Expert Systems with Applications 39 (2012) 10283–10291

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis

Wenhao Zhang, Hua Xu *, Wei Wan

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

ARTICLE INFO

Keywords:
Product weakness
Business intelligence
Sentiment analysis
Feature grouping

ABSTRACT

Finding the weakness of the products from the customers' feedback can help manufacturers improve their product quality and competitive strength. In recent years, more and more people express their opinions about products online, and both the feedback of manufacturers' products or their competitors' products could be easily collected. However, it's impossible for manufacturers to read every review to analyze the weakness of their products. Therefore, finding product weakness from online reviews becomes a meaningful work. In this paper, we introduce such an expert system, Weakness Finder, which can help manufacturers find their product weakness from Chinese reviews by using aspects based sentiment analysis. An aspect is an attribute or component of a product, such as price, degerm, moisturizing are the aspects of the body wash products. Weakness Finder extracts the features and groups explicit features by using morpheme based method and Hownet based similarity measure, and identify and group the implicit features with collocation selection method for each aspect. Then utilize sentence based sentiment analysis method to determine the polarity of each aspect in sentences. The weakness of product could be found because the weakness is probably the most unsatisfied aspect in customers' reviews, or the aspect which is more unsatisfied when compared with their competitor's product reviews. Weakness Finder has been used to help a body wash manufacturer find their product weakness, and our experimental results demonstrate the good performance of the Weakness Finder.

1. Introduction

Finding the weakness of the products from the customers' feedbacks can help the manufacturers improve their products design and make a better business strategy. In the past, this work must be done by collecting customers' feedbacks from the investigation questionnaires, which is both expensive and time consuming. Nowadays, as the emergence of Web 2.0 websites provides a good platform for the customers to write their opinions online (Hu & Liu, 2006), such as Amazon.com, Cnet.com, Taobao.com, etc. We can easily gather large amount of the customers' reviews about the products online. However, the volume of these reviews is too heavy for human to read reviews one by one. Therefore, helping the manufacturers automatically and intelligently find their product weakness from these opinions becomes exceedingly valuable work. In this paper, we proposed aspects based sentiment analysis system, Weakness Finder, to help a well-known cosmetic manufacturer find their body wash product weaknesses.

Aspect, also called *feature*, is usually the property or function of the product, for example, *price*, *degerm*, and *moisturizing* are the aspects of the body wash products. We know the weaknesses of the products are always the properties or the functions which the customers were not very satisfied. So the weaknesses are the subset of aspects for a product and should be identified before finding the weakness. Intuitively, for the reviews of the given product, the aspect mentioned in negative reviews much more than that in positive reviews might be the product weakness. Furthermore, when compared with competitors' product, the aspect was much more widely discussed in negative reviews but less in positive reviews might be the weakness. Following this idea, we should analyze the problem in two dimensions, aspects and sentiment, so that we can measure how many reviews comment the aspects with undesirable attitudes and how many of them comment with positive attitudes.

The aspects could be identified as two kinds of features, *explicit features* and *implicit features* (Liu, 2010). If a feature or any of its synonyms appears in a sentence, the feature is called an *explicit feature* in this sentence, such as "*cost*" in Example 1 is the synonym of price, it can explicitly show it mentions the "*price*" aspect as an explicit feature. On the contrary, the *implicit features* are the

* Corresponding author. Tel.: +86 10 62796450; fax: +86 10 62782266.
E-mail address: xuhua@mail.tsinghua.edu.cn (H. Xu).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.02.166

features not appearing directly in a review, but can imply the aspect, such as "*cheap*", "*cost-saving*" and "*affordable*" can imply the price aspect. The relationship between the aspect and its corresponding feature words can be illustrated by Fig. 1.

