As mentioned in the first section, we first need to identify potential opinion relations in sentences and estimate associations between opinion targets and opinion words. Similar to Liu's method [Liu *et al.*, 2012], we formulate this task as a monolingual word alignment process. We assume opinion targets to be nouns or noun phrases, and opinion words may be adjectives or verbs. Every sentence is replicated to generate parallel corpus, and the word alignment algorithm is applied to the monolingual scenario to align a noun (noun phase) with its modifiers for mining opinion relations in sentences. Thus, given a sentence with n

words $S = \{w_1, w_2, ..., w_n\}$, the word alignment $A = \{(i, a_i)|i \in [1, n], a_i \in [1, n]\}$ can be obtained as $A^* = \underset{A}{argmax}\, P(A|S)$, where (i, a_i) means that a noun (noun phrase) at position i is aligned with its modifier at position a_i. For alignment model, we select IBM-3 [Brown *et al.*, 1993] which has been proved to have better performance than other models [Liu *et al.*, 2012]. So 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 corpora. If a word modifies a noun (noun phrase) frequently, they will has higher value of $t(w_j|w_{a_j})$. $d(j|a_j, n)$ 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 "one-to-many" relation which means that a word can modify (be modified by) several words. ϕ_i denotes the number of words that are aligned with w_i.

3.1 Partially-Supervised Word Alignment Model

To improve alignment performance, we make partial supervision on the statistic model and incorporate partial alignment links into the alignment process. Here, the partial alignment links are regarded as the ground truth and the constrains for the alignment model. Given the partial alignment links $\hat{A} = \{(i, a_i)|i \in [1, n], a_i \in [1, n]\}$, the optimal word alignment A^* mentioned above is rewritten as follows.

$$A^* = \underset{A}{argmax}\, P(A|S, \hat{A}) \quad (2)$$

Because of the lack of the labeled full alignment data, we employ constrained EM algorithm to train the model parameters. In the E-step, we collect the statistics from all possible alignments, and ignore the evidences which are inconsistent with the ground truth to guarantee that no statistic from inconsistent alignments is collected for model parameter estimation. Thus we have

$$P(w_i|w_{a_i}) = \begin{cases} \lambda, & A \ \textit{is inconsistent with}\ \hat{A} \\ P(w_i|w_{a_i}) + \lambda, & \textit{otherwise} \end{cases} \quad (3)$$

where λ is the smoothing factor which means that we make soft constraints on the alignment model, and some incorrect partial alignment links generated through high-precision patterns (Section 3.2) may be revised possibly. Then, in M-step, we make count collections and normalize to produce the model parameters for the next iteration.

In practice, we can easily perform constrained EM algorithm as Equation (3) for simpler models such as in IBM-1, IBM-2, by ruling out the inconsistent evidences. However, for IBM-3 model, because of the consideration of fertility of the words, enumerating all possible alignments is NP-complete [Gao *et al.*, 2010]. Thus GIZA++ [Och and Ney, 2003] provides a greedy hill-climbing algorithm to resolve this problem. To improve the speed of hill climbing, the search space is constrained on the "neighbor alignments" defined as the alignments that can be generated from current alignment by one of the operators: 1) Move operator $m_{i,j}$, which changes $a_j = i$; 2) Swap operator s_{j_1,j_2}, which exchanges a_{j_1} and a_{j_2}. Specially, two matrix respectively called Moving Matrix M and Swapping Matrix S are introduced which record all possible move and swap costs between two different alignments.

To incorporate the partial alignment links into IBM-3 model as constrains, we adopt the algorithm of [Gao *et al.*, 2010] which includes three main steps: 1) We first train a simple alignment model (IBM-1, IBM-2, HMM etc.) for generating initial alignments for IBM-3 model; 2) Then, we update the initial alignment model to make it consistent to the partial alignment links; 3) We finally optimize the alignment under the constraints iteratively by employing operation matrix M and S, where we set the corresponding cost value of the invalid operation in M and S to be negative, so that the invalid operators will never be picked.

Moreover, to capture opinion relations in sentences, we only focus on alignment between opinion target candidats (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, in our model, we introduce other constrains: 1) Nouns/noun phrases (adjectives/verbs) must be aligned with adjectives/verbs (nouns/noun phrases) or null words. Aligning to null words means that this word has no any modifier or modifies nothing; 2) Other unrelated words can only align with themselves.

