Using Pointwise Mutual Information to Identify Implicit Features in Customer Reviews*

Qi Su, Kun Xiang, Houfeng Wang, Bin Sun, and Shiwen Yu

Institute of Computational Linguistics, School of Electronic Engineering and Computer Science, Peking University, 100871, Beijing, China
{sukia, xk, wanghf, bswen, yusw}@pku.edu.cn

Abstract. This paper is concerned with automatic identification of implicit product features expressed in product reviews in the context of opinion question answering. Utilizing a polarity lexicon, we map each adjectives in the lexicon to a set of predefined product features. According to the relationship between those opinion-oriented words and product features, we could identify what feature a review is regarding without the appearance of explicit feature nouns or phrases. The results of our experiments proved the validity of this method.

Keywords: pointwise mutual information, customer review, implicit feature.

1 Introduction

In recent years a large amount of research has been done on the identification of semantic orientation in text. The task is also known as opinion mining, opinion extraction, sentiment analysis, polarity detection and so on. It has focused on classifying reviews as positive or negative and ranking them according to the degree of polarity. The technology has proved to have many potential applications. For example, it could be integrated with search engines to provide quick statistical summary of whether a product was recommended or not in the World Wide Web, thus help consumers in making their purchasing decision and assist manufacturers in performing market analysis.

Our work studies the problem of answering natural language questions on customer reviews of products. The research is based on our previous projects of personalized information retrieval and natural language question answering(QA) system. Some similar works include multi-perspective question answering [1] [2] [3] and opinion question answering [4]. For a QA system, the questions from users could be of two types: objective questions and subjective questions. Subjective questions are those related to opinion, attitude, review, and etc. The identification of semantic orientation from corpora provides a possible way to answer user's subjective questions. According to the observation, those opinion

* This work was supported by National Natural Science Foundation of China (No. 60435020, No.60675035).

Y. Matsumoto et al. (Eds.): ICCPOL 2006, LNAI 4285, pp. 22–30, 2006.

related questions asked by users could be either general or feature-based. The following are such two query instances with opinions:

(1)"宝马汽车好不好？"(*How is BMW?*)
(2)"宝马汽车的动力性怎样？"(*How is the power of BMW?*)

For sentence (1), users want to get a general opinion about the automobile "BMW", while for sentence (2), users only want to get those opinions about the power of BMW and don't care other features. So, for our task, we need to first identify product features on which customers express their opinions.

Given a set of product reviews, the task of identifying semantic orientation could be divided into four subtasks [5]: (1)Identify product features expressed in the reviews; (2)Identify opinions regarding product features(including those general opinions); (3)Calculate the semantic orientation of opinions; (4)Rank the reviews based on their strength of semantic orientation.

Most of existing researches focused on subtask (3). In general, there are two approaches to semantic orientation calculation: namely, the rule based approach and the machine learning based approach. The former calculates the semantic orientation of each opinion-oriented word/phrase appeared in reviews and uses the average semantic orientation to represent the sentiment orientation of reviews. According to the value of average semantic orientation, it could rank reviews (subtask 4) easily. The latter, which is more widely employed, applies supervised automatic classification technology to classify reviews into bipolar orientation of positive or negative. Utilizing the semantic orientation of opinion-oriented word/phrase, this approach is proved to get an improved accuracy than using usual bag-of-word or n-gram features.

Although important, researches on the identification of product features are relatively few. Hu [6] classified product features that customers have expressed their opinion on as explicit features and implicit features. In their definition, the features which appear explicitly in opinion sentences are explicit features; the features which do not appear explicit in sentences are implicit features. But they did not propose a method for identifying implicit features. In this paper we address the issue of automatically identifying implicit features from domain dependent product reviews. To the best of our knowledge, no previous work has been conducted on exactly this problem. The existing researches on feature identification such as [5] [6] [7]mainly focus on finding features that appear explicitly as nouns or noun phrases in product reviews. The identification of implicit feature is a harder task than identifying explicit one. According to the observation, opinion-oriented word/phrase could usually be a good indicator to the feature which it modifies. So the key issues to the task are to define a feature set related to specific domain and map those indicators to the set of predefined features.

