Bootstrapping Both Product Features and Opinion Words from Chinese Customer Reviews with Cross-Inducing[1]

Bo Wang
Institute of Computational Linguistics
Peking University
Beijing, 100871, China
wangbo@pku.edu.cn

Houfeng Wang
Institute of Computational Linguistics
Peking University
Beijing, 100871, China
wanghf@pku.edu.cn

Abstract

We consider the problem of identifying product features and opinion words in a unified process from Chinese customer reviews when only a much small seed set of opinion words is available. In particular, we consider a problem setting motivated by the task of identifying product features with opinion words and learning opinion words through features alternately and iteratively. In customer reviews, opinion words usually have a close relationship with product features, and the association between them is measured by a revised formula of mutual information in this paper. A bootstrapping iterative learning strategy is proposed to alternately both of them. A linguistic rule is adopted to identify low-frequent features and opinion words. Furthermore, a mapping function from opinion words to features is proposed to identify implicit features in sentence. Empirical results on three kinds of product reviews indicate the effectiveness of our method.

1 Introduction

With the rapid expansion of network application, more and more customer reviews are available online, which are beneficial for product merchants to track the viewpoint of old customers and to assist potential customers to purchase products. However, it's time-consuming to read all reviews in person. As a result, it's significant to mine customer reviews automatically and to provide users with opinion summary.

In reality, product features and opinion words play the most important role in mining opinions of customers. One customer review on some cell phone is given as follows:

(a) "外型漂亮，屏幕大，拍照效果好。"(The **appearance** is beautiful, the **screen** is big and the **photo effect** is OK.)

Product features are usually nouns such as "外型" (appearance) and "屏幕" (screen) or noun phrases such as "拍照效果" (photo effect) expressing which attributes the customers are mostly concerned. Opinion words (opword is short for "opinion word") are generally adjectives used to express opinions of customers such as "漂亮" (beautiful), "大" (big) and "好" (well). As the core part of an opinion mining system, this paper is concentrated on identifying both product features and opinion words in Chinese customer reviews.

There is much work on feature extraction and opinion word identification. Hu and Liu (2004) makes use of association rule mining (Agrawal and Srikant, 1994) to extract frequent features, the surrounding adjectives of any extracted feature are considered as opinion words. Popescu and Etzioni (2005) has utilized statistic-based point-wise mutual information (PMI) to extract product features. Based on the association of opinion words with product features, they take the advantage of the syntactic dependencies computed by the MINIPAR parser (Lin, 1998) to identify opinion words. Tur-

---

ney (2002) applied a specific unsupervised learning technique based on the mutual in-formation between document phrases and two seed words "excellent" and "poor".

Nevertheless, in previous work, identifying product features and opinion words are always considered two separate tasks. Actually, most product features are modified by the surrounding opinion words in customer reviews, thus they are highly context dependent on each other, which is referred to as context-dependency property henceforth. With the co-occurrence characteristic, identifying product features and opinion words could be combined into a unified process. In particular, it is helpful to identify product features by using identified opinion words and vice versa. That implies that such two subtasks can be carried out alternately in a unified process. Since identifying product features are induced by opinion words and vice versa, this is called cross-inducing.

As the most important part of a feature-based opinion summary system, this paper focuses on learning product features and opinion words from Chinese customer reviews. Two sub-tasks are involved as follows:

Identifying features and opinion words: Resorting to context-dependency property, a bootstrapping iterative learning strategy is proposed to identify both of them alternately.

Identifying implicit features: Implicit features occur frequently in customer reviews. An implicit feature is defined as a feature that does not appear in an opinion sentence. The association between features and opinion words calculated with the revised mutual information is used to identify implicit features.

This paper is sketched as follows: Section 2 describes the approach in detail; Experiment in section 3 indicates the effectiveness of our approach. Section 4 presents related work and section 5 concludes and presents the future work.

