Walk and Learn: A Two-Stage Approach for Opinion Words and Opinion Targets Co-Extraction

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

ABSTRACT

This paper proposes a novel two-stage method for opinion words and opinion targets co-extraction. In the first stage, a *Sentiment Graph Walking* algorithm is proposed, which naturally incorporates syntactic patterns in a graph to extract opinion word/target candidates. In the second stage, we adopt a self-Learning strategy to refine the results from the first stage, especially for filtering out noises with high frequency and capturing long-tail terms. Preliminary experimental evaluation shows that considering pattern confidence in the graph is beneficial and our approach achieves promising improvement over three competitive baselines.

Categories and Subject Descriptors

I.2.7 [**Natural Language Processing**]: Text analysis

Keywords

Sentiment Analysis; Opinion Words; Opinion Targets

1. INTRODUCTION

Extracting opinion words and opinion targets are two key tasks in Opinion Mining or Sentiment Analysis, which have attracted much attention from both the research community and industry in recent years. Opinion words and opinion targets often co-occur in reviews and there exists modified relation (called *opinion relation* in this paper) between them. For example, in the sentence "It has a clear screen", "clear" is an opinion word and "screen" is an opinion target, and there is an opinion relation between the two words.

Previous works [3, 5] exploited syntactic patterns to identify opinion relations, which had achieved superior performance over co-occurrence-based method [1]. However, syntax-based methods still have some limitations: (i) As an example, the phrase "everyday at school" can be matched by a syntactic pattern "JJ-{prep}-{pobj}-NN", but it bears no sentiment orientation. We call such relations that match opinion patterns but express no opinion ***false opinion relations***. (ii) In another case, the phrase "wonderful time" can be matched by a pattern "JJ-{amod}-NN". This phrase does express a positive opinion but unfortunately "time" is not a valid opinion target for most domains such as MP3. Thus, ***false opinion targets*** are extracted. (iii) We further notice that previous works [1, 3, 5] often rank opinion targets by term frequency. Hence, they have the difficulty in identifying the ***long-tail opinion targets***.

To address the problems stated above, this paper proposes a novel two-stage method named *Walk and Learn*. In the first stage, we propose a graph-based algorithm called ***Sentiment Graph Walking*** to cope with the *false opinion relation* problem. Concretely, syntactic patterns are incorporated in a Sentiment Graph and random walking is used to estimate confidence of patterns. Thus, terms extracted by low-confidence patterns will have low confidence accordingly. This could potentially improve the extraction accuracy. In the second stage, we adopt a **self-Learning strategy**, which aims to filter out *false opinion targets* and extract *long-tail opinion targets* from the results of the first stage. Preliminary experimental evaluation on two domains of real world reviews shows that our approach gives promising improvement over three competitive baselines.

2. THE FIRST STAGE: SENTIMENT GRAPH WALKING ALGORITHM

Opinion Pattern Extraction. For each sentence, we first obtain its dependency tree. Following [1, 3, 5], all adjectives in the sentence are taken as opinion word candidates (OC) and all nouns are regarded as opinion target candidates (TC). Then candidates are replaced by wildcards TC or OC. Every shortest path between wildcard pairs (OC,TC) or (TC,TC) in dependency tree is extracted as an opinion pattern, which captures opinion relation between an OC and a TC or two TCs. Other words in the path are replaced by POS tags and at most two POSs are allowed in each pattern.

Sentiment Graph Construction. We propose *Sentiment Graph*, which is a weighted, directed graph $G = (V, E, W)$. $V = \{V_{oc} \cup V_{tc} \cup V_p\}$ is the set of vertices, where V_{oc}, V_{tc} and V_p denote the set of opinion word/target/pattern candidates respectively. $E = \{E_{po} \cup E_{pt}\} \subseteq \{V_p \times V_{oc}\} \cup \{V_p \times V_{tc}\}$ is the weighted, bi-directional edge set in G. Note that there are no edges between V_{oc} and V_{tc}. $W : E \to \mathbb{R}^+$ is the weight function which assigns non-negative weight to each edge. For $v_a, v_b \in V$, $(e : v_a \to v_b) \in E$, the weight function $w(v_a, v_b) = freq(v_a, v_b)/freq(v_a)$, where $freq(\cdot)$ is the frequency of a candidate extracted by opinion patterns or co-occurrence frequency among candidates.

Confidence Estimation. Random Walking (RW) algorithm is employed to estimate confidence of candidates. Let $\mathbf{M}_{oc\_p}$ denotes the transition matrix from V_{oc} to V_p, similarly, we have $\mathbf{M}_{tc\_p}$, $\mathbf{M}_{p\_oc}$, $\mathbf{M}_{p\_tc}$. Let $\mathbf{c}^t_{oc}$, $\mathbf{c}^t_{tc}$ and $\mathbf{c}^t_p$ denote row vectors after walking t steps for confidence of

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

opinion words/targets/patterns. Initially $\mathbf{c}_{oc}^{0}$ is uniformly distributed on a few opinion word seeds, then the following formula are updated iteratively until $\mathbf{c}_{tc}^{t}$ and $\mathbf{c}_{oc}^{t}$ converge:

$$\mathbf{c}_{p}^{t+1} = \mathbf{c}_{oc}^{t} \times \mathbf{M}_{oc\_p} + \mathbf{c}_{tc}^{t} \times \mathbf{M}_{tc\_p} \quad (1)$$

$$\mathbf{c}_{oc}^{t+1} = \mathbf{c}_{p}^{t} \times \mathbf{M}_{p\_oc}\ ,\ \mathbf{c}_{tc}^{t+1} = \mathbf{c}_{p}^{t} \times \mathbf{M}_{p\_tc} \quad (2)$$

3. THE SECOND STAGE: REFINING RESULTS BY SELF-LEARNING

Opinion Targets Refinement. In the results of the first stage, there are still some issues need to be addressed: (i) The *false opinion targets* problem remains unsolved, because there exist many opinion expressions containing non-target terms such as "good thing" and "nice people" in reviews. As a result, many frivolous general noun noises are included. (ii) *Long-tail opinion targets* may have low degree in Sentiment Graph. Hence their confidence will be low although they may be extracted by some high-confidence patterns. To address these issues, we exploit a semi-supervised classifier TSVM [2] to refine opinion targets as follows.