We take a mutual information approach to address the problem. In this paper we first present our approaches in the identification of opinion-oriented words/phrases, and map the adjective ones to the corresponding implicit product features in reviews. By identifying implicit features, we could undoubtedly get more reasonable semantic orientation scores on different product features.

2 Definition of Features in Product Reviews

2.1 Explicit Features and Implicit Features

Product features include attributes, such as "加速性"(*acceleration*) for automobile products, and parts, such as "刹车"(*brake*). A feature can appear either explicitly or implicitly in product reviews. Features that can be extracted directly from the reviews are Explicit Features, e.g., 动力性(*power*) in "宝马汽车的动力性不错"(*BMW power is not bad*). Sometimes users express their opinions without explicit feature words. But we could still deduce the features which their opinions toward from the opinion sentences. Those kinds of features are Implicit Features. For example, in the sentence of "宝马汽车很漂亮"(*BMW is beautiful*), we could judge that the feature which users talk about is BMW's *exterior* although the word doesn't appear explicitly.

The identification of explicit features is relatively easy. Using frequent nouns and noun phrases appeared in reviews with some pruning strategies, the experiment in [6] could reach the best average precision of 79%. Evaluating each noun phrase by computing mutual information between the phrase and meronymy discriminators associated with the product class, the experiment in [5] achieved 86% in precision and 89% in recall.

In the research of [8], they represented a method to identify implicit features by tagging the mapping of specific feature values to their actual feature. Our work is close to their research in the sense that we also use mapping rules, but it is also different in that we propose an automatic method to generate the mapping rules and use opinion-oriented words.

2.2 Feature Structures in the Domain of Automobile Review

Our work in this paper is focused on identifying implicit features in product reviews. Since there are no feature words appeared in reviews, we need firstly to define a set of review features for automatic opinion mining. The definition of features are domain dependent. In our experiment, we consider the domain of automobile review.

According to the investigation of several automobile review websites, we found that people usually evaluate a car from several aspects, such as 动力性(Power), 操控性(Handling), 外观(Exterior) , 内饰(Interior), 经济性(Economy), 工艺性(Craft), 市场性(Marketability) and etc. So in this paper we define the product feature sets of automobile according to the 7 items above. This is a very detailed review feature proposal. In fact, many automobile review websites classify automobile review features on a rough level. For example, they combine the features of 动力性(Power) and 操控性(Handling) as 性能(Performance), combine the features of 外观(Exterior) and 内饰(Interior) as 设计(Design). From the rough classification schemes, we may have an impression that the feature of 动力性(Power) and 操控性(Handling) may be hard to divide, and so as the feature of 外观(Exterior) and 内饰(Interior). For our task of implicit feature identification, they are also problems. For example, when users speak of *an exquisite automobile*, they may consider both exterior design and interior design.

So, we propose that a feature indicator should be mapped to several implicit features. It also seems to accord with our instinct.

Product features which appear explicitly in automobile reviews can also be classified into the feature structure. For example, *acceleration* belongs to the feature of Power, *brake* belongs to the feature of Handling.

3 Feature Identification

3.1 The Method of Pointwise Mutual Information

Pointwise Mutual Information (PMI) is an ideal measure of word association norms based on information theory [9]. Researchers have applied this measurement to many natural language processing problems such as word clustering. PMI compares the probability of observing two items together with the probabilities of observing two items independently. So it can be used to estimate whether the two items have a genuine association or just be observed by chance.

If two words $word_1$ and $word_2$ have probabilities $P(word_1)$ and $P(word_2)$, then their mutual information $PMI(word_1,word_2)$ is defined as [10]:

$$PMI(word_1, word_2) = \log\left(\frac{P(word_1 \& word_2)}{P(word_1)P(word_2)}\right) \tag{1}$$

Usually, word probabilities $P(word_1)$, $P(word_2)$ and joint probabilities $P(word_1 \& word_2)$ can be estimated by counting the number of observations of $word_1$, $word_2$ and the co-occurrence of $word_1$ and $word_2$ in a corpus normalizing by the size of the corpus. The co-occurrence range of $word_1$ and $word_2$ is usually limited in a window of w words.

