Extracting Opinion Targets and Opinion Words from Online Reviews with Graph Co-ranking

Kang Liu, Liheng Xu and Jun Zhao
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
{kliu, lhxu, jzhao}@nlpr.ia.ac.cn

Abstract

Extracting opinion targets and opinion words from online reviews are two fundamental tasks in opinion mining. This paper proposes a novel approach to collectively extract them with graph co-ranking. First, compared to previous methods which solely employed opinion relations among words, our method constructs a heterogeneous graph to model two types of relations, including semantic relations and opinion relations. Next, a co-ranking algorithm is proposed to estimate the confidence of each candidate, and the candidates with higher confidence will be extracted as opinion targets/words. In this way, different relations make cooperative effects on candidates' confidence estimation. Moreover, word preference is captured and incorporated into our co-ranking algorithm. In this way, our co-ranking is personalized and each candidate's confidence is only determined by its preferred collocations. It helps to improve the extraction precision. The experimental results on three data sets with different sizes and languages show that our approach achieves better performance than state-of-the-art methods.

1 Introduction

In opinion mining, extracting opinion targets and opinion words are two fundamental subtasks. Opinion targets are objects about which users' opinions are expressed, and opinion words are words which indicate opinions' polarities. Extracting them can provide essential information for obtaining fine-grained analysis on customers' opinions. Thus, it has attracted a lot of attentions (Hu and Liu, 2004b; Liu et al., 2012; Moghaddam and Ester, 2011; Mukherjee and Liu, 2012).

To this end, previous work usually employed a collective extraction strategy (Qiu et al., 2009; Hu and Liu, 2004b; Liu et al., 2013b). Their intuition is: opinion words usually co-occur with opinion targets in sentences, and there are strong modification relationship between them (called *opinion relation* in (Liu et al., 2012)). If a word is an opinion word, other words with which that word having opinion relations will have highly probability to be opinion targets, and vice versa. In this way, extraction is alternatively performed and mutual reinforced between opinion targets and opinion words. Although this strategy has been widely employed by previous approaches, it still has several limitations.

1) **Only considering opinion relations is insufficient**. Previous methods mainly focused on employing opinion relations among words for opinion target/word co-extraction. They have investigated a series of techniques to enhance opinion relations identification performance, such as nearest neighbor rules (Liu et al., 2005), syntactic patterns (Zhang et al., 2010; Popescu and Etzioni, 2005), word alignment models (Liu et al., 2012; Liu et al., 2013b; Liu et al., 2013a), etc. However, we are curious that whether merely employing opinion relations among words is enough for opinion target/word extraction? We note that there are additional types of relations among words. For example, "*LCD*" and "*LED*" both denote the same aspect "*screen*" in TV set domain, and they are topical related. We call such relations between homogeneous words as semantic relations. If we have known "*LCD*" to be an opinion target, "*LED*" is naturally to be an opinion target. Intuitively, besides opinion relations, semantic relations may provide additional rich clues for indicating opinion targets/words. Which kind of relations is more effective for opinion targets/words extraction? Is it beneficial to consider these two types of relations together for the extraction? To our best knowl-

Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 314–324,
Baltimore, Maryland, USA, June 23-25 2014. ©2014 Association for Computational Linguistics

edge, these problems have seldom been studied before (see Section 2).

2) **Ignoring word preference**. When employing opinion relations to perform mutual reinforcing extraction between opinion targets and opinion words, previous methods depended on opinion associations among words, but seldom considered *word preference*. Word preference denotes a word's preferred collocations. Intuitively, the confidence of a candidate being an opinion target (opinion word) should mostly be determined by its word preferences rather than all words having opinion relations with it. For example

"*This camera's price is expensive for me.*"

"*It's price is good.*"

"*Canon 40D has a good price.*"

In these three sentences, "*price*" is modified by "*good*" more times than "*expensive*". In traditional extraction strategy, opinion associations are usually computed based on the co-occurrence frequency. Thus, "*good*" has more strong opinion association with "*price*" than "*expensive*", and it would have more contributions on determining "*price*" to be an opinion target or not. It's unreasonable. "*Expensive*" actually has more relatedness with "*price*" than "*good*", and "*expensive*" is likely to be a word preference for "*price*". The confidence of "*price*" being an opinion target should be influenced by "*expensive*" in greater extent than "*good*". In this way, we argue that the extraction will be more precise.

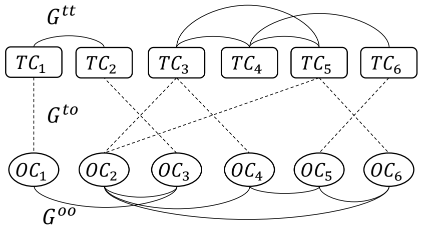


Figure 1: Heterogeneous Graph: OC means opinion word candidates. TC means opinion target candidates. Solid curves and dotted lines respectively mean semantic relations and opinion relations between two candidates.

Thus, to resolve these two problems, we present a novel approach with graph co-ranking. The collective extraction of opinion targets/words is performed in a co-ranking process. First, we operate over a heterogeneous graph to model semantic relations and opinion relations into a unified model. Specifically, our heterogeneous graph is composed of three subgraphs which model different relation types and candidates, as shown in Figure 1. The first subgraph G^{tt} represents semantic relations among opinion target candidates, and the second subgraph G^{oo} models semantic relations among opinion word candidates. The third part is a bipartite subgraph G^{to}, which models opinion relations among different candidate types and connects the above two subgraphs together. Then we perform a random walk algorithm on G^{tt}, G^{oo} and G^{to} separately, to estimate all candidates' confidence, and the entries with higher confidence than a threshold are correspondingly extracted as opinion targets/words. The results could reflect which type of relation is more useful for the extraction.