Group explicit and implicit features for each aspect consists of two tasks. Feature words should be identified from the reviews, and then group the feature words into their corresponding aspects. Numerous solutions have been proposed for the former task, and most of the solutions are statistics based methods (Hu & Liu, 2004; Liu, Hu, & Cheng, 2005; Moghaddam & Ester, 2010; Popescu & Etzioni, 2005; Zhan, Loh, & Liu, 2009). These methods typically apply some rules to restrict the high frequency nouns in order to find the explicit features. And the limitation is that they ignore the low frequency feature words. For the feature grouping task, the intuitive method is based on the word knowledge resources (Alvarez & Lim, 2007; Hughes & Ramage, 2007; Yang & Powers, 2005) to find the words' synonyms, such as thesauri, WordNet. The more natural way is to use an unsupervised learning method to group the explicit features, and the similarities are measured by distributional properties of words (Bollegala, Matsuo, & Ishizuka, 2007; Fellbaum, 1998; Lin, 1998; Zhai, Liu, Xu, & Jia, 2011), but whether these methods work well in Chinese reviews is to be validated. Implicit features identification and grouping are very tough work, (Hu & Liu, 2004) mentioned the association rule method can help to find implicit features rules. And most papers ignore implicit features identification in their experiment, because it's a tough task.

Example 1. "*产品也很***便宜***|***价格***很***划算***|我能***买得起***" (The product is very **cheap** The **price** is **cost-saving** The product is **affordable**).

The sentiment analysis plays an important role, and it has been extensively discussed since 1990s (Liu, 2010; Pang & Lee, 2008; Tang, Tan, & Cheng, 2009). There are mainly two main approaches to do sentiment analysis, one is based on semantic analysis (Ding, Liu, & Yu, 2008; Hu & Liu, 2006; Popescu & Etzioni, 2005; Yu & Hatzivassiloglou, 2003), and the other one is based on machine learning (Mullen & Collier, 2004; Pang, Lee, & Vaithyanathan, 2002; Rushdi Saleh, Martín-Valdivia, Montejo-Ráez, & Ureña-López, 2011; Zhang, Ye, Zhang, & Li, 2011), and the methods based on machine learning are commonly used in document-level sentiment analysis. However, The Weakness Finder needs to compare customers' attitudes for each aspect, so the system needs sentence-level sentiment analysis to judge the polarity of each sentence and the corresponding aspect. In traditional sentence level sentiment analysis, they focus on determining the sentiment degree by considering the sentiment words, negation words and transition words. However, Adverb of degree could also be used to measure the intensity of the sentiment. At the same time, it is a signal to express the customer's sentiment in a sentence especially in Chinese.

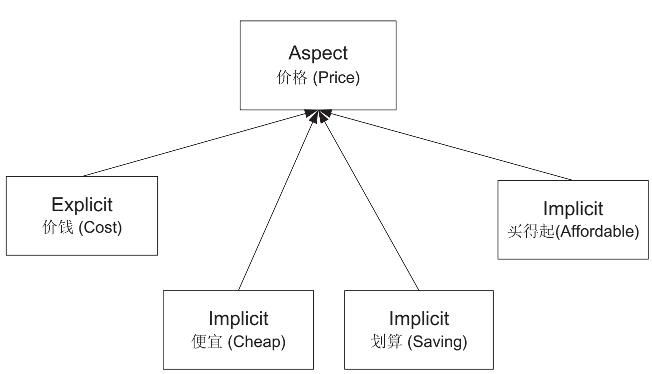

Fig. 1. The relationship between aspect and its explicit and implicit feature words.

Because the sentiment words list could not contain all sentiment words and the neural words with sentiment are also difficult to find, the adverb of degree could be used to identify the sentiment orientation by considering the context information.

Weakness Finder employs the aspects based sentiment analysis method to find the weakness and the expert system could be decomposed into 5 subtasks:

1. The product reviews crawling, preprocess and candidate feature words extraction.
2. Group products feature words into corresponding aspects.
3. Determine sentiment orientation of each aspect in reviews' sentences.
4. Find product weakness from specific product reviews.
5. Find product weakness by comparing with their competitors' products.

This paper introduces the expert system Weakness Finder which embodies a solution to the above 5 subtasks. We crawled the cosmetic reviews from the Internet in order to help a well-known cosmetic manufacturer find their body wash product weaknesses, the evaluation of the effect of Weakness Finder has also been discussed. And in general, we have the following contributions in this paper:

1. To the best of our knowledge, Weakness Finder is the first complete general expert system which aims at helping the manufacturers to find their product weakness of the products from Chinese products reviews.
2. Implicit features identification is a very tough task, however, Weakness Finder employs collocation statistics based method to find the implicit features.
3. Weakness Finder groups products feature words into corresponding aspects for Chinese reviews by applying semantic methods.
4. The impact of adverbs of degree has been taken into consideration in the procedure of sentence level sentiment analysis.

