Grammatical Phrase-Level Opinion Target Extraction on Chinese Microblog Messages

Haochen Zhang, Yiqun Liu, Min Zhang, and Shaoping Ma*

State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science and Technology,
Tsinghua University, Beijing 100084, China
zhang-hc10@mails.tsinghua.edu.cn, {yiqunliu,z-m,msp}@tsinghua.edu.cn

Abstract. Microblog is one of the most widely used web applications. Weibo, which is a microblog service in China, produces plenty of opinionated messages every second. Sentiment analysis on Chinese Weibo impacts many aspects of business and politics. In this work, we attempt to address the opinion target extraction, which is one of the most important aspects of sentiment analysis. We propose a unified approach that concentrates on phrase-level target extraction. We assume that a target is represented as a subgraph of the sentence's dependency tree and define the grammatical relations that point to the target word as **TAR-REL**s. We conduct the extraction by classifying grammatical relations with a cost-sensitive classifier that enhances performance of unbalanced data and figuring out the target subgraph by connecting and recovering **TAR-REL**s. Then we prune the noisy targets by empirically summarized rules. The evaluation results indicate that our approach is effective to the phrase-level target extraction on Chinese microblog messages.

Keywords: sentiment analysis, opinion target, phrase-level target extraction, grammatical relation.

1 Introduction

Opinion targets are the entities and the properties that carry authors' opinion. In product review corpus, target is also known as aspect or feature of the product. Detecting and extracting these targets is one of the most important steps to analyze the authors' sentiment. It helps to figure out the entities that users are discussing.

Opinion targets often contain attributives or modifiers. For instance, 'screen' is a target but too ambiguous. Author often describes it with some modifiers like 'iPad screen' or 'computer screen'. These phrases are indeed the complete opinion targets.

* This work was supported by Natural Science Foundation (60903107, 61073071), National High Technology Research and Development (863) Program (2011AA01A207). This work has been done at the NUS–Tsinghua EXtreme search centre (NExT).

G. Zhou et al. (Eds.): NLPCC 2013, CCIS 400, pp. 432–439, 2013.

However, some existing work ignores the completeness of target and only focuses on word-level target extraction [1–5]. Meanwhile, the other work addresses completing phrase-level targets by extending candidates, which is difficult to control the boundaries of the targets.

We proposed a extraction approach that unifies discovering candidates and extending them to the phrase-level target. It is based on the assumption that a phrase-level target is the subgraph of the sentence's dependency tree. We figure out the subgraph by classifying grammatical relations to **TAR-REL**s and **O-REL**s and connecting **TAR-REL**s.

By participating the evaluation task of the 2nd Conference on Natural Language Processing & Chinese Computing(NLP&CC 2013), our approach shows the capacity to extract phrase-level opinion targets from Chinese microblog messages.

2 Related Work

Opinion target extraction is one of the most important aspects of sentiment analysis, which has been investigated by many researchers. There are mainly several different types of approaches, which are summarized as follow:

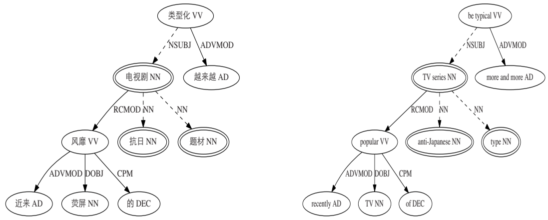
1. Qiu *et al.* [1] and Liu *et al.* [6] propose iterative extraction approaches that adopt targets and opinions to expand each other iteratively.
2. Sayeed *et al.* [3] investigates the structured information and proposes a word-level extraction approach that introduces syntactic features. Ku *et al.* [7] concentrates on morphological structures of Chinese content and adopts SVM and CRFs to classify words and identify targets.
3. Several work [6, 8] addresses the target extraction in multilingual corpus. They attempt to employ language characteristics to enhance performance on both language corpus.
4. Probabilistic models, especially topic models, are employed to extract targets [2, 4, 5]. These approaches performance well and are domain-independent. However, phrase-level target and informal corpus cannot be well solved.

In this work, we address this problem by a grammatical method since we want to discover phrase-level target directly from dependency tree of the sentence.

3 Problem Definition

In this work, we concentrate on the completeness of the target. The completeness of the target extraction requires more efforts on detecting the targets' boundary, which encloses all attributives and descriptive modifiers of the targets. By statistics on the annotated dataset, there are average 2.02 segmented words in a target and only 35% targets consist of only one word. Therefore completeness is quite significant to the target extraction.