Second, a co-ranking algorithm, which incorporates three separate random walks on G^{tt}, G^{oo} and G^{to} into a unified process, is proposed to perform candidate confidence estimation. Different relations may cooperatively affect candidate confidence estimation and generate more global ranking results. Moreover, we discover each candidate's preferences through topics. Such word preference will be different for different candidates. We add word preference information into our algorithm and make our co-ranking algorithm be personalized. A candidate's confidence would mainly absorb the contributions from its word preferences rather than its all neighbors with opinion relations, which may be beneficial for improving extraction precision.

We perform experiments on real-world datasets from different languages and different domains. Results show that our approach effectively improves extraction performance compared to the state-of-the-art approaches.

2 Related Work

There are many significant research efforts on opinion targets/words extraction (sentence level and corpus level). In sentence level extraction, previous methods (Wu et al., 2009; Ma and Wan, 2010; Li et al., 2010; Yang and Cardie, 2013) mainly aimed to identify all opinion target/word mentions in sentences. They regarded it as a sequence labeling task, where several classical models were used, such as CRFs (Li et al., 2010) and SVM (Wu et al., 2009).

This paper belongs to corpus level extraction, and aims to generate a sentiment lexicon and a target list rather than to identify mentions in sentences. Most of previous corpus-level methods adopted a co-extraction framework, where opinion targets and opinion words reinforce each other according to their opinion relations. Thus, how to improve opinion relations identification performance was their main focus. (Hu and Liu, 2004a) exploited nearest neighbor rules to mine opinion relations among words. (Popescu and Etzioni, 2005) and (Qiu et al., 2011) designed syntactic patterns to perform this task. (Zhang et al., 2010) promoted Qiu's method. They adopted some special designed patterns to increase recall. (Liu et al., 2012; Liu et al., 2013a; Liu et al., 2013b) employed word alignment model to capture opinion relations rather than syntactic parsing. The experimental results showed that these alignment-based methods are more effective than syntax-based approaches for online informal texts. However, all aforementioned methods only employed opinion relations for the extraction, but ignore considering semantic relations among homogeneous candidates. Moreover, they all ignored word preference in the extraction process.

In terms of considering semantic relations among words, our method is related with several approaches based on topic model (Zhao et al., 2010; Moghaddam and Ester, 2011; Moghaddam and Ester, 2012a; Moghaddam and Ester, 2012b; Mukherjee and Liu, 2012). The main goals of these methods weren't to extract opinion targets/words, but to categorize all given aspect terms and sentiment words. Although these models could be used for our task according to the associations between candidates and topics, solely employing semantic relations is still one-sided and insufficient to obtain expected performance.

Furthermore, there is little work which considered these two types of relations globally (Su et al., 2008; Hai et al., 2012; Bross and Ehrig, 2013). They usually captured different relations using co-occurrence information. That was too coarse to obtain expected results (Liu et al., 2012). In addition, (Hai et al., 2012) extracted opinion targets/words in a bootstrapping process, which had an error propagation problem. In contrast, we perform extraction with a global graph co-ranking process, where error propagation can be effectively alleviated. (Su et al., 2008) used heterogeneous relations to find implicit sentiment associations among words. Their aim was only to perform aspect terms categorization but not to extract opinion targets/words. They extracted opinion targets/words in advanced through simple phrase detection. Thus, the extraction performance is far from expectation.

3 The Proposed Method

In this section, we propose our method in detail. We formulate opinion targets/words extraction as a co-ranking task. All nouns/noun phrases are regarded as opinion target candidates, and all adjectives/verbs are regarded as opinion word candidates, which are widely adopted by pervious methods (Hu and Liu, 2004a; Qiu et al., 2011; Wang and Wang, 2008; Liu et al., 2012). Then each candidate will be assigned a confidence and ranked, and the candidates with higher confidence than a threshold will be extracted as the results.

Different from traditional methods, besides opinion relations among words, we additionally capture semantic relations among homogeneous candidates. To this end, a heterogeneous undirected graph $G = (V, E)$ is constructed. $V = V^t \cup V^o$ denotes the vertex set, which includes opinion target candidates $v^t \in V^t$ and opinion word candidates $v^o \in V^o$. E denotes the edge set, where $e_{ij} \in E$ means that there is a relation between two vertices. $E^{tt} \subset E$ represents the semantic relations between two opinion target candidates. $E^{oo} \subset E$ represents the semantic relations between two opinion word candidates. $E^{to} \subset E$ represents the opinion relations between opinion target candidates and opinion word candidates. Based on different relation types, we used three matrices $M_{tt} \in \mathbb{R}^{|V^t|\times|V^t|}$, $M_{oo} \in \mathbb{R}^{|V^o|\times|V^o|}$ and $M_{to} \in \mathbb{R}^{|V^t|\times|V^o|}$ to record the association weights between any two vertices, respectively. Section 3.4 will illustrate how to construct them.

