Incorporating Language Patterns and Domain Knowledge into Feature-opinion Extraction

Erqiang Zhou[1,2], Xi Luo[1], and Zhiguang Qin[1]

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China, No.2006, Xiyuan Ave, Chengdu 611731, Sichuan, P.R.China
zhoueq@uestc.edu.cn, xiluouestc@gmail.com, qinzg@uestc.edu.cn
[2] Guangdong Key Laboratory of Popular High Performance Computers, Shenzhen Key Laboratory of Service Computing and Applications, No. 3688, Nanhai Ave, Shenzhen 518060, Guangdong, P.R.China

Abstract. We present a hybrid method for aspect-based sentiment analysis of Chinese restaurant reviews. Two main components are employed so as to extract feature-opinion pairs in the proposed method: domain independent language patterns found in Chinese and a lexical base built for restaurant reviews. The language patterns focus on the general knowledge which is implicit contained in Chinese, thus can be used directly by other domains without any modification. The lexical base, on the other hand, targets for particular characteristics of a given domain and acts as a plug-in part in our prototype system, thus does not affect the portability when applying the proposed approach in practice. Empirical evaluation shows that our method performs well and it can gain a progressive result when each component takes into effective.

Keywords: Opinion Mining, Sentiment Analysis, Restaurant Review

1 Introduction

As a kind of user generated content, online reviews become more important whenever for the consumers and the sellers of e-commerce or for a service provider and its user. Based on the reviews of online products or services, consumers or users can evaluate the aspects with which they are concerned in detail in addition to the overall scores; sellers or service providers can know which aspects need to be improved for attracting more clients.

Realizing the importance of extracting feature-opinions for practical requirements, much efforts [2,5,10,11,16] have been devotes into this field since the pioneer work done by Hu and Liu [4]. However, automatic acquiring feature opinion pairs from texts is still a challenging problem.

In fact, in addition to a few researches [3,12,14,15], most of existing work in this research field focused on English reviews and has not attempted to handle texts written in other languages. Furthermore, to the best of our knowledge, the corpus of Chinese restaurant reviews, which is the target of our work, has not been fully investigated in the field of sentiment analysis.

Although the domain of restaurant reviews share main characteristics with other domains for feature-opinion extraction, it has particular distinguishing points. For example, some Chinese dish names not only tell the ingredients of certain dishes, but also imply how they are cooked or which regional style they may have, such as '蒜苗回锅肉' (garlic sprouts with twice-cooked pork), '水煮牛肉' (water cooked beef), '重庆辣子鸡' (Chongqing spicy deep-fried chicken), etc. This fact leads to the difficulty of identifying such kind of entity names. Another fact which distinguishes the corpus of restaurant from product or movies reviews is that the quality of dishes and services provided by restaurants is time sensitive. In other words, the features of a product or a movie does not change once it is put on the market, while foods and services are hand work thus a recent opinion is more important than the viewpoint of a long time ago.

Based on all these facts, a fully investigation of Chinese restaurant reviews has both practical and scientific values. In this paper, we study the problem of feature-opinion extraction from online reviews of Chinese restaurants. We address the problem by employing following two steps. Firstly, we build a knowledge base for the domain of restaurant reviews. The base consists of two parts: 1) the feature words and their potential aspects, which can be classified into following five categories: taste, environment, service, price and overall evaluations; 2) the possible opinion words for remarking restaurant features. Secondly, we develop a rule sets for capturing the general patterns that are used when people expressing their opinions.

Our experiments are conducted on a real life dataset, which consists of 5,251 restaurants and each restaurant has 26 reviews in average. The overall extracted review texts is 125M bytes, we randomly selected 5 restaurants and manually labelled the feature-opinion pairs of all their reviews for testing. It should be noted that our test standard is whether the automatically extracted feature-opinion pairs match the manually labelled ones. Experiments show that the proposed method can reach a reasonable and acceptable result: the average Precision, Recall and F-score is 0.55, 0.83 and 0.66 respectively.

2 Related Work

There are two main tasks involved in the research of extracting feature-opinion: feature identification and opinion extraction. Hu and Liu [4] utilized frequent item to identify product features and considered adjectives which modifies feature words as the opinion words. Zhuang *et al.* [16] integrated WordNet, statistical analysis and movie knowledge into a multi-knowledge base so as to extract feature opinion pairs and summarize movie reviews. Somprasertsri and Lalitrojwong [11] extracted product feature and opinion by considering syntactic and semantic information, that is by applying dependency relations and ontological knowledge with probabilistic based model.