We find that most top-ranked general noun noises are the most frequently used terms in common texts. Therefore, we create a small domain-independent General Noun Corpus (GNC) from large web corpora such as Google-n-gram[1] to cover some most frequently used general nouns. Then N target candidates with the highest confidence but not in GNC are regarded as positive labeled examples, other N terms from GNC which are also top ranked in target list are selected as negative labeled examples. Other target candidates are regarded as the unlabeled examples.

Let $\mathbf{x_i} = (o_1, \ldots, o_j, p_1, \ldots, p_k)^T$ denotes feature vector of a target candidate t_i, the values of opinion word feature o_j and opinion pattern feature p_k are:

$$x(o_j) = conf(o_j) \times \frac{\sum_{p_k} freq(t_i, o_j, p_k)}{freq(o_j)} \quad (3)$$

$$x(p_k) = conf(p_k) \times \frac{\sum_{o_j} freq(t_i, o_j, p_k)}{freq(p_k)} \quad (4)$$

where $conf(\cdot)$ denotes confidence score estimated by RW. Thus, a long-tail target is determined by its own contexts, whose weights are learnt from frequent opinion targets.

Opinion Words Refinement. We use the classified opinion target list T to further refine opinion words by $s(o_j) = \sum_{t_i \in T} \sum_{p_k} s(t_i) conf(p_k) freq(t_i, o_j, p_k)/freq(t_i)$, where $s(t_i)$ is confidence score exported by TSVM.

4. EXPERIMENTAL EVALUATION

Datasets. Two domains of real world English reviews [4] are selected to evaluate our approach. Two annotators were required to annotate out opinion words/targets. If conflicts happened, a third annotator would make final judgement.

Evaluation Settings. Three methods *Hu* [1], *DP* [3] and *Zhang* [5] are selected as baselines. Several variants of our approach are given. *Ours-Full* is the full implementation of our method. *Ours-Bigraph* constructs a bi-graph between opinion words and targets, so opinion patterns are not included in the graph. *Ours-Stage1* only uses the first stage. *Ours-Stage2* only contains the second stage so $conf(\cdot)$ in Eq. (3) and (4) are set to 1. Minipar[2] is employed for parsing. Opinion seeds used are same as in [3] and N is 50. Precision(P) and Recall(R) are used as the evaluation metrics.

Method	Opinion Targets				Opinion Words			
	MP3		Hotel		MP3		Hotel	
	P	R	P	R	P	R	P	R
Hu	0.53	0.55	0.55	0.57	0.48	0.65	0.51	0.68
DP	0.66	0.57	0.66	0.60	0.58	0.62	0.60	0.66
Zhang	0.65	0.62	0.64	0.66	–	–	–	–
Ours-Bigraph	0.55	0.68	0.58	0.70	0.54	0.68	0.57	0.69
Ours-Stage1	0.62	0.68	0.63	0.71	0.59	**0.69**	0.61	**0.71**
Ours-Stage2	0.53	0.54	0.52	0.57	0.49	0.61	0.50	0.66
Ours-Full	**0.73**	**0.71**	**0.75**	**0.73**	**0.64**	0.67	**0.67**	0.69

Table 1: Performance on two domains.

Discussion on Results. Experimental results are shown in Table 1. *Zhang* do not extract opinion words so their results for opinion words are omitted. We can see that *Ours-Full* outperforms the three baselines. *Ours-Stage1* outperforms *Ours-Bigraph*, especially in precision. We believe it is because *Ours-Stage1* estimated pattern confidence so *false opinion relations* are reduced. Therefore, the consideration of pattern confidence is beneficial as expected. *Ours-Full* achieves much better performance than *Ours-Stage1*, which alleviates the shortcoming of *false opinion target* problem. Also, *Ours-Stage2* has much worse performance than *Ours-Full*, showing the confidence scores estimated in the first stage are indispensable and indeed key to the learning of the second stage. Furthermore, the average recall of long-tail targets[3] of *Hu*, *DP*, *Zhang* and *Ours-Full* are 0.45, 0.48, 0.52 and 0.63 respectively, which shows that our method improves the limitation of *long-tail opinion target* problem.

5. ACKNOWLEDGEMENTS

Thanks to Prof. Yulan He for her insightful advices. 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).

6. REFERENCES

[1] M. Hu and B. Liu. Mining and summarizing customer reviews. In *KDD '04*, pages 168–177.

[2] T. Joachims. Transductive inference for text classification using support vector machines. In *ICML '99*, pages 200–209.

[3] G. Qiu, B. Liu, J. Bu, and C. Chen. Expanding domain sentiment lexicon through double propagation. In *IJCAI'09*, pages 1199–1204.

[4] H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis without aspect keyword supervision. In *KDD '11*, pages 618–626.

[5] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain. Extracting and ranking product features in opinion documents. In *COLING '10*, pages 1462–1470.

[1] http://books.google.com/ngrams. In practice, we selected 1000 most frequent nouns in Google-1-gram.

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

[3] We conservatively regard 60% opinion targets with the lowest frequency as the "long-tail" terms.