3.1 Only Considering Opinion Relations

To estimate the confidence of each candidate, we use a random walk algorithm on our graph to perform co-ranking. Most previous methods (Hu and Liu, 2004a; Qiu et al., 2011; Wang and Wang, 2008; Liu et al., 2012) only considered opinion relations among words. Their basic assumption is as follows.

> **Assumption 1:** If a word is likely to be an opinion word, the words which it has opinion relation with will have higher confidence to be opinion targets, and vice versa.

In this way, candidates' confidences (v^t or v^o) are collectively determined by each other iteratively. It equals to making random walk on subgraph $G^{to} = (V, E^{to})$ of G. Thus we have

$$\begin{aligned} C_t &= (1-\mu) \times M_{to} \times C_o + \mu \times I_t \\ C_o &= (1-\mu) \times M_{to}^T \times C_t + \mu \times I_o \end{aligned} \tag{1}$$

where C_t and C_o respectively represent confidences of opinion targets and opinion words. $m_{i,j}^{to} \in M_{to}$ means the association weight between the ith opinion target and the jth opinion word according to their opinion relations.

It's worthy noting that I_t and I_o respectively denote prior confidences of opinion target candidates and opinion word candidates. We argue that opinion targets are usually domain-specific, and there are remarkably distribution difference of them on different domains (in-domain D_{in} vs. out-domain D_{out}). If a candidate is salient in D_{in} but common in D_{out}, it's likely to be an opinion target in D_{in}. Thus, we use a domain relevance measure (DR) (Hai et al., 2013) to compute I_t.

$$DR(t) = \frac{R(t, D_{in})}{R(t, D_{out})} \tag{2}$$

where $R(t, D) = \frac{\bar{w}_t}{s_t} \times \sum_{j=1}^{N}(w_{tj} - \frac{1}{W_j} \times \sum_{k=1}^{W_j} w_{kj})$ represents candidate relevance with domain D. $w_{tj} = (1 + logTF_{tj}) \times log\frac{N}{DF_t}$ is a TF-IDF-like weight of candidate t in document j. TF_{tj} is the frequency of the candidate t in the jth document, and DF_t is document frequency. N means the document number in domain D. $R(t, D)$ includes two measures to reflect the salient of a candidate in D. 1) $w_{tj} - \frac{1}{W_j} \times \sum_{k=1}^{W_j} w_{kj}$ reflects how frequently a term is mentioned in a particular document. W_j denotes the word number in document j. 2) $\frac{\bar{w}_t}{s_t}$ quantifies how significantly a term is mentioned across all documents in D. $\bar{w}_t = \frac{1}{N} \times \sum_{k=1}^{N} w_{tk}$ denotes average weight across all documents for t. $s_t = \sqrt{\frac{1}{N} \times \sum_{j=1}^{N} (w_{tj} - \bar{w}_j)^2}$ denotes the standard variance of term t. We use the given reviews as in-domain collection D_{in} and Google n-gram corpus[1] as out-domain collection D_{out}. Finally, each entry in I_t is a normalized $DR(t)$ score. In contrast, opinion words are usually domain-independent. Users may use same words to express theirs opinions, like "good", "bad", etc. But there are still some domain-dependent opinion words, like "delicious" in the restaurant domain, "powerful" in the car domain. It's difficult to discriminate them from other words by using statistical information. So we simply set all entries in I_o to be 1. $\mu \in [0, 1]$ in Eq.1 determines the impact of the prior confidence on results.

3.2 Only Considering Semantic Relations

To estimate candidates' confidences by only considering semantic relations among words, we make two separately random walks on the subgraphs of G, $G^{tt} = (V, E^{tt})$ and $G^{oo} = (V, E^{oo})$. The basic assumption is as follows:

> **Assumption 2:** If a word is likely to be an opinion target (opinion word), the words which it has strong semantic relation with will have higher confidence to be opinion targets (opinion words).

In this way, the confidence of the candidate is determined only by its homogeneous neighbours. There is no mutual reinforcement between opinion targets and opinion words. Thus we have

$$\begin{aligned} C_t &= (1-\nu) \times M_{tt} \times C_t + \nu \times I_t \\ C_o &= (1-\nu) \times M_{oo} \times C_o + \nu \times I_o \end{aligned} \tag{3}$$

where ν has the same role as μ in Eq.1.

3.3 Considering Semantic Relations and Opinion Relations Together

To jointly model semantic relations and opinion relations for opinion targets/words extraction, we couple two random walking algorithms mentioned above together. Here, **Assumption 1** and **Assumption 2** are both satisfied. Thus, an opinion target/word candidate's confidence is collectively determined by its neighbours according to different relation types. Meanwhile, each item may make influence on it's neighbours. It's an iterative reinforcement process. Thus, we have

$$\begin{aligned} C_t &= (1-\lambda-\mu) \times M_{to} \times C_o \\ &\quad + \lambda \times M_{tt} \times C_t + \mu \times I_t \\ C_o &= (1-\lambda-\mu) \times M_{to}^T \times C_t \\ &\quad + \lambda \times M_{oo} \times C_o + \mu \times I_o \end{aligned} \tag{4}$$

where $\lambda \in [0, 1]$ determines which type of relations dominates candidate confidence estimation. $\lambda = 0$ means that each candidate's confidence is estimated by only considering opinion relations among words, which equals to Eq.1. Otherwise, when $\lambda = 1$, candidate confidence estimation only

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

considers semantic relations among words, which equals to Eq.3. μ, I_o and I_t have the same meaning in Eq.1. Our algorithm will run iteratively until it converges or in a fixed iteration number $Iter$. In experiments, we set $Iter = 200$.