From the alignment results, we can obtain a set of word pairs composed of a noun (noun phrase) and the corresponding modified words. Then the alignment probabilities between potential opinion target w_t and potential opinion word w_o can be estimated using $P(w_t|w_o) = \frac{Count(w_t, w_o)}{Count(w_o)}$, where $P(w_t|w_o)$ means the alignment probabilities between two words. Similarly, we can obtain alignment probability $P(w_o|w_t)$ if we change the alignment direction in the alignment process. Then, the association between an opinion target candidate and its modifier is estimated as follows.

$$Association(w_t, w_o) = (\alpha * P(w_t|w_o) + (1-\alpha) P(w_o|w_t))^{-1} \quad (4)$$

where α is the harmonic factor to combine these two alignment probabilities, and we set $\alpha = 0.5$.

3.2 Obtaining Partial Alignment Links

The other important problem is how to obtain partial alignment links for our PSWAM. We resort to manually labeling. But this strategy is time-consuming and isn't suitable for multiple domains. Therefore, as mentioned in the first section, we make use of some high-precision-low-recall syntactic patterns to capture the dependency relations between words for generating initial partial alignment links. Then these initial links will be feeded into the alignment model.

To guarantee the used syntactic patterns to be high-precision, we constrain that the syntactic patterns only capture direct dependency relations which is defined in [Qiu *et al.*, 2011]. A direct dependency indicates that one word depends on the other word without any additional words in their dependency path or these two words both depend on a third word directly. As shown in the left part ((a) and (b)) of Figure 2, A directly depends on B in (a), and A and B both directly depend on C in (b). [Qiu *et al.*, 2011] also defined some indirect dependency relations. We don't use them because introducing indirect dependency relations may decrease the precision. Specifically, we employ Minipar[1] as the sentence parser. The right part in Figure 2 shows the used syntactic patterns corresponding to two direct dependency relation types. A is potential opinion words, where $\{O\} = \{adjective, verb\}$. B is potential opinion target words, where $\{T\} = \{noun, nounphrase\}$. C is any words. Moreover, to obtain precision alignment links, we constrain the dependency relation labels in our patterns to be $R1 = \{mod, pnmod\}$ and $R2 = \{subj\}$, where mod, $pnmod$ and $subj$ is the dependency relation labels output by Minipar.

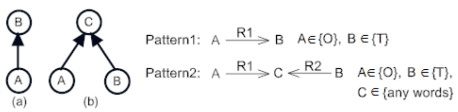

Figure 2: Examples of Used Syntactic Patterns

4 Estimating Candidate Confidence

After mining the associations between words, we compute the confidence of each opinion target candidate, and the candidates with higher confidence will be extracted as the opinion targets. Here, opinion words are regarded as the important indicators. We assume that two candidates are likely to belong to the similar category, if they are modified by similar opinion words. If we have known the one of them to be an opinion target, the other one has high probability to be an opinion target. Thus, we can forward the opinion target confidences among different target candidates through opinion words, which indicates that existing graph-based algorithms are applicable.

To model the mined associations between words, we construct a bipartite graph, 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 3, 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 them, which are obtained using Eq. (4).

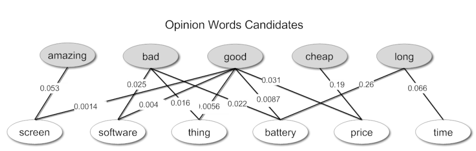

Figure 3: Modeling Opinion Relations between Words in a Bipartite Graph

[1] http://webdocs.cs.ualberta.ca/lindek/minipar.htm

To estimate opinion target confidence, we employ a random walk based algorithm named as Adsorption [Baluja *et al.*, 2008; Talukdar *et al.*, 2008] on our graph. In Adsorption, each vertex has two roles, forwarding confidences and collecting confidences. The algorithm iteratively computes the weighted average of opinion target confidences from neighboring vertices, then uses the random walk to estimate the confidences of other vertices. However, as mentioned in the first section, in the standard random walk process, the higher-degree vertices will usually have more impacts on other vertexes, which may introduce noises. Therefore, we make modification on Adsorption and put penalty on these higher-degree vertices in the random walk process to decrease the probability of generating errors. Specifically, when the random walk reaches a target vertex v, we believe that there are three choices: (a) continue the random walk to the neighbors of v; (b) abandon the random walk; (c) stop the walk and emit a confidence according to the prior knowledge. We assume the probabilities of these three events are $P_{con}(v)$, $P_{abnd}(v)$ and $P_{inj}(v)$ respectively. Thus we have