The quality of the PMI algorithm largely depends on the size of training data. If there is no co-occurrence of $word_1$ and $word_2$ in the corpus, the accuracy of PMI becomes an issue. The PMI-IR algorithm introduced by Turney in [11] used PMI to analyze statistical data returned by the query of Information Retrieval(IR). So, the corpus used by PMI-IR algorithm is the document collection which is indexed by IR system. The PMI-IR algorithm is proposed originally for recognizing synonyms in TOEFL test. Then Turney [12] used this method in their sentiment classification experiments, where the sentiment orientation $\hat{\sigma}(w)$ of word/phrase w is estimated as follows.

$$\hat{\sigma}(w) = PMI(w, positive) - PMI(w, negative) \tag{2}$$

In equation (2), *positive* and *negative* means a set of Reference Words Pair (RWP)[13] with the sentiment orientation of positive or negative respectively. Choosing the RWP of *excellent* and *poor*, the equation above can be written as:

$$\hat{\sigma}(w) = \log\frac{hits(w, excellent)/\ hits(excellent)}{hits(w, poor)/\ hits(poor)}, \tag{3}$$

Where $hits(query)$ is the number of hits (the number of documents retrieved) when the query *query* is given to IR system. And $hits(query1, query2)$ is the number of hits "*query1* NEAR *query2*".

3.2 Feature-Based PMI Algorithm

In Turney's research, semantic orientation was calculated by the word association with a positive paradigm word minus a negative paradigm word. For our research, we expand the RWP of two words to a set of features. In our case, we have a set $S_{feature}$ which holds different features.

$$S_{feature} = \{feature_1, feature_2, ..., feature_n\} \tag{4}$$

For each selected word w, we calculate the PMI between w and $feature_i$ ($feature_i \in S_{feature}$). Using the PMI method, we get the word association between w and different features, then map w to one or several features according to the probability.

The PMI-IR algorithm use a NEAR operator to simulate the co-occurrence window limit of two words in PMI algorithm. But using NEAR operator has its limitation. The two NEAR words may be distributed on the different sentences or clauses, or even two paragraphs. Thus their semantics may be non-sequential and not represent the correct co-occurrence relationship. For keeping the two words in a semantic sequence, we suggest an improvement of PMI-IR algorithms for our task.

We use conjunction operator to construct queries. Because we have limit *word* to be adjectives in the polarity lexicon. As features are always nouns/noun phrases, so the query of *"word feature"* would form a normal collocation relation.

$$Score_{conjunction}(word, feature) = \log \frac{p\ (word\ feature)}{p\ (word)\ p\ (feature)} \tag{5}$$

Since we calculate the association between w and the feature set $S_{feature}$, we can drop $p(word)$. Using World Wide Web as the corpus (using the query results returned by search engine), the equation(5) can be simplified as follows:

$$Score_{conjunction}(word, feature) = \log \frac{hits(\text{“}word\ feature\text{”}) + \varepsilon}{hits(feature)} \tag{6}$$

Where ε is a parameter to prevent the numerator from getting zero when there is no hits returned for the query of *"word feature"*.

4 Experiment

4.1 Opinion Words Collection

For collecting opinion words, we first download automobile review webpages from the internet, and then extract the opinion words from them manually. Thus we get a small/basic polarity lexicon. Using these words as seeds, we enlarge our polarity lexicon utilizing synonym and antonym sets in the Chinese Concept Dictionary(CCD), a Chinese version of the online electronic dictionary Wordnet. The consideration of utilizing CCD is, in general, words share the same orientation as their synonyms and opposite orientation as their antonyms [6]. For example, the positive word 可爱(*lovely*) has the synonym set of {优美 伶俐 可爱 喜人 柔情 深情 漂亮 甜蜜 痛快 秀丽 绝妙 讨人喜欢} and the antonym set of{可恨 可恶 可憎 讨厌}. The two sets take the opposite orientation of positive and negative accordingly. In our research, according to the structural characteristic of CCD and the expansion results, we only use the result of synonym expansion, and enlarge the polarity lexicon more than 5 times larger.

Most of the opinion words in our lexicon are adjectives. Being good modifiers for product features and carriers of sentiment, in our experiment, only the adjectives in polarity lexicon were chosen to identify implicit product features. We map each adjective to the set of product features which it could modified.