Obtaining Word Preference. The co-ranking algorithm in Eq.4 is based on a standard random walking algorithm, which randomly selects a link according to the association matrix M_{to}, M_{tt} and M_{oo}, or jumps to a random node with prior confidence value. However, it generates a global ranking over all candidates without taking the node preference (word preference) into account. As mentioned in the first section, each opinion target/word has its preferred collocations, it's reasonable that the confidence of an opinion target (opinion word) candidate should be preferentially determined by its preferences, rather than all of its neighbors with opinion relations.

To obtain the word preference, we resort to topics. We believe that if an opinion word ${v_i}^o$ is topical related with a target word ${v_j}^t$, ${v_i}^o$ can be regarded as a word preference for ${v_j}^t$, and vice versa. For example, "*price*" and "*expensive*" are topically related in phone's domain, so they are a word preference for each other.

Specifically, we use a vector $P^{T_i} = [P_1^{T_i}, ..., P_k^{T_i}, ..., P_{|V^o|}^{T_i}]_{1\times|V^o|}$ to represent word preference of the ith opinion target candidate. $P_k^{T_i}$ means the preferred probability of the ith potential opinion target for the kth potential opinion words. To compute $P_k^{T_i}$, we first use *Kullback-Leibler* divergence to measure the semantic distance between any two candidates on the bridge of topics. Thus, we have

$$D(v_i, v_j) = \frac{1}{2}\Sigma_z(KL_z(v_i||v_j) + KL_z(v_j||v_i))$$

where $KL_z(v_i||v_j) = p(z|v_i)log\frac{p(z|v_i)}{p(z|v_j)}$ means the KL-divergence from candidate v_i to v_j based on topic z. $p(z|v) = p(v|z)\frac{p(z)}{p(v)}$, where $p(v|z)$ is the probability of the candidate v to topic z (see Section 3.4). $p(z)$ is the probability that topic z in reviews. $p(v)$ is the probability that a candidate occurs in reviews. Then, a logistic function is used to map $D(v_i, v_j)$ into $[0, 1]$.

$$SA(v_i, v_j) = \frac{1}{1 + e^{D(v_i,v_j)}} \tag{5}$$

Then, we calculate $P_k^{T_i}$ by normalize $SA(v_i, v_j)$ score, i.e. $P_k^{T_i} = \frac{SA(v_i^t, v_k^o)}{\sum_{p=1}^{|V^o|} SA(v_i^t, v_p^o)}$. For demonstration, we give some examples in Table 1, where each entry denotes a $SA(v_i, v_j)$ score between two candidates. We can see that using topics can successfully capture the preference information for each opinion target/word.

	expensive	good	long	colorful
price	0.265	0.043	0.003	0.000
LED	0.002	0.035	0.007	0.098
battery	0.000	0.015	0.159	0.001

Table 1: Examples of Calculated Word Preference

And we use a vector $P^{O_j} = [P_1^{O_j}, ..., P_q^{O_j}, ..., P_{|V^t|}^{O_j}]_{1\times|V^t|}$ to represent the preference information of the jth opinion word candidate. Similarly, we have $P_q^{O_j} = \frac{SA(v_q^t, v_j^o)}{\sum_{k=1}^{|V^t|} SA(v_k^t, v_j^o)}$.

Incorporating Word Preference into Coranking. To consider such word preference in our co-ranking algorithm, we incorporate it into the random walking on G^{to}. Intuitively, preference vectors will be different for different candidates. Thus, the co-ranking algorithm would be personalized. It allows that the candidate confidence propagates to other candidates only in its preference cluster. Specifically, we make modification on original transition matrix $M_{to} = (M_1^{to}, M_2^{to}, ..., M_{|V^t|}^{to})$ and add each candidate's preference in it. Let $\hat{M}_{to} = (\hat{M}_1^{to}, \hat{M}_2^{to}, ..., \hat{M}_{|V^t|}^{to})$ be the modified transition matrix, which records the associations between opinion target candidates and opinion word candidates. Here $M_k^{to} \in \mathbb{R}^{1\times|V^o|}$ and $\hat{M}_k^{to} \in \mathbb{R}^{1\times|V^o|}$ denotes the kth column vector in M_{to} and $\hat{M}_{to}$, respectively. And let $Diag(P^{T_k})$ denote a diagonal matrix whose eigenvalue is vector P^{T_k}, we have

$$\hat{M}_k^{to} = M_k^{to} Diag(P^{T_k})$$

Similarly, let $U_k^{to} \in \mathbb{R}^{1\times|V^t|}$ and $\hat{U}_k^{to} \in \mathbb{R}^{1\times|V^t|}$ denotes the kth row vector in M_{to}^T and $\hat{M}_{to}^T$, respectively. $Diag(P^{O_k})$ denote a diagonal matrix whose eigenvalue is vector P^{O_k}. Then we have

$$\hat{U}_k^{to} = U_k^{to} Diag(P^{O_k})$$

In this way, each candidate's preference is incorporated into original associations based on opinion relation M_{to} through $Diag(P^{O_k})$ and $Diag(P^{T_k})$. And candidates' confidences will mainly come from the contributions of its preferences. Thus, C_t and C_o in Eq.4 become:

$$
\begin{aligned}
C_t &= (1-\lambda-\mu)\times \hat{M}_{to}\times C_o \\
&\quad + \lambda \times M_{tt}\times C_t + \mu \times I_t \\
C_o &= (1-\lambda-\mu)\times \hat{M}_{to}^{T}\times C_t \\
&\quad + \lambda \times M_{oo}\times C_o + \mu \times I_o
\end{aligned}
\tag{6}
$$

3.4 Capturing Semantic and Opinion Relations

In this section, we explain how to capture semantic relations and opinion relations for constructing transition matrices M_{tt}, M_{oo} and M_{to}.