$$L_v^{i+1} = P_{con}(v) \times D_v^i + P_{inj}(v) \times I_v + P_{abnd}(v) \times L_\phi \quad (5)$$

where L_v^{i+1} represents the opinion target confidence estimation of v in the $i+1^{th}$ iteration. And

$$D_v^i = \sum_u \frac{W(v,u)}{\sum_u W(v,u)} L_u^i$$

which is the weighted average confidence for vertex v in the i^{th} iteration entirely based on its neighbors' confidence. And I_v is defined as the prior confidence of each candidate for opinion target. Similar to [Liu *et al.*, 2012], we set $I_v = \frac{tf(v)idf(v)}{\sum_v tf(v)idf(v)}$, where $tf(v)$ is the term frequency of v in the corpus, and $df(v)$ is computed by using the Google n-gram corpus[2]. L_ϕ represents lack of information about the opinion target confidence of the vertex v. $P_{abnd}(v)$ is used to mitigates the effect of the transition into unrelated region in the graph when reaching higher-degree vertex [Talukdar *et al.*, 2008]. Moreover, we adopt the heuristics from [Talukdar *et al.*, 2008] to set the random walk probabilities. Let $c_v = \frac{log(\beta)}{log(\beta+exp(H(v)))}$ and $\beta = 2$. $H(v) = -\sum_u p_{uv} \times log(p_{uv})$ denotes the entropy of vertex v, where $p_{uv} = \frac{W(u,v)}{\sum_u W(u,v)}$. If v has higher degree, $H(v)$ will be higher and c_v will be lower. In this way, the contribution of higher-degree vertices will be restricted. And $j_v = (1 - c_v) \times \sqrt{H(v)}$. $z_v = max(c_v + j_v, 1)$. Therefore, we can set

$$P_{inj}(v) = \frac{j_v}{z_v},\ P_{con}(v) = \frac{c_v}{z_v}$$

$$P_{abnd}(v) = 1 - P_{con}(v) - P_{inj}(v)$$

The algorithm is ran until convergence which is achieved when the confidence on each node ceases to change in a tolerance value. Finally, the candidates with higher confidence will be extracted as opinion targets.

[2] http://books.google.com/ngrams/datasets

Methods	Camera			Car			Laptop			Phone		
	P	R	F	P	R	F	P	R	F	P	R	F
Hu	0.63	0.65	0.64	0.62	0.58	0.60	0.51	0.67	0.58	0.69	0.60	0.64
DP	0.71	0.70	0.70	0.72	0.65	0.68	0.58	0.69	0.63	0.78	0.66	0.72
Zhang	0.71	0.78	0.74	0.69	0.68	0.68	0.57	0.80	0.67	0.80	0.71	0.75
Liu	0.75	0.81	0.78	0.71	**0.71**	0.71	0.61	**0.85**	0.71	0.83	0.74	0.78
Ours	**0.77**	**0.82**	**0.79**	**0.74**	**0.71**	**0.72**	**0.66**	**0.85**	**0.74**	**0.85**	**0.75**	**0.80**

Table 1: Experimental Results on *COAE 2008*

Methods	Hotel			MP3			Restaurant		
	P	R	F	P	R	F	P	R	F
Hu	0.60	0.65	0.62	0.61	0.68	0.64	0.64	0.69	0.66
DP	0.67	0.69	0.68	0.69	0.70	0.69	0.74	0.72	0.73
Zhang	0.67	0.76	0.71	0.67	0.77	0.72	0.75	0.79	0.77
Liu	0.71	0.80	0.75	0.70	0.82	0.76	0.80	0.84	0.82
Ours	**0.76**	**0.83**	**0.79**	**0.74**	**0.84**	**0.79**	**0.85**	**0.85**	**0.85**

Table 2: Experimental Results on *Large*

5 Experiments

5.1 Datasets and Evaluation Metrics

We select three datasets which were already used in [Liu *et al.*, 2012] to evaluate our approach. The first dataset is *COAE2008 dataset2*[3], which contains Chinese reviews about four kinds of products including camera, car, laptop and phone. The second dataset is *Large*, which includes three corpora with different languages from three domains including hotel, mp3 and restaurant. The detailed statistical information of these two datasets can be found in [Liu *et al.*, 2012]. The third dataset is *Customer Review Datasets* including English reviews of five products, which was also used in [Hu and Liu, 2004a; Qiu *et al.*, 2011]. The detailed information can be found in [Hu and Liu, 2004a].