2 The Approach

Figure 1 illustrates the framework of an opinion summary framework, the principal parts related to this paper are shown in bold. The first phase "identifying features and opinion words", works iteratively to identify features with the opinion words identified and learn opinion words through the product features identified alternately. Then,

one linguistic rule is used to identify low-frequent features and opinion words. After that, a mapping function is designed to identify implicit features.
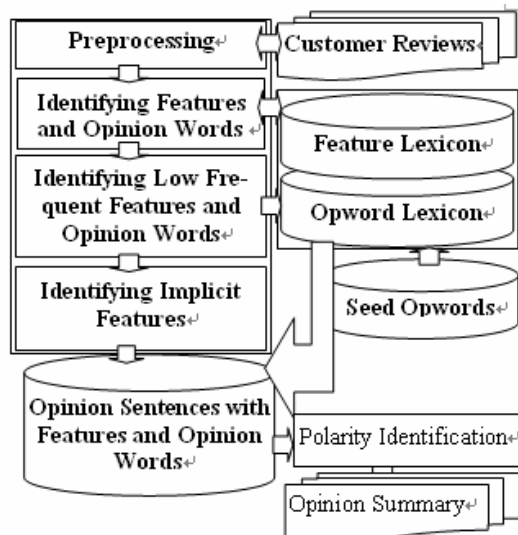
Figure 1. The framework of an opinion summary system

2.1 Iterative Learning Strategy

Product features and opinion words are highly context-dependent on each other in customer reviews, i.e., the feature "机身" (body) for digital camera often co-occur with some opinion words such as "大" (big) or "小巧" (delicate) while the feature "性价比" (the proportion of performance to price) often co-occurs with the opinion word "高" (high).

Product features can be identified resorting to the surrounding opinion words identified before and vice versa. A bootstrapping method that works iteratively is proposed in algorithm 1.

Algorithm 1 works as follows: given the seed opinion words and all the reviews, all noun phrases (noun phrases in the form "noun+") form *CandFeaLex* (the set of feature candidates) and all adjectives compose of *CandOpLex* (the set of the candidates of opinion words). The sets *ResFeaLex* and *ResOpLex* are used to store final features and opinion words. Initially, *ResFeaLex* is set empty while *ResOpLex* is composed of all the seed opinion words. At each iterative step, each feature *candidate* in *CandFeaLex* is scored by its context-dependent association with each *opword* in *ResOpLex*, the candidate whose score is above the pre-specified threshold *Threshold_{feature}* is added to *Res*

Algorithm 1. Bootstrap learning product features and opinion words with cross-inducing

Bootstrap-Learning (ReviewData, SeedOpLex, Threshold$_{feature}$, Threshold$_{opword}$)
1 Parse(ReviewData);
2 ResFeaLex = {}, ResOpLex = SeedOpLex;
3 CandFeaLex = all noun phrases in ReviewData;
4 CandOpLex = all adjectives in ReviewData;
5 **while** (CandFeaLex≠{} && CanOpLex≠{})
6   **do for** each candfea∈CandFeaLex
7   **do for** each opword∈ResOpLex
8   **do** calculate RMI(candfea,opword) with ReviewData;
9   score(canfea)=$\Sigma_{opword\in ResOpLex}$RMI(candfea,opword)/|ResOpLex|;
10   sort CandFeaLex by score;
11   **for** each candfea∈CandFeaLex
12   **do if** (score(candfea)> Threshold$_{feature}$)
13   **then** ResFeaLex=ResFeaLex+{candfea};
14   CanFeaLex=CandFeaLex – {candfea};
15   **for** each candop∈CandOpLex
16   **do for** each feature∈ResFeaLex
17   **do** calculate RMI(candop,feature) with D;
18   score(candop)=$\Sigma_{feature\in ResFeaLex}$RMI(feature,candop)/|ResFeaLex| ;
19   sort CandOpLex by score;
20   **for** each candop∈CandOpLex
21   **do if** (score (candop)>Threshold$_{opword}$)
22   **then** ResOpLex=ResOpLex+{candop };
23   CanOpLex=CandOpLex – {candop};
24   **if** (neither candfea and candop is learned) **then** break;
25 **return** ResFeaLex, ResOpLex;