4.2 Mapping Opinion Words to Features

The IR system used in our experiments is Google. Google provides API to refer queries conveniently and rapidly. Google API returns estimated hit counts rather than actual values. This would add some noise into our model, but still has the necessary stability for our estimation.

For each product feature in the above mentioned feature set $S_{feature}$, we choose one or several representative nouns as feature words to calculate the PMI between them and opinion words. Such as, for the feature of 动力性(Power), we choose the words of 动力(power),引擎(engine),马力(horsepower) and etc.

Table 1 shows PMI scores for some word examples.

Table 1. PMI scores for some word examples

Words	Power	Handling	Exterior	Interior	Economy	Craft	Marketability
漂亮(beautiful)	-12.01	-14.15	-4.79	-8.67	-12.87	-9.41	-12.76
强劲(powerful)	-4.26	-13.20	-9.91	-7.96	-13.04	-10.46	-8.76
一般(ordinary)	-7.86	-9.14	-9.51	-7.47	-6.24	-5.20	-6.15
昂贵(expensive)	-15.50	-15.76	-13.40	-9.86	-16.09	-11.78	-5.58
考究(exquisite)	-17.70	-15.76	-10.00	-8.23	-16.09	-3.94	-16.21
灵活(flexible)	-10.27	-4.13	-11.00	-8.35	-10.16	-14.65	-12.15
别致(unique)	-17.70	-15.76	-9.92	-9.60	-16.09	-9.49	-16.78
方便(convenient)	-12.29	-5.07	-9.51	-9.47	-12.37	-11.44	-12.40
不足(inadequate)	-4.37	-10.15	-11.88	-10.27	-8.99	-9.75	-8.75
完美(perfect)	-9.88	-9.64	-10.03	-7.95	-4.86	-8.41	-11.88

For each opinion word, we have got its PMI scores on each $feature_i \in S_{feature}$. We need to further map them to one or several most suitable features according to their PMI values.

For this purpose, we first define a function to describe the difference between $\hat{S}(\hat{S} \subset S_{feature})$ and $S_{feature}$-$\hat{S}$ for a word w.

$$score_{diff}(\hat{S}, S_{feature}, w) = \frac{\sum_{f \in \hat{S}}(f, w)}{|\hat{s}|} - \frac{\sum_{f \in (S_{feature} - \hat{S})}(f, w)}{|S_{feature} - \hat{s}|} \quad (7)$$

Then we calculate a series of $score_{diff}$ for each word w using an algorithm as described in the following pseudo code:

Algorithm 1. The calculation of $score_{diff}$

Set $S_1 = \emptyset$, $S_2 = S_{feature}$
while $S_2 \neq \emptyset$ **do**
  $f_i = \arg\max_{f \in S_2} Score(f_i, w)$
  Set $S_1 = S_1 + \{f_i\}$,$S_2 = S_2 - \{f_i\}$
  Calculate $Score_{diff}(S_1, w)$
end

Table 2 is the $score_{diff}$ of some word examples.

Table 2. Value of $score_{diff}$ of some word examples

考究(exquisite)		强劲(powerful)	一般(ordinary)	灵活(flexible)
Craft	10.06	Power 6.30	Craft 2.53	Handling 6.97
Interior	9.07	Interior 4.96	Marketability 2.37	Interior 5.41
Exterior	9.05	Marketability 4.66	Economy 2.63	Economy 4.47
Handling	7.18	Exterior 4.51	Interior 2.57	Power 4.37
Economy	6.15	Craft 4.85	Power 2.74	Exterior 4.62
Marketability	6.00	Handling 4.14	Handling 2.5	Marketability 5.31

We use value *gap* to describe the margin between two $diff_{score}$s and set a experiential threshold $\xi = 1.0$. When the margin is greater than ξ, we say that the two adjacent $diff_{score}$ have a *gap*. With each word and its $diff_{score}$, we find every gap value from high to low of a set of $diff_{score}$, and judge the feature class which the word should belong to according to the sequence.