It is worth notice more and more ontologies or knowledge bases are used for the tasks of opinion mining recently. For example, O'Leary [8] utilized knowledge gained from tags on blogs for blog mining and showed that there is a need for domain specific terms so as to capture a richer understanding of mood of a blog. Peñalver-Martínez *et al.* [9] proposed a method for improving feature-based opinion mining by employing ontologies in the selection of features. Freitas and Vieira [3] identified the polarity of Portuguese reviews according to features described in domain (movie and hotel) on-

tologies. Ittoo and Bouma [5] showed that open-domain corpora, like Wikipedia, can be exploited as knowledge bases for extracting causal relations from domain-specific texts so as to overcome data sparsity issues. Yin *et al.* [14] presented a linguistic model, which is based on a automatically constructed ontology, for identifying the basic appraisal expression in Chinese product reviews.

In addition to methods of using ontologies, several kinds of rules are designed in the field of opinion mining. Zhai *et al.* [15] presented an unsupervised approach to identify people's opinions on topics and their aspects (the so called evaluative opinions) by proposing several rules. Jiao *et al.* [6] employed Chinese dependency grammar to set several rules for extracting candidate feature-opinion word pairs.

Statistical topic models are also often employed for tackling opinion mining tasks. Brody and Elhadad [1] applied the LDA model to the unsupervised features extract: they treated one sentence as a document and adjusted the parameters of LDA so as to make it suit the sentence-document model. After finding feature words in some topic, Brody and Elhadad treated the nearest adjective words as the opinion words. Li *et al.* [7] introduced a dependency-sentiment-LDA, which relaxes the sentiment independent assumption and is an extension of a joint sentiment and topic model, Sentiment-LDA, which is also proposed by the authors. Xu *et al.* [13] propose a generative topic model, the Joint Aspect/Sentiment (JAS) model, to jointly extract aspects and aspect-dependent sentiment lexicons from online customer reviews.

However, according to our experiments, the LDA-based topic models fail to find the feature words from our Chinese restaurant reviews. One possible reason may be the so called data sparsity issue: the dish names and the recipe of foods are candidates words of restaurant features, but they usually varies in restaurant reviews. We handle this problem by semi-automatically building an lexical base for the domain of Chinese restaurant reviews.

3 Opinion-Feature Extraction

Figure 1 gives an overview of the proposed method. Three main steps involves in the identification of feature-opinion pairs: 1) pre-processing; 2) lexical base building and; 3) feature-opinion pair extraction. As to the first step, we use ICTCLAS[3] to POS-tag all reviews, and then each review is split into sentences for further processing. The second step and the third step will be explained in Section 3.1 and Section 3.2 respectively.

3.1 Lexical Base Building

The lexical base consists three parts: aspect words, dish names, words that can match with aspect words and dish names. Accordingly, the building process can be decomposed into three tasks: 1) to collect dish names; 2) to identify aspect words and; 3) to extract the matching words by designing rules.

As to the first task, a crawling spider is developed to obtain candidate dish names from the website of *www.dianping.com* since it allows users to write their own recommending dishes for each restaurant. However, some irrelevant phrases or opinions (such

[3] `http://www.ictclas.org/ictclas_download.aspx`

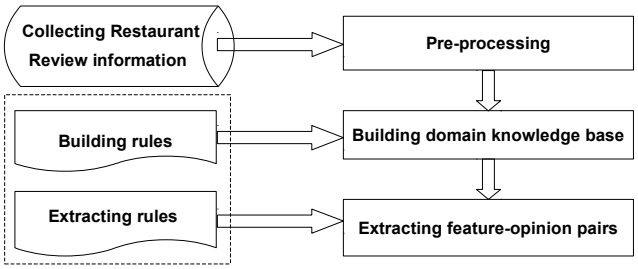

Fig. 1. Overview of the proposed method

as: 都好吃/delicious, 不喜欢/dislike, etc.) were also given by users, thus a cleanup step is needed for removing such kind of noises. In our research, we treat the word that are rarely used in dish name as seed words for identifying the noise phrases. Finally we collect 6433 dish names altogether.