In the experiments, reviews are first segmented into sentences according to punctuation. Then sentences are tokenized, part-of-speech tagged by using Stanford NLP tool[4], and parsed by using Minipar toolkit. The method in [Zhu *et al.*, 2009] is used to identify noun phrases. We select precision, recall and F-measure as the evaluation metrics.

5.2 Our Methods vs. State-of-the-art Methods

For comparison, we select the following the state-of-the-art methods as baselines.

- **Hu** is the method described in [Hu and Liu, 2004a], which extracted frequent opinion target words based on association mining rules.
- **DP** is the method proposed by [Qiu *et al.*, 2011], which used syntax-based patterns to capture opinion relations in sentences, and used Double Propagation algorithm to extract the opinion targets.
- **Zhang** is the method proposed by [Zhang *et al.*, 2010], which is an extension of DP. They extracted opinion targets candidates using syntactic patterns and other specific patterns. Then HITS [Kleinberg, 1999] algorithm combined with candidate frequency is employed to rank the results for opinion target extraction.
- **Liu** is the method described in [Liu *et al.*, 2012], which used unsupervised WAM to mining the associations between words. Then a standard random walk based algorithm is exploited to estimate the candidate confidences for extracting opinion targets.

The parameter settings in these baselines are the same as the settings in original papers. The overall performance results are respectively shown in Table 1, 2 and 3, where "P" denotes precision, "R" denotes recall and "F" denotes F-measure. **Ours** denotes our method, in which we use partially-supervised word alignment model to identify opinion relations in sentences, and a modified graph-based algorithm to extract opinion targets. We set $\phi_{max} = 2$ in Eq.(1). From the experimental results, we can obtain the following observations.

1) **Ours** outperforms other baselines in most datasets. This indicates that our method based on PSWAM is effective for opinion target extraction.

2) The methods based on word alignment model (**Ours** and **Liu**) achieves performance improvement over other baselines, especially outperforming syntax-based methods (**DP** and **Zhang**). It indicates that the methods based on the word alignment model, which formulate identifying opinion relations as an alignment process, can effective avoid parsing errors for informal texts on the Web without using parsing. So it can mine more precise opinion relations from sentences. The similar observation is also obtained by [Liu *et al.*, 2012].

3) Moreover, **Ours** outperforms **Liu**. We believe there are two reasons. First, **Ours** identifies opinion relations by performing WAM under partial supervision. We employ high-precision syntactic patterns to obtain partial alignment links and regard them as constrains for supervising alignment model. This strategy is effective for improving the precision of opinion relations identification. Second, to estimate the confidence of each candidate in the graph, we make penalty on the higher-degree vertices to decrease the probability of the random walk running into the unrelated regions. In this way, some errors can be effectively avoided, so that the precision can be improved further.

4) In Table 3, **Ours** makes comparable results with baselines in *Customer Review Datasets*, although there is a little loss in precision in some domains. We believe the reason is that the size of *Customer Review Datasets* is too small. As

[3]http://ir-china.org.cn/coae2008.html

[4]http://nlp.stanford.edu/software/tagger.shtml

Methods	D1			D2			D3			D4			D5		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hu	0.75	0.82	0.78	0.71	0.79	0.75	0.72	0.76	0.74	0.69	0.82	0.75	0.74	0.80	0.77
DP	**0.87**	0.81	0.84	**0.90**	0.81	0.85	**0.90**	0.86	0.88	**0.81**	0.84	0.82	**0.92**	0.86	**0.89**
Zhang	0.83	0.84	0.83	0.86	**0.85**	0.85	0.86	0.88	0.87	0.80	**0.85**	0.82	0.86	0.86	0.86
Liu	0.84	**0.85**	0.84	0.87	**0.85**	**0.86**	0.88	0.89	0.88	**0.81**	**0.85**	**0.83**	0.89	**0.87**	0.88
Ours	0.86	0.84	**0.85**	0.88	0.83	0.85	0.89	**0.90**	**0.89**	**0.81**	0.83	0.82	0.91	**0.87**	**0.89**