As illustrated in Figure 1, nodes of the dependency tree correspond to words of the sentence, and directed edges correspond to grammatical relations. Each



(a) 近来风靡荧屏的抗日题材电视剧越来越类型化。

(b) The anti-Japanese type TV series that are popular on TV are more and more typical.

Fig. 1. Illustration of dependency tree and subgraph-based target

relation has a grammatical relation type. The start word of directed edge is named as governor, while the end word is named as dependent.

We intuitively proposed an assumption that a target is the subgraph of the dependency tree, which is illustrated in Figure 1–(a). The double-circle nodes correspond to words inside the target. Therefore, phrase-level target extraction is then converted to figuring out the target subgraph from the dependency tree.

To figure out expected subgraphs, we classify the directed edges, i.e. the grammatical relations. All relations whose dependents belong to targets are defined as **TAR-REL**, while other relations are defined as **O-REL**. In Figure 1, all **TAR-REL**s are represented in dashed lines.

4 Approach

4.1 Word-Level Sentimental Signal

Before target extraction, we first estimate sentimental signal of each word. Cui *et al.* [9] proposed a graph-based approach that propagates the sentimental signal within the co-occurrence graph and assigns word a positive score and a negative score after the propagation is convergent. The sentimental signal then is estimated by the average of the positive score and the negative score.

4.2 Grammatical Relations Classification

In this section, we explain how to classify grammatical relations by a supervised cost-sensitive learning method.

We adopt nine features, which consist of three categories: general features, sentimental features and grammatical features. These features are extracted from the relation and its corresponding words, which are described in detail as follow:

1. **General Feature:** General features consist of dependent's location (indexed by segmented tokens and denoted as Loc_{dep}), signal of whether dependent

is a named-entity (denoted as NE_{dep}) and dependent and governor's word frequency (denoted as $Freq_{dep}$ and $Freq_{gov}$).
2. **Sentimental Feature:** Sentimental features are dependent and governor's sentimental signals estimated by graph-based propagation. We intuitively assume that targets carry more opinion than those functional or descriptive expressions. They are denoted as $Senti_{dep}$ and $Senti_{gov}$.
3. **Grammatical Feature:** We adopt the relation's dependency type, dependent's POS tag and governor's POS tag as the grammatical features (denoted as Rel, POS_{dep} and POS_{gov}). Grammatical information, especially POS tags, is widely adopted to extract opinion targets.

We employ the Naïve Bayes, one of the most widely adopted classifiers, as the base classifier taking all the above features to classify the relations.

Notice that the **TAR-REL**s and **O-REL**s are not balanced, where there are over 10 times more **O-REL**s. We need a proper boosting meta-algorithm to enhance the performance. MetaCost [10] is a state-of-the-art cost-sensitive algorithm that wraps an arbitrary base classifier. It relabels the result of a base classifier according to Bayes optimal prediction that reduces the conditional risk. By specifying different cost, we actually give a bias on a specific class. In this work, cost of false **TAR-REL**s is assigned to 10 empirically and cost of false **O-REL**s is assigned to 1. Greater cost of false **TAR-REL**s results in bias on the **TAR-REL** class and finally produces more target candidates.

4.3 Phrase-Level Target Candidates

We connect the tagged relations to generate phrase-level target subgraphs. As illustrated by Figure 1, adjacent **TAR-REL** relations are assembled to one target subgraph, and all dependents within this subgraph generate the target. Notice that target excludes the governor of the root relation.

Since these relations are classified independently, targets are probably partially misclassified. we summarize two types of misclassification and reclassify these relations to **TAR-REL**s. We distinguish the misclassification with the following criteria.

1. The **O-REL** that is surrounded by **TAR-REL**s. The governor and dependent of this relation both belong to other two **TAR-REL** relations. Figure 2–(a) illustrates this type of misclassification. E2 is surrounded by **TAR-REL**s and is reclassified as **TAR-REL**. Notice that E5 does not satisfy this criterion and remains **O-REL**.
2. The **O-REL** whose dependent is a leaf node and governor is within a no-less-than-3-word target graph. Figure 2–(b) illustrates this type of misclassification. N1, N2, N4 and N5 construct a 4-word target and N3 is a leaf node. Thus E2 is reclassified as **TAR-REL**.

Afterward, we achieve target subgraphs represented in tree structures. Dependents of relations inside a subgraph generate a target candidate.