As to the second task, we define a noun as a *frequent-noun* if its *sentence appear ratio*, which is defined as a division between the number of sentences which has the *frequent-noun* and the number of all review sentences, is great than 1%. The generated *frequent-nouns* are treated as candidate aspect words in this paper.

To perform the third task, following two rules are used:

Rule 1: {(F, N)—(A, V)}. This rule means that if a noun N or a dish name F appears near a word which was labeled as an adjective A or a verb V in one *short sentence*[4], the noun term N or dish name F and adjective A or verb V are considered as a candidate pair of matching words.

Rule 2: {F—N}. If a food term F appears with a noun term N in a short sentence, F and N are considered as a candidate pair of matching words. The noun N has certain chance of being an aspect of the food.

We define the number when term N or dish name F and adjective A or verb V co-occurs as the *Matched Degree* (MD). The bigger an MD is, the more chance of being a pair the term N (or F) and the term A (or V) have. A global threshold value of MD is used in our research.

The rules are quite intuitive, and they are mainly used for finding the matching words. For example, in POS-tagged sentence '环境/n (environment) 和/c (and) 味道/n (taste) 很/d (very) 不错/a (nice)', '环境' (environment) and '味道' (taste) are both noun terms, '不错' (nice) is an adjective word which co-occurs with '环境' (environment) and '味道' (taste). So the MD of '环境(environment)—不错(nice)' and '味道(taste)—不错(nice)' will all increase one number each.

3.2 Extracting Feature-opinion Pairs

The extraction of feature-opinion pairs is performed by exploring following rules:

[4] Short sentences are obtained by splitting a natural sentence by punctuations.

Rule 3: if a noun N appears with and an adjective A (or a verb V), and there is no other[5] nouns, adjectives or verbs between N and A, then we make use of the MD of matching pair to decide whether N and A can be treated as a feature-opinion pair.

Rule 4: if a dish name F and an adjective A (or a verb V) appear in a *short sentence*, decide whether F and A can be treated as a feature-opinion pair according to the global threshold value of MD.

Rule 5: if a frequent-noun follows a dish name and an adjective comes along with the frequent-word, the frequent-noun might be an attribute of the dish. Then we check the MD of pair (dish name, frequent-noun), if the MD satisfies the global threshold, we extract the pair (dish name, frequent-noun) as a feature-opinion pair; otherwise the following two pairs (dish name, adjective) and (frequent-noun, adjective) should be checked.

For example, in POS-tagged sentence '回锅肉/ny (Double cooked pork) 和/c (and) 环境/n (environment) 很/d (very) 好/a (good)', '回锅肉(Double cooked pork)' is a dish name since it has a tag 'ny', and '环境(environment)' is a frequent-noun. So we first check whether the MD of pair (回锅肉(Double cooked pork), 环境(environment)) satisfy the global threshold value, if not, the two pairs (回锅肉(Double cooked pork, 好(good)) and (环境(environment), 好(good)) should be check. In our lexical base, these two pairs both satisfy the global threshold, therefore, they are treated as feature-opinion pairs.

4 Empirical Evaluation

The dataset that we used for evaluating the proposed was crawled from one of the most popular reviewing website (www.dianping.com) in China. The dataset consists of two parts: The first part includes all restaurant reviews of Chengdu city (about 125MB). The overall number of restaurants is 22579 while only 5251 ones have been reviewed more than 10 times. The reviews of these 5251 restaurants is used for building our lexical base.

As to the second part, we firstly filter restaurants which have more than 100 reviews, then the most recent 100 reviews of 5 random selected restaurants are chosen as the second part of our dataset. The five restaurants are 陈麻婆豆腐(CMP),大嘴霸王排骨(DZ),好伦哥(HLG), 红杏酒家(HX), 陋室茶居烧烤五花肉(LS). Two students are employed to annotate all sentences and find the feature-opinion pairs. The Kappa scores for inter-rater agreements range from 0.75 to 0.79, which indicate good agreement. The details of the labeled feature-opinions pairs are given in Table 1.

4.1 Experiment Settings

Our method RuleMD is compared with three baseline methods: SimpleNoun, SimpleVerb, RuleNoMD which are described as follows.