FeaLex and subtracted from *CandFeaLex*. Similarly, opinion words are processed in this way, but the scores are related to features in *ResFeaLex*. The iterative process continues until neither *ResFeaLex* nor *ResOpLex* is altered. Any feature candidate and opinion word candidate, whose relative distance in sentence is less than or equal to the specified window size *Minimum-Offset*, are regarded to co-occur with each other. The association between them is calculated by the revised mutual information denoted by *RMI*, which will be described in detail in the following section and employed to identify implicit features in sentences.

2.2 Revised Mutual Information

In customer reviews, features and opinion words usually co-occur frequently, features are usually modified by the surrounding opinion words. If the absolute value of the relative distance in a sentence for a feature and an opinion word is less than *Minimum-Offset*, they are considered context-dependent.

Many methods have been proposed to measure the co-occurrence relation between two words such as χ^2 (Church and Mercer,1993) , mutual information (Church and Hanks, 1989; Pantel and Lin, 2002), t-test (Church and Hanks, 1989), and log-likelihood (Dunning,1993). In this paper a revised formula of mutual information is used to measure the association since mutual information of a low-frequency word pair tends to be very high.

Table 1 gives the contingency table for two words or phrases w_1 and w_2, where *A* is the number of reviews where w_1 and w_2 co-occur; *B* indicates the number of reviews where w_1 occurs but does not co-occur with w_2; *C* denotes the number of reviews where w_2 occurs but does not co-occur with w_1; *D* is number of reviews where neither w_1 nor w_2 occurs; $N = A + B + C + D$.

With the table, the revised formula of mutual information is designed to calculate the association of w_1 with w_2 as formula (1).

	w_2	$\sim w_2$
w_1	A	B
$\sim w_1$	C	D

Table 1: Contingency table

$$RMI(w_1, w_2) = freq(w_1, w_2) \times \log \frac{p(w_1, w_2)}{p(w_1) \bullet p(w_2)}$$

$$= A \times \log \frac{N \times A}{(A+B) \times (A+C)} \quad (1)$$

2.3 Identifying Low-Frequent Features and Opinion Words

In Chinese reviews, one linguistic rule "noun+ adverb* adjective+" occurs frequently and most of the instances of the rule are used to express positive or negative opinions on some features, i.e., "机身/noun 比较/adverb 小巧/adjective" (The body is rather delicate) , where each Chinese word and its part-of-speech is separated by the symbol "/".

Intuitively, this linguistic rule can be used to improve the output of the iterative learning. For each instance of the rule, if "noun+" exists in *ResFeaLex*, the "adjective" part would be added to *ResOpLex*, and if "adjective+" exists in *ResOpLex*, the noun phrase "noun+" part will be added to *ResFeaLex*. After that, most low-frequent features and opinion words will be recognized.

2.4 Identifying Implicit Features

The context-dependency property indicates the context association between product features and opinion words. As a result, with the revised mutual information, the implicit features can be deduced from opinion words. A mapping function *f: opword*→*feature* is used to deduce the mapping feature for opword , where *f(opword)* is defined as the feature with the largest association with opinion word.

If an opinion sentence contains opinion words, but it does not have any explicit features, the mapping function *f: opword*→*feature* is employed to generate the implicit feature for each opinion word and the feature is considered as an implicit feature in the opinion sentence. Two instances are given in (b) and (c), where the implicit features are inserted in suitable positions and they are separated in parentheses. Since *f* ("漂亮" (beautiful)) = "外观" (appearance) and *f* ("时尚" (fashionable)) = "外观" (appearance), "外观" (appearance) is an implicit feature in (b). Similarly, the implicit features

in (c) are "性能" (performance) and "图像" (picture).

(b) (外观)漂亮而且(外观)时尚。It's (**appearance**) <u>beautiful</u> and (**appearance**) <u>fashionable</u>.