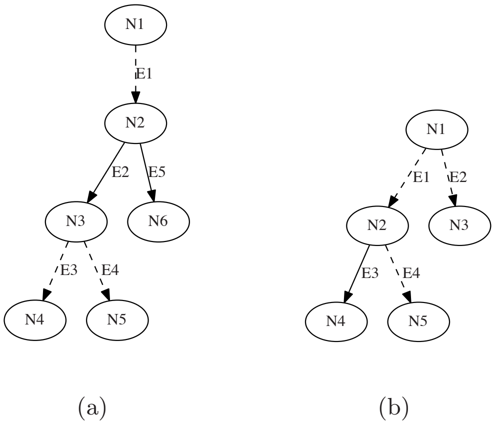


Fig. 2. Illustration of two different misclassification types. Solid and dashed lines represent original **O-REL** and **TAR-REL** respectively.

4.4 Rule-Based Pruning

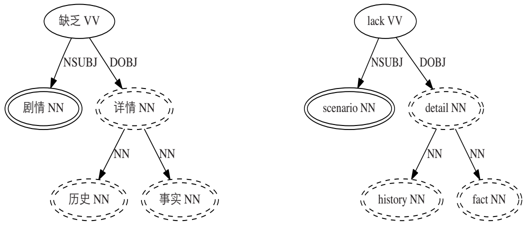
We propose grammatical rules and textual rules to prune these noises after reviewing the dataset and empirical summarization.

The grammatical rules are based on our observation of microblog messages. Microblog messages are more informal and less objective. Therefore subjectives in opinionated microblog sentences are quite more probable to be opinion targets comparing to co-occurred objectives or predicatives, which helps us to establish the grammatical rule: Given a governor and all relations governed by it, if there exists subjective relation and the corresponding dependent is tagged as target, then all targets connected to the governor by objective/predicative relations are pruned.

Figure 3 illustrates one typical case adopting this rule. The double circle nodes are extracted by previous steps. Both subjective and objective are recognized as targets. By following our rule, the objective is pruned, illustrated as dashed border. In this work, we take {**NSUBJ, TOP**} as subjective relations and {**DOBJ, LOBJ, POBJ, ATTR**} as objective/predicative relations.

We also summarize several textual rules by conducting overview of the false targets. These textual rules are presented as follow.

1. If there is only one word in the target, the word must be noun or pronoun.
2. Target's length must be less than $0.8L_{clause}$, where L_{clause} is the length of the clause.
3. Target must be continuously written in the message. As we extract targets from dependency tree, the target may not sequentially adjacent.



(a) 剧情缺乏详细历史事实。 (b) The scenario lacks details of history fact.

Fig. 3. Illustration of typical case that follows grammatical rule

4.5 Target Polarity Classification

As we have estimated word-level sentimental signal in previous sections, we take the average score of all words' sentiment scores within the target as the target's lexicon sentimental score. Meanwhile, we take the sentimental signal of the adjacent adjective of the target as the target's modifier sentimental score.

Finally we take the signal with maximum absolute value between lexicon sentimental score and modifier sentimental score. If the signal is greater than upper threshold or less than lower threshold, the target is classified as positive or negative respectively. Otherwise, the target is classified as neutral (Other). In this work, upper and lower threshold are both assigned to 0.

5 Experiment

We conduct our grammatical phrase-level target extraction approach on the Chinese microblog sentiment analysis contest held by the 2nd Conference on Natural Language Processing & Chinese Computing(NLP&CC 2013).

5.1 Dataset

The dataset is provided by contest holder and are crawled from Tencent Weibo. It consists of two parts: one is pre-released smaller dataset with only 2 topics, 200 Weibo messages and 473 sentences, which is employed as training set; the other is contest dataset with 10 topics, 12382 Weibo messages, 27499 sentences and 1000 randomly sampled annotated messages.

In training dataset, all sentences are annotated whether they are opinionated and targets are extracted from the sentences. In contest dataset, randomly sampled messages are annotated whether they are opinionated and only these annotated sentences are required to be processed. Notice that only opinionated sentences contain opinion targets.

5.2 Ground Truth and Evaluation Metrics

The ground truth targets are provided by contest holder. A target is represented as a tuple including target's content, offset of the target within microblog message and target's polarity.

There are two types of evaluation metrics: Strict Evaluation and Lenient Evaluation. Precision, Recall and F_1-measure are employed to evaluate approach performance.

In strict evaluation, extracted targets are correct if and only if their offset, textual content, polarity exactly equal to annotations.

In lenient evaluation, extracted targets do not have to exactly equal to annotated targets. We adopt coverage to estimate the true positive of the precision and recall.