SimpleNoun: This method only uses the *frequent-nouns* as feature words. If an adjective appears with a *frequent-noun* in a *short sentence*, the adjective is regarded as an opinion word. If there are more than one adjective words, only the closest one is regarded as the opinion word.

[5] Words like adverbs, conjunction and auxiliary words can appear between N and A.

Table 1. Restaurant details

Shop	Labeled-Pairs	Kappa
CMP	284	0.776
DZ	330	0.752
HLG	308	0.754
HX	386	0.791
LS	265	0.784

Table 2. F-score of restaurants

Index	F-score					
Shop	CMP	DZ	HLG	HX	LS	**Arg**
SimpleNoun	0.49	0.55	0.53	0.58	0.46	**0.52**
SimpleVerb	0.51	0.60	0.60	0.65	0.56	**0.58**
RuleNoMD	0.53	0.64	0.65	0.68	0.61	**0.62**
RuleMD	0.56	0.67	0.68	0.71	0.65	**0.66**

Table 3. Comparison Results

Index	Precision						Recall					
Shop	CMP	DZ	HLG	HX	LS	**Arg**	CMP	DZ	HLG	HX	LS	**Arg**
SimpleNoun	0.40	0.46	0.43	0.48	0.36	**0.43**	0.62	0.69	0.67	0.73	0.66	**0.67**
SimpleVerb	0.40	0.48	0.46	0.51	0.40	**0.45**	0.69	0.81	0.84	0.91	0.88	**0.83**
RuleNoMD	0.43	0.53	0.52	0.53	0.46	**0.49**	0.72	0.82	0.86	0.93	0.90	**0.85**
RuleMD	0.48	0.57	0.57	0.59	0.52	**0.55**	0.67	0.80	0.85	0.89	0.88	**0.83**

SimpleVerb: This method extends SimpleAdj by considering the verbs which have a 'v' POS-tag. A verb is considered as an opinion words if it is a *frequent-verb*, in other words, its *sentence appear ratio* is great than 1%.

RuleNoMD: This method extends SimpleVerb by incorporating the modified rules described in Section 3.2. The modification is made by disregarding the global threshold value of MD for further selecting the extracted feature-opinion pairs. That is to say, all word pairs that conform to the rule patterns will be treated as feature-opinion pairs. The purpose of testing this method is to check whether the rule pattern can improve performance than SimpleVerb, and furthermore, to check whether the lexical base has effects in feature-opinion extraction.

RuleMD: This is our proposed method and it extends SimpleVerb by incorporating the rules described in Section 3.2.

4.2 Evaluation Results

The comparison results are shown in Table 2 and Table 3, where 'Arg' represents the average result of the five restaurants. Below we discuss some detailed observations:

1: On F-Score, SimpleVerb is better than SimpleNoun. The reason is that certain verbs, such as '喜欢' (like), '失望' (disappoint), directly express the users' opinions. From Table 3 we can see that SimpleVerb's recall is much higher (16%) than SimpleNoun, and SimpleVerb's precision is also improved by 2%. This fact shows that the adding of verbs plays an important role in opinion extractions. Accordingly, the F-score is improved by 6%.

2: On F-Score, RuleNoMD is better than SimpleVerb. In addition, the precision improves by 4% and recall has 2% advances. The reason is that certain feature opinion pairs are not adjacent in a sentence, and the rules given in Section 3.2 is effective for capturing such kind of pairs.

3: On F-Score, RuleMD is better than RuleNoMD because of the using of our lexical base. It is worth notice that while the precision improves by 6%, recall decrease by

2%. This result shows that while we improve the precision by filtering some extracted feature-opinion pairs, a small part of correct pairs are also affected, consequently, ther recall of RuleMD is lower than RuleNoMD.

The proposed method has one importance parameter: the global threshold value of MD. Figure 2 shows the influences of this parameter on the overall performance. In Figure 2, the proposed method achieves the best average performance when the threshold is around 5.