(c) (性能)很 <u>稳定</u>，而且(图像)很 <u>清晰</u>。It's (**performance**) very <u>stable</u> and (**picture**) very clear.

3 Experiment

3.1 Data Collection

We have gathered customer reviews of three kinds of electronic products from http://it168.com: digital camera, cell-phone and tablet. The first 300 reviews for each kind of them are downloaded. One annotator was asked to label each sentence with product features (including implicit features) and opinion words. The annotation set for features and opinion words are shown in table 2.

Product Name	No. of Features	No. of Opinion Words
digital camera	135	97
cell-phone	155	125
tablet	96	83

Table 2. Annotation set for product features and opinion words

Unlike English, Chinese are not separated by any symbol. Therefore, the reviews are tokenized and tagged with part-of-speech by a tool ICTCLAS[2].One example of the output of this tool is as (d).

(d) 开机/n 速度/n 还/d 满/d 快/a ，/w 镜头/n 保护盖/n 拉开/v 就/d 可以/v 进入/v 拍摄/n 状态/n ，/w 模式/n 选择/vn 切换/vn 也/d 很/d 方便/a 。/w

The seed opinion words employed in the iterative learning are: "清晰" (clear), "快" (quick), "白" (white), "差劲" (weak). "好" (good), "不错" (good), "高" (high), "小" (little), "多" (many), "长" (long). Empirically, *Threshold_{feature}* and *Threshold_{opword}* in Algorithm 1 is set to 0.2, *Minimum-Offset* is set to 4.

Product Name	On Set			On Sentence		
	Precision	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
digital camera	64.03%	45.92%	53.49%	46.62%	65.72%	54.55%
cell-phone	54.43%	43.87%	48.58%	34.17%	55.15%	42.19%
tablet	51.45%	59.38%	55.13%	41.39%	60.21%	49.06%
average	**56.64%**	**49.72%**	**52.40%**	**40.73%**	**60.36%**	**48.60%**

Table 3. Evaluation of apriori algorithm

Type	Product Name	On Set			On Sentence		
		Precision	**Recall**	**F-Score**	**Precision**	**Recall**	**F-score**
feature	**digital camera**	73.57%	54.81%	62.82%	55.80%	68.69%	61.58%
		78.20%	73.33%	75.69%	54.71%	70.80%	63.49%
	cell-phone	80.92%	45.81%	58.50%	47.31%	58.59%	52.35%
		82.30%	66.46%	73.53%	49.22%	61.63%	54.73%
	tablet	72.73%	57.29%	64.09%	49.79%	61.03%	54.84%
		77.99%	73.96%	75.92%	52.54%	64.43%	57.88%
	average	**75.74%**	**52.64%**	**61.80%**	**50.97%**	**62.77%**	**56.26%**
		79.50%	**71.25%**	**75.05%**	**52.16%**	**65.62%**	**58.70%**
opword	**digital camera**	89.02%	38.02%	53.28%	72.35%	50.24%	59.30%
		87.31%	60.94%	71.78%	69.40%	85.28%	76.53%
	cell-phone	87.95%	30.80%	45.63%	66.44%	42.84%	52.09%
		88.49%	51.90%	65.43%	63.14%	79.51%	70.39%
	tablet	77.94%	30.64%	43.98%	61.30%	42.69%	50.34%
		80.73%	50.87%	62.41%	63.92%	81.02%	71.46%
	average	**84.97%**	**33.15%**	**47.63%**	**66.70%**	**45.26%**	**53.91%**
		85.51%	**54.57%**	**66.54%**	**65.49%**	**81.94%**	**72.79%**

Table 4. Evaluation of iterative learning (the upper) and the combination of iterative learning and the linguistic rule (the lower).

3.2 Evaluation Measurement

As Hu and Liu (2004), the features mined form the result set while the features in the manually annotated corpus construct the answer set. With the two sets, precision, recall and f-score are used to evaluate the experiment result on set level.