Capturing Semantic Relations: For capturing semantic relations among homogenous candidates, we employ topics. We believe that if two candidates share similar topics in the corpus, there is a strong semantic relation between them. Thus, we employ a LDA variation (Mukherjee and Liu, 2012), an extension of (Zhao et al., 2010), to discover topic distribution on words, which sampled all words into two separated observations: opinion targets and opinion words. It's because that we are only interested in topic distribution of opinion targets/words, regardless of other useless words, including conjunctions, prepositions etc. This model has been proven to be better than the standard LDA model and other LDA variations for opinion mining (Mukherjee and Liu, 2012).

After topic modeling, we obtain the probability of the candidates (v^t and v^o) to topic z, i.e. $p(z|v^t)$ and $p(z|v^o)$, and topic distribution $p(z)$. Then, a symmetric *Kullback-Leibler* divergence as same as Eq.5 is used to calculate the semantical associations between any two homogenous candidates. Thus, we obtain $SA(v^t, v^t)$ and $SA(v^o, v^o)$, which correspond to the entries in M_{tt} and M_{oo}, respectively.

Capturing Opinion Relations: To capture opinion relations among words and construct the transition matrix M_{to}, we used an alignment-based method proposed in (Liu et al., 2013b). This approach models capturing opinion relations as a monolingual word alignment process. Each opinion target can find its corresponding modifiers in sentences through alignment, in which multiple factors are considered globally, such as co-occurrence information, word position in sentence, etc. Moreover, this model adopted a partially supervised framework to combine syntactic information with alignment results, which has been proven to be more precise than the state-of-the-art approaches for opinion relations identification (Liu et al., 2013b).

After performing word alignment, we obtain a set of word pairs composed of a noun (noun phrase) and its corresponding modified word. Then, we simply employ Pointwise Mutual Information (PMI) to calculate the opinion associations among words as the entries in M_{to}. $OA(v^t, v^o) = log\frac{p(v^t,v^o)}{p(v^t)p(v^o)}$, where v^t and v^o denote an opinion target candidate and an opinion word candidate, respectively. $p(v^t, v^o)$ is the co-occurrence probability of v^t and v^o based on the opinion relation identification results. $p(v^t)$ and $p(v^o)$ give the independent occurrence probability of of v^t and v^o, respectively

4 Experiments

4.1 Datasets and Evaluation Metrics

Datasets: To evaluate the proposed method, we used three datasets. The first one is *Customer Review Datasets* (CRD), used in (Hu and Liu, 2004a), which contains reviews about five products. The second one is *COAE2008* dataset2[2], which contains Chinese reviews about four products. The third one is *Large*, also used in (Wang et al., 2011; Liu et al., 2012; Liu et al., 2013a), where two domains are selected (Mp3 and Hotel). As mentioned in (Liu et al., 2012), *Large* contains 6,000 sentences for each domain. Opinion targets/words are manually annotated, where three annotators were involved. Two annotators were required to annotate out opinion words/targets in reviews. When conflicts occur, the third annotator make final judgement. In total, we respectively obtain 1,112, 1,241 opinion targets and 334, 407 opinion words in Hotel, MP3.

Pre-processing: All sentences are tagged to obtain words' part-of-speech tags using Stanford NLP tool[3]. And noun phrases are identified using the method in (Zhu et al., 2009) before extraction.

Evaluation Metrics: We select precision(P), recall(R) and f-measure(F) as metrics. And a significant test is performed, i.e., a t-test with a default significant level of 0.05.

4.2 Our Method vs. The State-of-the-art Methods

To prove the effectiveness of the proposed method, we select some state-of-the-art methods for comparison as follows:

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

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

Methods	D1			D2			D3			D4			D5			Avg.
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	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	0.758
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	0.856
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	0.846
SAS	0.80	0.79	0.79	0.82	0.76	0.79	0.79	0.74	0.76	0.77	0.78	0.77	0.80	0.76	0.78	0.778
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	0.858
Hai	0.77	**0.87**	0.83	0.79	**0.86**	0.82	0.79	0.89	0.84	0.72	0.88	0.79	0.74	0.88	0.81	0.818
CR	0.84	0.86	0.85	0.87	0.85	0.86	0.87	**0.90**	0.88	**0.81**	0.87	**0.83**	0.89	0.88	0.89	0.862
CR_WP	0.86	0.86	**0.86**	0.88	**0.86**	**0.87**	0.89	**0.90**	**0.89**	**0.81**	0.87	**0.83**	0.91	**0.89**	**0.90**	**0.870**