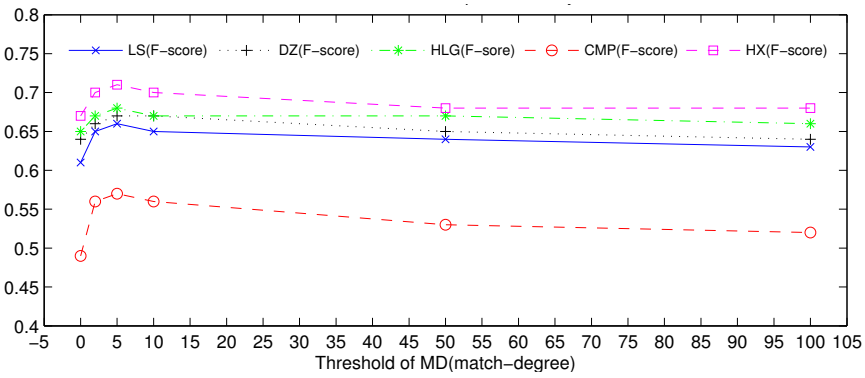

Fig. 2. Influence of the global threshold value of MD

5 Conclusion

This paper presented a hybrid approach that incorporates domain independent language patterns of Chinese and a lexical base for addressing the problem of feature-opinion extraction. Experiments showed that our method is effective and can gain a reasonable result. Currently, the introduced method has only been tested on the domain of restaurant reviews. In future work, we will further adapt our method (such as developing an adaptive architecture, refining the designed rules) so that it can be ported to other domain with minimal efforts.

Acknowledgements This work was jointly supported by the National Natural Science Foundation of China (No. 61133016, No. 61300094) and the Fundamental Research Funds for the Central Universities (No. ZYGX2013J070).

References

1. Brody, S., Elhadad, N.: An unsupervised aspect-sentiment model for online reviews. In: Proceeding of HLT '10. pp. 804–812 (2010)
2. Bross, J., Ehrig, H.: Automatic construction of domain and aspect specific sentiment lexicons for customer review mining. In: Proceedings of CIKM '13. pp. 1077–1086 (2013)

3. Freitas, L.A., Vieira, R.: Ontology based feature level opinion mining for portuguese reviews. In: Proceedings of WWW '13 Companion. pp. 367–370 (2013)
4. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of SIGKDD '04. pp. 168–177 (2004)
5. Ittoo, A., Bouma, G.: Minimally-supervised learning of domain-specific causal relations using an open-domain corpus as knowledge base. Data & Knowledge Engineering 85(0), pp. 57–79 (2013)
6. Jiao, F., Dong, G., Li, Q., Zhu, J.: Opinminer: Extracting feature-opinion pairs with dependency grammar from chinese product reviews. In: Proceedings of WISA '12. pp. 217–222 (2012)
7. Li, F., Huang, M., Zhu, X.: Sentiment analysis with global topics and local dependency. In: Proceedings of AAAI '10. pp. 1371–1376 (2010)
8. O'Leary, D.E.: Blog mining-review and extensions:"from each according to his opinion". Decision Support Systems 51(4), pp. 821–830 (2011)
9. Peñalver-Martínez, I., Valencia-García, R., García-Sánchez, F.: Ontology-guided approach to feature-based opinion mining. In: Proceedings of NLDB'11. pp. 193–200 (2011)
10. Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of HLT '05. pp. 339–346 (2005)
11. Somprasertsri, G., Lalitrojwong, P.: Mining feature-opinion in online customer reviews for opinion summarization. Journal of Universal Computer Science 16(6), pp. 938–955 (2010)
12. Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., Swen, B., Su, Z.: Hidden sentiment association in chinese web opinion mining. In: Proceedings of WWW '08. pp. 959–968 (2008)
13. Xu, X., Cheng, X., Tan, S., Liu, Y., Shen, H.: Aspect-level opinion mining of online customer reviews. Communications, China 10(3), pp. 25–41 (March 2013)
14. Yin, P., Wang, H., Guo, K.: Feature pinion pair identification of product reviews in chinese: a domain ontology modeling method. New Review of Hypermedia and Multimedia 19(1), pp. 3–24 (2013)
15. Zhai, Z., Liu, B., Zhang, L., Xu, H., Jia, P.: Identifying evaluative sentences in online discussions. In: Proceedings of AAAI '11. pp. 933–938 (2011)
16. Zhuang, L., Jing, F., Zhu, X.Y.: Movie review mining and summarization. In: Proceedings of CIKM '06. pp. 43–50 (2006)