In our work, the evaluation is also conducted on sentence for three factors: Firstly, each feature or opinion word may occur many times in reviews but it just occurs once in the corresponding answer set; Secondly, implicit features should be evaluated on sentence; Besides, to generate an opinion summary, the features and the opinion words should be identified for each opinion sentence.

On sentence, the features and opinion words identified for each opinion sentence are compared with the annotation result in the corresponding sentence. Precision, recall and f-score are also used to measure the performance.

3.3 Evaluation

Hu and Liu (2004) have adopted associate rule mining to mine opinion features from customer reviews in English. Since the original corpus and source code is not available for us, in order to make comparison with theirs, we have re-implemented their algorithm, which is denoted as apriori method as follows. To be pointed out is that, the two pruning techniques proposed in Hu and Liu (2004): compactness pruning and redundancy pruning, were included in our experiment. The evaluation on our test data is listed in table 3. The row indexed by average denotes the average performance of the corresponding column and each entry in it is bold.

Table 4 shows our testing result on the same data, the upper value in each entry presents the result for iterative learning strategy while the lower values denote that for the combination of iterative learning and the linguistic rule. The average row

shows the average performance for the corresponding columns and each entry in the row is shown in bold.

On feature, the average precision, recall and f-score on set or sentence increase according to the order apriori < iterative < ite+rule, where apriori indicates Hu and Liu's method, iterative represents iterative strategy and iterative+rule denotes the combination of iterative strategy and the linguistic rule. The increase range from apriori to iterative+rule of f-score on set gets to 22.65% while on sentence it exceeds 10%. The main reason for the poor performance on set for apriori is that many common words such as "电脑" (computer), "中国" (China) and "时间" (time of use) with high frequency are extracted as features. Moreover, the poor performance on sentence for apriori method is due to that it can't identify implicit features. Furthermore, the increase in f-score from iterative to ite+rule on set and on sentence shows the performance can be enhanced by the linguistic rule.

Table 4 also shows that the performance in learning opinion words has been improved after the linguistic rule has been used. On set, the average precision increases from 84.97% to 85.51% while the average recall from 33.15% to 54.57%. Accordingly, the average f-score increase significantly by about 18.91%.

On sentence, although there is a slow decrease in the average precision, there is a dramatic increase in the average recall, thus the average f-score has increased from 53.91% to 72.79%. Furthermore, the best f-score (66.54%) on set and the best f-score (72.79%) on sentence indicate the effectiveness of *ite+rule* on identifying opinion words.

4 Related Work

Our work is much related to Hu's system (Hu and Liu,2004), in which association rule mining is used to extract frequent review noun phrase as features. After that, two pruning techniques: compactness pruning and redundancy pruning, are utilized. Frequent features are used to find potential opinion words (adjectives) and WordNet synonyms/antonyms in conjunction with a set of seed words are used in order to find actual opinion words. Finally, opinion words are used to extract associated infrequent features. The system only extracts explicit features. Our work differs from

hers at two aspects: (1) their method can't identify implicit features which occur frequently in opinion sentences; (2) Product features and opinion words are identified on two separate steps in Hu's system but they are learned in a unified process here and induced by each other in this paper.

Popescu and Etzioni (2005) has used web-based point-wise mutual information (PMI) to extract product features and use the identified features to identify potential opinion phrases with co-occurrence association. They take advantage of the syntactic dependencies computed by the MINIPAR parser. If an explicit feature is found in a sentence, 10 extraction rules are applied to find the heads of potential opinion phrases. Each head word together with its modifier is returned as a potential opinion phrase. Our work is different from theirs on two aspects: (1) Product features and opinion words are identified separately but they are learned simultaneously and are boosted by each other here. (2) They have utilized a syntactic parser MINIPAR, but there's no syntactic parser available in Chinese, thus the requirement of our algorithm is only a small seed opinion word lexicon. Although co-occurrence association is used to derive opinion words from explicit features in their work, the way how co-occurrence association is represented is different. Besides, the two sub-tasks are boosted by each other in this paper.