Table 2: Results of Opinion Targets Extraction on *Customer Review Dataset*

Methods	Camera			Car			Laptop			Phone			Mp3			Hotel			Avg.
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	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	0.61	0.68	0.64	0.60	0.65	0.62	0.587
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	0.69	0.70	0.69	0.67	0.69	0.68	0.683
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	0.67	0.77	0.72	0.67	0.76	0.71	0.712
SAS	0.72	0.72	0.72	0.71	0.64	0.67	0.59	0.72	0.65	0.78	0.69	0.73	0.69	0.75	0.72	0.69	0.74	0.71	0.700
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	0.70	0.82	0.76	0.71	0.80	0.75	0.749
Hai	0.68	**0.84**	0.76	0.69	**0.75**	0.72	0.58	**0.86**	0.72	0.75	0.76	0.76	0.65	0.83	0.74	0.62	0.82	0.75	0.742
CR	0.75	0.83	0.79	0.72	0.74	0.73	0.60	0.85	0.70	0.83	**0.77**	**0.80**	0.70	**0.84**	0.76	0.71	**0.83**	0.77	0.758
CR_WP	**0.78**	**0.84**	**0.81**	**0.74**	**0.75**	**0.74**	**0.64**	0.85	**0.73**	**0.84**	0.76	**0.80**	**0.74**	**0.84**	**0.79**	**0.74**	0.82	**0.78**	**0.773**

Table 3: Results of Opinion Targets Extraction on *COAE 2008* and *Large*

Hu extracted opinion targets/words using association mining rules (Hu and Liu, 2004a).

DP used syntax-based patterns to capture opinion relations in sentences, and then used a bootstrapping process to extract opinion targets/words (Qiu et al., 2011),.

Zhang is proposed by (Zhang et al., 2010). They also used syntactic patterns to capture opinion relations between words. Then a HITS (Kleinberg, 1999) algorithm is employed to extract opinion targets.

Liu is proposed by (Liu et al., 2013a), an extension of (Liu et al., 2012). They employed a word alignment model to capture opinion relations among words, and then used a random walking algorithm to extract opinion targets.

Hai is proposed by (Hai et al., 2012), which is similar to our method. They employed both of semantic relations and opinion relations to extract opinion words/targets in a bootstrapping framework. But they captured relations only using cooccurrence statistics. Moreover, word preference was not considered.

SAS is proposed by (Mukherjee and Liu, 2012), an extended lda-based model of (Zhao et al., 2010). The top K items for each aspect are extracted as opinion targets/words. It means that only semantic relations among words are considered in SAS. And we set aspects number to be 9 as same as (Mukherjee and Liu, 2012).

CR: is the proposed method in this paper by using co-ranking, referring to Eq.4. CR doesn't consider word preference.

CR_WP: is the full implementation of our method, referring to Eq.6.

Hu, DP, Zhang and Liu are the methods which only consider opinion relations among words. SAS is the methods which only consider semantic relations among words. Hai, CR and CR_WP consider these two types of relations together. The parameter settings of state-of-the-art methods are same as their original paper. In CR and CR_WP, we set $\lambda = 0.4$ and $\mu = 0.1$. The experimental results are shown in Table 2, 3, 4 and 5, where the last column presents the average F-measure scores for multiple domains. Since Liu and Zhang aren't designed for opinion words extraction, we don't present their results in Table 4 and 5. From experimental results, we can see.

1) Our methods (CR and CR_WP) outperform other methods not only on opinion targets extraction but on opinion words extraction in most domains. It proves the effectiveness of the proposed method.

2) CR and CR_WP have much better performance than Liu and Zhang, especially on Recall. Liu and Zhang also use a ranking framework like ours, but they only employ opinion relations for extraction. In contrast, besides opinion relations, CR and CR_WP further take semantic relations into account. Thus, more opinion targets/words can be extracted. Furthermore, we observe that CR and CR_WP outperform SAS. SAS only exploits semantic relations, but ignores opinion relations among words. Its extraction is performed separately and neglects the reinforcement between opinion targets and opinion words. Thus, SAS has worse performance than our methods. It demonstrates the usefulness of considering multiple relation types.

Methods	D1			D2			D3			D4			D5			Avg.
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	F
Hu	0.57	0.75	0.65	0.51	0.76	0.61	0.57	0.73	0.64	0.54	0.62	0.58	0.62	0.67	0.64	0.624
DP	0.64	0.73	0.68	0.57	0.79	0.66	0.65	0.70	0.67	0.61	0.65	0.63	0.70	0.68	0.69	0.666
SAS	0.64	0.68	0.66	0.55	0.70	0.62	0.62	0.65	0.63	0.60	0.61	0.60	0.68	0.63	0.65	0.632
Hai	0.62	**0.77**	0.69	0.52	**0.80**	0.64	0.60	0.74	0.67	0.56	**0.69**	0.62	0.66	**0.70**	0.68	0.660
CR	0.62	0.75	0.68	0.57	0.79	0.67	0.64	**0.75**	0.69	0.63	**0.69**	0.66	0.68	0.69	0.69	0.678
CR_WP	**0.65**	0.75	**0.70**	**0.59**	**0.80**	**0.68**	**0.65**	0.74	**0.70**	**0.66**	0.68	**0.67**	**0.71**	**0.70**	**0.70**	**0.690**