From Table 2, the biggest *gap* for 考究(exquisite) should between the feature of Exterior and Handling. So we map the word 考究(exquisite) to feature sets of Craft,Interior,Exterior. For the word of 强劲(powerful), the biggest *gap* should between Power and Interior. So we map the word 强劲(powerful) to the feature Power. As for the word 一般(ordinary), we find that there is no *gap* between every two features. So we consider that the word 一般(ordinary) is a general descriptive adjective.

Based on this method, we map the adjective words in our polarity lexicon to the predefined product feature set. Table 3 gives some experimental example results.

Table 3. Mapping results of some example words

words	feature sets
漂亮(beautiful)	Exterior
强劲(powerful)	Power
一般(ordinary)	-
昂贵(expensive)	Marketability
考究(exquisite)	Craft,Interior,Exterior
灵活(flexible)	Handling
别致(unique)	Craft,Interior,Exterior
方便(convenient)	Handling
不足(inadequate)	Power
完美(perfect)	Economy

5 Conclusion

Feature identification in product reviews is the first step of opinion QA and other opinion mining tasks. In product reviews, the appearance of feature nouns or noun phrases is usually an important clue which aids in mapping to predefined product features. But according to our observation of product review webpages, in many cases, those explicit feature nouns or noun phrases do not appeared in the context. Here comes the importance of the identification of implicit product features. According to the mapping between opinion words and product features, we could judge what features a review is regarding to, without the explicit appearance of feature nouns or noun. In this paper, we describe a method which uses PMI to calculate the association between opinion-oriented adjectives and a set of product review features. We do not conduct a quantitative evaluation of our method, because the relationship between some adjectives, especially those general ones and product features are likely to be highly contextual. But the results of our experiments prove the validity of this method intuitionistically.

Using the method proposed in this paper, we supply our polarity lexicon with the corresponding product feature information. With this resource, we could score a product review from different aspect of features. Future work should include the weighting of explicit features and implicit features when both of the feature types appear in a review.

References

1. Cardie, C., Wiebe, J.,Wilson, T. and et al.: Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering. Proceedings of the AAAI Spring Symposium on New Directions in Question Answering(2003)
2. Wiebe, J., Breck, E., Buckley, C and et al.:Recognizing and Organizing Opinions Expressed in the World Press.Proceedings of the 2003 AAAI Spring Symposium on New Directions in Question Answering(2003)

3. Stoyanov, V., Cardie, C., Wiebe, J.:Multi-Perspective Question Answering Using the OpQA Corpus.Proceedings of HLT-EMNLP 2005(2005)
4. Yu, H., Hatzivassiloglou, V.: Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences.Proceedings of EMNLP-03, 8th Conference on Empirical Methods in Natural Language Processing(2003)129–136
5. Popescu, A. M. and Etzioni, O.: Extracting Product Features and Opinions from Reviews. Proceedings of HLT-EMNLP(2005)
6. Hu, M. Q. and Liu, B.: Mining and Summarizing Customer Reviews. In KDD(2004)168-177
7. Kobayashi, N., Inui, K., Tateishi, K and et al.:Collecting Evaluative Expressions for Opinion Extraction. In IJCNLP(2004)596-605.
8. Liu, Bing., Hu, M. Q., and Cheng, J. S.:Opinion observer: analyzing and comparing opinions on the Web. Proceedings of WWW '05, the 14th international conference on World Wide Web(2005)342–351
9. Thomas, M. Cover. and Joy A. Thomas.: Elements of Information Theory. John Wiley(1991)
10. Church, K. W. and Hanks, P.: Word Association Norms, Mutual Information and Lexicography. Proceedings of the 26th Annual Conference of the Association for Computational Linguistics(1989)
11. Turney, P.D.: Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. Proceedings of the Twelfth European Conference on Machine Learning. Berlin: Springer-Verlag(1991) 491-502
12. Turney, P. D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proc. of the 40th Annual Meeting of the Associationfor Computational Linguistics(2002) 417-424
13. Ye, Q., Shi, W., Li, Y. J.: Sentiment Classification for Movie Reviews in Chinese by Improved Semantic Oriented Approach. Proceedings of the 39th Annual Hawaii International Conference on System Sciences(2006)
14. Bethard, S., Yu, H., Thornton, A. and et al.:Automatic extraction of opinion propositions and their holders. In 2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text(2004)