Table 3: Experimental Results on *Customer Review Dataset*

a result, word alignment model may suffer from data sparseness for association estimation. Nevertheless, the average recall is improved. Moreover, compared the results in Table 1 with the results in Table 2, we can observe that **Ours** may obtains larger improvements with the increase of the data size, which indicates the proposed method is more appropriate for larger corpora.

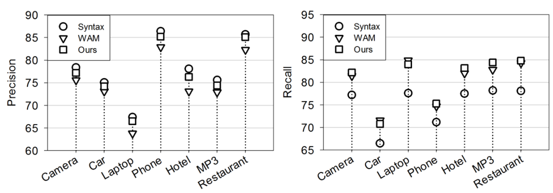

Figure 4: Experimental Comparison among different opinion relation identification methods

5.3 Effect of Partially-Supervised Word Alignment Model

In this subsection, we aim to prove the effectiveness of our PSWAM for capturing opinion relations in sentences. For comparison, we design two baselines, named as **Syntax** and **WAM**. They respectively use syntactic patterns used in Section 3.2 and unsupervised WAM of [Liu *et al.*, 2012] to identify opinion relations in sentences. Then the same method Eq.4) is used to estimate associations between opinion targets and opinion words. At last the same graph-based algorithm proposed in Section 4 is used to extract opinion targets. Due to the limitation of the space, the experimental results only on *COAE2008 dataset2* and *Large* are shown in Figure 4.

In Figure 4, we observe that **Syntax** has better precision but worse recall than other two methods based on word alignment model (**Ours** and **WAM**). It's because that **Syntax** exploiting high-precision-low-recall syntactic patterns can only capture the part of opinion relations in sentences, which may lose many potential opinion targets. On the other side, (**Ours** and **WAM**) can capture more opinion relations by using word alignment model, so that they have better recall. Moreover, **Ours** outperforms **WAM**. It's because that alignment performance is improved under the partially supervised framework, which uses high-precision partial alignment links as the constrains. Thus, it can not only make significant improvements on recall, but obtain competitive performance on precision compared with **Syntax**, which proves the effectiveness of our PSWAM.

5.4 Effect of Our Graph-based Method

In this experiment, we aim to prove the effectiveness of our graph-based method for opinion target extraction. We design two baselines, named as **PSWAM_HITS** and **PSWAM_RW**. Both of them also use PSWAM to mine associations between opinion targets and opinion words. Then, **PSWAM_HITS** uses the HITS algorithm used in [Zhang *et al.*, 2010] to extract opinion targets. **PSWAM_RW** uses the the random walk algorithm used in [Liu *et al.*, 2012] to extract opinion targets. Figure 5 gives the experimental results on *COAE2008 dataset2* and *Large*. In Figure 5, we can observe that our graph-based algorithm outperforms other approaches. We believe the main reason is that we make penalty on the higher-degree vertices in the graph according to the vertex entropy to decrease the probability of the random walk running into the unrelated regions. Some errors introduced by *general words* can be filtered, so that the performance can be improved.

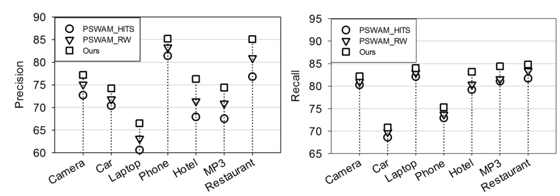

Figure 5: Experimental Comparison among different ranking methods

6 Conclusions and Future Work

This paper proposes a novel approach to extract opinion targets by using PSWAM. Compared with previous syntax-based methods, PSWAM can effectively avoid parsing errors when dealing with informal sentences in online reviews. And, compared with the methods using unsupervised alignment model, PSWAM can capture opinion relations more precisely under the constrains of partial alignment links. The experimental results proves the effectiveness of our method. In future work, we wouldn't extract opinion targets only using opinion relations. Other semantic relations, such as topical assocations, can be employed, which is expected to benefit for performance improvement.