Table 4: Results of Opinion Words Extraction on *Customer Review Dataset*

Methods	Camera			Car			Laptop			Phone			Mp3			Hotel			Avg.
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	F
Hu	0.72	0.74	0.73	0.70	0.71	0.70	0.66	0.70	0.68	0.70	0.70	0.70	0.48	0.67	0.56	0.52	0.69	0.59	0.660
DP	0.80	0.73	0.76	0.79	0.71	0.75	0.75	0.69	0.72	0.78	0.68	0.73	0.60	0.65	0.62	0.61	0.66	0.63	0.702
SAS	0.73	0.70	0.71	0.75	0.68	0.71	0.72	0.68	0.69	0.71	0.66	0.68	0.64	0.62	0.63	0.66	0.61	0.63	0.675
Hai	0.76	0.74	0.75	0.72	**0.74**	0.73	0.69	**0.72**	0.70	0.72	0.70	0.71	0.61	**0.69**	0.64	0.59	0.68	0.64	0.690
CR	**0.80**	**0.75**	**0.77**	0.77	**0.74**	0.75	0.73	0.71	0.72	0.75	0.71	0.73	0.63	**0.69**	0.64	0.63	0.68	0.66	0.710
CR_WP	**0.80**	**0.75**	**0.77**	**0.80**	**0.74**	**0.77**	**0.77**	0.71	**0.74**	**0.78**	**0.72**	**0.75**	**0.66**	0.68	**0.67**	**0.67**	**0.69**	**0.68**	**0.730**

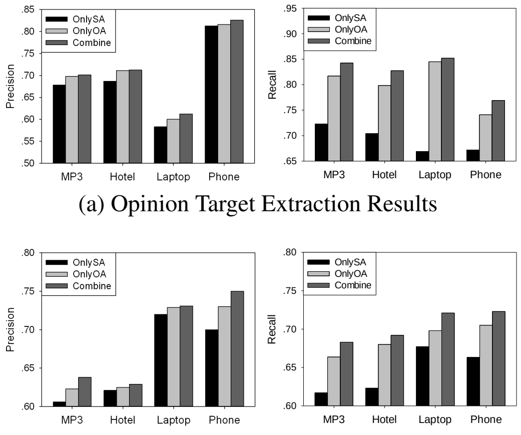
Table 5: Results of Opinion Words Extraction on *COAE 2008* and *Large*

3) CR and CR_WP both outperform Hai. We believe the reasons are as follows. First, CR and CR_WP considers multiple relations in a unified process by using graph co-ranking. In contrast, Hai adopts a bootstrapping framework which performs extraction step by step and may have the problem of error propagation. It demonstrates that our graph co-ranking is more suitable for this task than bootstrapping-based strategy. Second, our method captures semantic relations using topic modeling and captures opinion relations through word alignments, which are more precise than Hai which merely uses co-occurrence information to indicate such relations among words. In addition, word preference is not handled in Hai, but processed in CR_WP. The results show the usefulness of word preference for opinion targets/words extraction.

4) CR_WP outperforms CR, especially on precision. The only difference between them is that CR_WP considers word preference when performing graph ranking for candidate confidence estimation, but CR does not. Each candidate confidence estimation in CR_WP gives more weights for this candidate's preferred words than CR. Thus, the precision can be improved.

4.3 Semantic Relation vs. Opinion Relation

In this section, we discuss which relation type is more effective for this task. For comparison, we design two baselines, called **OnlySA** and **OnlyOA**. OnlyOA only employs opinion relations among words, which equals to Eq.1. OnlySA only employs semantic relations among words, which equals to Eq.3. Moreover, **Combine** is our method which considers both of opinion relations and semantic relations together, referring to Eq.4 with $\lambda = 0.5$. Figure 2 presents experimental results. The left graph presents opinion targets extraction results and the right graph presents opinion words extraction results. Because of space limitation, we only shown the results of four domains (MP3, Hotel, Laptop and Phone).



(a) Opinion Target Extraction Results

(b) Opinion Word Extraction Results

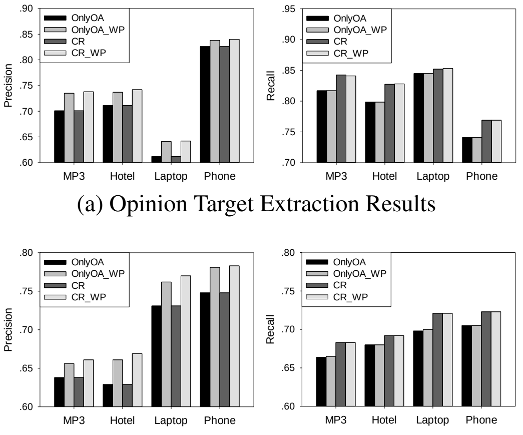
Figure 2: Semantic Relations vs. Opinion Relations

From results, we observe that OnlyOA outperforms OnlySA in all domains. It demonstrates that employing opinion relations are more useful than semantic relations for co-extracting opinion targets/words. And it is necessary to utilize the mutual reinforcement relationship between opinion words and opinion targets. Moreover, Combine outperforms OnlySA and OnlyOA in all domains. It indicates that combining different relations among words together is effective.

4.4 The Effectiveness of Considering Word Preference

In this section, we try to prove the necessity of considering word preference in Eq.6. Besides the comparison between CR and CR_WP performed in the main experiment in Section 4.2, we further incorporate word preference in aforementioned **OnlyOA**, named as **OnlyOA_WP**, which only employs opinion relations among words and equals to Eq.6 with $\lambda = 0$. Experimental results are shown in Figure 3. Because of space limitation, we only show the results of the same domains in section 4.3,

Form results, we observe that CR_WP outperforms CR, and OnlyOA_WP outperforms OnlyOA in all domains, especially on precision. These observations demonstrate that considering word preference is very important for opinion targets/words extraction. We believe the reason is that exploiting word preference can provide more fine information for opinion target/word candidates' confidence estimation. Thus the performance can be improved.