On identifying opinion words, Morinaga et al (2002)has utilized information gain to extract classification features with a supervised method; Hatzivassiloglou and Wiebe (1997) used textual junctions such as "fair and legitimate" or "simplistic but well-received" to separate similarity- and oppositely-connoted words; Other methods are present in (Riloff et al, 2003; Riloff and Wiebe, 2003; Gamon and Aue, 2005; Wilson et al, 2006) The principal difference from previous work is that, they have considered extracting opinion words as a separate work but we have combined identifying features and opinion words in a unified process. Besides, the opinion words are identified for sentences but in their work they are identified for reviews.

5 Conclusion

In this paper, identifying product features and opinion words are induced by each other and are combined in a unified process. An iterative learn-

ing strategy based on context-dependence property is proposed to learn product features and opinion words alternately, where the final feature lexicon and opinion word lexicon are identified with very few knowledge (only ten seed opinion words) and augmented by each other alternately. A revised formula of mutual information is used to calculate the association between each feature and opinion word. A linguistic rule is utilized to recall low-frequent features and opinion words. Besides, a mapping function is designed to identify implicit features in sentence. In addition to evaluating the result on set, the experiment is evaluated on sentence. Empirical result indicates that the performance of iterative learning strategy is better than apriori method and that features and opinion words can be identified with cross-inducing effectively. Furthermore, the evaluation on sentence shows the effectiveness in identifying implicit features.

In future, we will learn the semantic orientation of each opinion word, calculate the polarity of each subjective sentence, and then construct a feature-based summary system.

References

Ana Maria Popescu and Oren Etzioni. 2005. *Extracting Product Features and Opinions from Reviews*. Proceedings of HLT-EMNLP (2005)

De-Kang Lin. 1998. *Dependency-Based Evaluation of MINIPAR*. In:Proceedings of the Workshop on the Evaluation of Parsing Systems, Granada, Spain, 1998, 298～312

Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. *Learning Subjective Nouns Using Extraction Pattern Bootstrapping*. Seventh Conference on Natural Language Learning (CoNLL-03). ACL SIGNLL. Pages 25-32.

Ellen Riloff and Janyce Wiebe. 2003. *Learning Extraction Patterns for Subjective Expressions*. Conference on Empirical Methods in Natural Language Processing (EMNLP-03). ACL SIGDAT. 2003, 105-112.

Kenneth Ward Church and Robert L. Mercer. 1993. *Introduction to the special issue on computational linguistics using large corpora*. Computational Linguistics 19:1-24

Kenneth Ward Church and Patrick Hanks. 1989. *Word Association Norms, Mutual Information and Lexicography*. Proceedings of the 26th Annual Conference of the Association for Computational Linguistics(1989).

Michael Gamon and Anthony Aue. 2005. *Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms*. In :ACL 2005 Workshop on Feature Engineering,2005.

Minqing Hu and Bing Liu. 2004. *Mining Opinion Features in Customer Reviews*. Proceedings of Nineteeth National Conference on Artificial Intellgience (AAAI-2004), San Jose, USA, July 2004.

Patrick Pantel and Dekang Lin. 2002. *Document Clustering with Committees*. In Proceedings of ACM Conference on Research and Development in Information Retrieval (SIGIR-02). pp. 199-206. Tampere, Finland.

Peter D. Turney. 2002. *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*. ACL 2002: 417-424

Rakesh Agrawal and Ramakrishan Srikant. 1994. *Fast algorithm for mining association rules*. VLDB'94, 1994.

Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, and Toshikazu Fukushima. 2002. *Mining Product Reputations on the WEB*, Proceedings of 8th ACM SIGKDD International Conference on Knowledge. Discover and Data Mining, (2002) 341-349

Ted Dunning. 1993. *Accurate methods for the statistics of surprise and coincidence*. Computational Linguistics 19:61-74

Theresa Wilson , Janyce Wiebe, and Rebecca Hwa. 2006. *Recognizing strong and weak opinion clauses*. Computational Intelligence 22 (2): 73-99.