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

[Baluja *et al.*, 2008] Shumeet Baluja, Rohan Seth, D. Sivakumar, Yushi Jing, Jay Yagnik, Shankar Kumar, Deepak Ravichandran, and Mohamed Aly. Video suggestion and discovery for youtube: taking random walks through the view graph. In Jinpeng Huai, Robin Chen, Hsiao-Wuen Hon, Yunhao Liu, Wei-Ying Ma, Andrew Tomkins, and Xiaodong Zhang 0001, editors, *WWW*, pages 895–904. ACM, 2008.

[Bing *et al.*, 2005] Liu Bing, Minqing Hu, and Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the web. In Allan Ellis and Tatsuya Hagino, editors, *WWW*, pages 342–351. ACM, 2005.

[Brown *et al.*, 1993] Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. The mathematics of statistical machine translation: parameter estimation. *Comput. Linguist.*, 19(2):263–311, June 1993.

[Ding *et al.*, 2008] Xiaowen Ding, Bing Liu, and Philip S. Yu. A holistic lexicon-based approach to opinion mining. In *Proceedings of the Conference on Web Search and Web Data Mining (WSDM)*, 2008.

[Gao *et al.*, 2010] Qin Gao, Nguyen Bach, and Stephan Vogel. 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 2010. Association for Computational Linguistics.

[Hu and Liu, 2004a] Mingqin Hu and Bing Liu. Mining opinion features in customer reviews. In *Proceedings of Conference on Artificial Intelligence (AAAI)*, 2004.

[Hu and Liu, 2004b] Minqing Hu and Bing Liu. 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, 2004. ACM.

[Jin and Huang, 2009] Wei Jin and Hay Ho Huang. A novel lexicalized hmm-based learning framework for web opinion mining. In *Proceedings of International Conference on Machine Learning (ICML)*, 2009.

[Kleinberg, 1999] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, September 1999.

[Li *et al.*, 2010] Fangtao Li, Chao Han, Minlie Huang, Xiaoyan Zhu, Yingju Xia, Shu Zhang, and Hao Yu. Structure-aware review mining and summarization. In Chu-Ren Huang and Dan Jurafsky, editors, *COLING*, pages 653–661. Tsinghua University Press, 2010.

[Li *et al.*, 2012] Fangtao Li, Sinno Jialin Pan, Ou Jin, Qiang Yang, and Xiaoyan Zhu. Cross-domain co-extraction of sentiment and topic lexicons. In *ACL (1)*, pages 410–419. The Association for Computer Linguistics, 2012.

[Liu *et al.*, 2012] Kang Liu, Liheng Xu, and Jun Zhao. 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 2012. Association for Computational Linguistics.

[Ma and Wan, 2010] Tengfei Ma and Xiaojun Wan. 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, 2010.

[Moore, 2005] Robert C. Moore. A discriminative framework for bilingual word alignment. In *HLT/EMNLP*. The Association for Computational Linguistics, 2005.

[Och and Ney, 2003] Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, 2003.

[Popescu and Etzioni, 2005] Ana-Maria Popescu and Oren Etzioni. 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, 2005. Association for Computational Linguistics.

[Qiu *et al.*, 2009] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Che. Expanding domain sentiment lexicon through double propagation. 2009.

[Qiu *et al.*, 2011] Guang Qiu, Bing Liu 0001, Jiajun Bu, and Chun Chen. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27, 2011.

[Talukdar *et al.*, 2008] Partha Pratim Talukdar, Joseph Reisinger, Marius Pasca, Deepak Ravichandran, Rahul Bhagat, and Fernando Pereira. Weakly-supervised acquisition of labeled class instances using graph random walks. In *EMNLP*, pages 582–590. ACL, 2008.

[Wu *et al.*, 2009] Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. Phrase dependency parsing for opinion mining. In *EMNLP*, pages 1533–1541. ACL, 2009.

[Zhang *et al.*, 2010] Lei Zhang, Bing Liu, Suk Hwan Lim, and Eamonn O'Brien-Strain. 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, 2010.

[Zhu *et al.*, 2009] Jingbo Zhu, Huizhen Wang, Benjamin K. Tsou, and Muhua Zhu. 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, 2009.