(a) Opinion Target Extraction Results

(b) Opinion Word Extraction Results

Figure 3: Experimental results when considering word preference

4.5 Parameter Sensitivity

In this subsection, we discuss the variation of extraction performance when changing λ and μ in Eq.6. Due to space limitation, we only show the F-measure of CR_WP on four domains. Experimental results are shown in Figure 4 and Figure 5. The left graphs in Figure 4 and 5 present the performance variation of CR_WP with varying λ from 0 to 0.9 and fixing $\mu = 0.1$. The right graphs in Figure 4 and 5 present the performance variation of CR_WP with varying μ from 0 to 0.6 and fixing $\lambda = 0.4$.

In the left graphs in Figure 4 and 5, we observe the best performance is obtained when $\lambda = 0.4$. It indicates that opinion relations and semantic relations are both useful for extracting opinion targets/words. The extraction performance is beneficial from their combination. In the right graphs in Figure 4 and 5, the best performance is obtained when $\mu = 0.1$. It indicates prior knowledge is useful for extraction. When μ increases, performance, however, decreases. It demonstrates that incorporating more prior knowledge into our algorithm would restrain other useful clues on estimating candidate confidence, and hurt the performance.

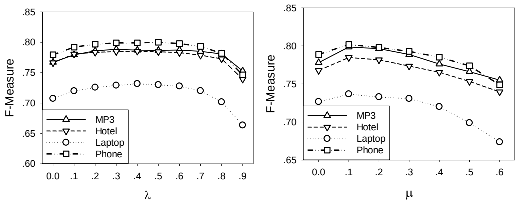


Figure 4: Opinion targets extraction results

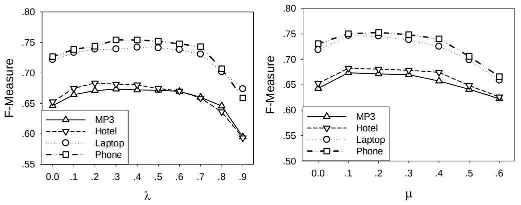


Figure 5: Opinion words extraction results

5 Conclusions

This paper presents a novel method with graph coranking to co-extract opinion targets/words. We model extracting opinion targets/words as a coranking process, where multiple heterogenous relations are modeled in a unified model to make cooperative effects on the extraction. In addition, we especially consider word preference in co-ranking process to perform more precise extraction. Compared to the state-of-the-art methods, experimental results prove the effectiveness of our method.

Acknowledgement

This work was sponsored by the National Basic Research Program of China (No. 2014CB340500), the National Natural Science Foundation of China (No. 61272332 and No. 61202329), the National High Technology Development 863 Program of China (No. 2012AA011102), and CCF-Tencent Open Research Fund.

References

Juergen Bross and Heiko Ehrig. 2013. Automatic construction of domain and aspect specific sentiment lexicons for customer review mining. In *Proceedings of the 22nd ACM international conference on Conference on information &#38; knowledge management*, CIKM '13, pages 1077–1086, New York, NY, USA. ACM.

Zhen Hai, Kuiyu Chang, and Gao Cong. 2012. One seed to find them all: mining opinion features via association. In *CIKM*, pages 255–264.

Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang. 2013. Identifying features in opinion mining via intrinsic and extrinsic domain relevance. *IEEE Transactions on Knowledge and Data Engineering*, 99(PrePrints):1.

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

Minqing 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.

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, Minqing 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. 2013a. Opinion target extraction using partially supervised word alignment model.

Kang Liu, Liheng Xu, and Jun Zhao. 2013b. Syntactic patterns versus word alignment: Extracting opinion targets from online reviews.

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.

Samaneh Moghaddam and Martin Ester. 2011. Ilda: Interdependent lda model for learning latent aspects and their ratings from online product reviews. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 665–674, New York, NY, USA. ACM.

Samaneh Moghaddam and Martin Ester. 2012a. Aspect-based opinion mining from product reviews. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 1184–1184, New York, NY, USA. ACM.

Samaneh Moghaddam and Martin Ester. 2012b. On the design of lda models for aspect-based opinion mining. In *CIKM*, pages 803–812.

Arjun Mukherjee and Bing Liu. 2012. Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12, pages 339–348, Stroudsburg, PA, USA. Association for Computational Linguistics.

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.

Qi Su, Xinying Xu, Honglei Guo, Zhili Guo, Xian Wu, Xiaoxun Zhang, Bin Swen, and Zhong Su. 2008. Hidden sentiment association in chinese web opinion mining. In Jinpeng Huai, Robin Chen, Hsiao-Wuen Hon, Yunhao Liu, Wei-Ying Ma, Andrew Tomkins, and Xiaodong Zhang 0001, editors, *WWW*, pages 959–968. ACM.

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.

Bishan Yang and Claire Cardie. 2013. Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*

Papers), pages 1640–1649, Sofia, Bulgaria, August. Association for Computational Linguistics.

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.

Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li. 2010. Jointly modeling aspects and opinions with a maxent-lda hybrid. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 56–65, Stroudsburg, PA, USA. Association for Computational Linguistics.

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.