Given any extracted target r and any annotated target r', the coverage $C(r, r')$ between them is defined as:

$$C(r, r') = \begin{cases} \frac{|s \cap s'|}{|s'|} & s \cap s' \neq \emptyset \\ 0 & otherwise \end{cases} \quad (1)$$

where s and s' is the corresponding range to r and r' respectively. Afterward, the overall coverage between proposed target set and annotated target set is the summation of all extracted target and annotated target pairs' coverage. Taking this coverage as the true positive, we get the lenient precision and recall.

5.3 Evaluation Results

There are two ways to measure evaluation results. Micro-average result is calculated on overall contest dataset directly, while macro-average result is the average performance calculated topically. Notice that micro-average has the bias on performance of topics with more targets.

Table 1. Evaluation results of the grammatical phrase-level target extraction

	Micro-average			Macro-average		
Evaluation	Precision	Recall	F_1-measure	Precision	Recall	F_1-measure
Strict	0.202	0.236	0.218	0.200	0.226	0.210
Lenient	0.288	0.340	0.312	0.283	0.323	0.299

Table 1 presents evaluation results on both strict and lenient criteria. The evaluation results are balanced between precision and recall. Furthermore, recall is better than precision in lenient evaluations. This is because we prefer more targets while extracting candidates. Notice that we do not consider pronouns and implicit targets, which hurt the performance on recall.

Our approach does not make significant improvement to lenient evaluation. It is mainly caused by the extraction strategy that we concentrate on the accurate boundaries of the targets, which prunes entirely error cases and partially correct cases equally.

6 Conclusion

We investigate the Chinese microblog messages and propose a grammatical phrase-level target extraction approach. The approach is based on the assumption that a target is the subgraph of the sentence's dependency tree. We figure out the target subgraphs directly by conducting classification of the **TAR-REL**s that connecting the targets. After recovering misclassified relations and pruning with some empirically summarized rules, we finally achieve targets from the microblog corpus.

The evaluation results indicate that our approach is feasible and effective on phrase-level extraction from Chinese microblog corpus. In future, we'd like to continue our research on investigating the extraction with structured information of sentences, attempting to involve contextual grammatical relations, adopting probabilistic model and more effective features for phrase-level target extraction.

Acknowledgments. The NExT Search Centre is supported by the Singapore National Research Foundation and Interactive Digital Media R&D Program Office, MDA under research grant (WBS: R-252-300-001-490).

References

1. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion Word Expansion and Target Extraction through Double Propagation. Computational Linguistics 37(1), 9–27 (2011)
2. Brody, S., Elhadad, N.: An Unsupervised Aspect-sentiment Model for Online Reviews. In: Brody, S., Elhadad, N. (eds.) HLT 2010, Stroudsburg, PA, USA, pp. 804–812 (2010)
3. Sayeed, A.B., Boyd-Graber, J., Rusk, B., Weinberg, A.: Grammatical Structures for Word-level Sentiment Detection. In: NAACL HLT 2012, Stroudsburg, PA, USA, pp. 667–676 (2012)
4. Mukherjee, A., Liu, B.: Aspect Extraction through Semi-supervised Modeling. In: ACL 2012, Stroudsburg, PA, USA, pp. 339–348 (2012)
5. Lin, C., He, Y.: Joint Sentiment/Topic Model for Sentiment Analysis. In: CIKM 2009, New York, NY, USA, pp. 375–384 (2009)
6. Liu, K., Xu, L., Zhao, J.: Opinion Target Extraction Using Word-based Translation Model. In: EMNLP-CoNLL 2012, Stroudsburg, PA, USA, pp. 1346–1356 (2012)
7. Ku, L.W., Huang, T.H., Chen, H.H.: Using Morphological and Syntactic Structures for Chinese Opinion Analysis. In: EMNLP 2009, Stroudsburg, PA, USA, pp. 1260–1269 (2009)
8. Zhou, X., Wan, X., Xiao, J.: Cross-language Opinion Target Extraction in Review Texts, pp. 1200–1205 (December 2012)
9. Cui, A., Zhang, H., Liu, Y., Zhang, M., Ma, S.: Lexicon-based Sentiment Analysis on Topical Chinese Microblog Messages. In: Semantic Web and Web Science, pp. 333–344. Springer Proceedings in Complexity (2013)
10. Domingos, P.: Metacost: A General Method for Making Classifiers Cost-sensitive. In: KDD 1999, New York, NY, USA, pp. 155–164 (1999)