In this paper, we introduce our system Weakness Finder to help manufacturers to find the weakness of their products by using aspects based sentiment analysis. The rest paper is organized as follows: in Section 3 we introduce the architecture and detailed algorithm of Weakness Finder, which includes explicit and implicit features identification, features grouping and sentiment analysis methods, then introduce the data analysis to find product weakness. In Section 4, we will show the experiment and algorithm performance on Chinese body wash product reviews. And in Section 5, we will give our conclusion and further work.

2. Problem definition and dataset

In this section, the problem statement and the experimental dataset have been introduced.

2.1. Problem statement

Let $P = \{P_1, P_2, \ldots, P_n\}$ be a set of products which are made by different manufacturers, like "Apple's iPhone", "Google's Nexus One", and "Samsung's Galaxy", they are all the cell phone but made by different companies. For each product P_i, there exists a set of reviews $R_i = \{R_1, R_2, \ldots, R_m\}$. And for each review R_j, it may consist of some sentimental sentences about the corresponding product's aspect. Therefore, let $A = \{A_1, A_2, \ldots, A_l\}$ be a set of aspects of product, such as "*price*", "*battery life*" and "*keyboard*", etc. The weakness of the product is the aspect which the customers are not satisfied

Table 1
Dataset statistics.

	Brand A	Brand B
Reviews number	526	561
Sentences number	1111	1646

with, and we know the weakness of products W must be the subset of A, i.e. $W \subseteq A$.

An aspect can be expressed as different feature words, including explicit feature words and implicit feature words, such as the aspect "*price*" could be described by an explicit feature word "*cost*" which is a synonym of "*price*", and could also be implied by the implicit feature word "*expensive*". So the aspect $A_k = \{f_1, f_2, \ldots, f_p\}$, in which f_u is the explicit or implicit feature words to describe Aspect A_k. Besides, for each aspect A_k, the sentiment of A_k can be shown as S_k. Therefore, the pair $O_k = \langle A_k, S_k \rangle$ could be detected to show the sentiment orientation for each aspect.

Problem Definition: Given a set of reviews for different manufacturers' products P. Firstly, the task is to identify and group the explicit and implicit features words f for each aspect A. Secondly, determine the pair of each aspect and its sentiment $O = \langle A, S \rangle$. Thirdly, find the product weaknesses W by both comparing the result O of each aspect for a specific product P_i and comparing the results O with different products P.

2.2. Dataset

The experimental data we used in this paper is crawled from two of the most popular cosmetic websites, Luce (http://www.luce.com) and Yoka (http://www.yoka.com), which were suggested by a big multinational cosmetic products manufacturer. We ask two persons annotate the data independently, they judge the sentiment of each feature in a sentence, and then annotate them like Example 2. The structure is as the following format:

Aspect1: *Feature1 Sentiment1 Feature2 Sentiment2*...
Aspect2: *Feature1 Sentiment1 Feature2 Sentiment2* ...

If the sentiment of the feature is positive, we give it score number 1, or if neutral, give it 0, otherwise, if the feature in the review shows the negative attitude, give it −1. The labeled data is shown in Example 2.

Example 2. "*这个木瓜的我最爱, 味道好, 保湿效果很棒, 且很容易清 洗干净*" ("This product with papaya is my favorite, the smelling is good, the performance of locking in moisture is great, and it's very easy to clean.") can be annotated as "*香味*: *味道 1*; *滋润*: *保湿 1*; *清洁*: *干净 1*;"

Finally, we ask another person to review both of their judgment and to double check the labeled dataset. Here are the statistics of two brands' body wash reviews data shown in Table 1. And Brand A is our target product which needs to be found weakness, and in the rest part of the paper, we will use the data to explain how Weakness Finder works.

3. The proposed technique of Weakness Finder

3.1. Weakness Finder overview

The architectural overview of Weakness Finder system is illustrated in Fig. 2. Actually, the process could be divided into 5 subtasks as we mentioned in Section 1. In order to find the weakness of the product, Weakness Finder needs to crawl the reviews data from the Internet at first. However, the raw data may contain the duplicated entries, misspelling words, and it is different from English, Chinese needs to be segments into words, so preprocess need to be done before data mining.

From Example 1, we knew that different feature words can refer to one aspect, such as "*cheap*", "*cost*", "*affordable*" all imply the aspect price. So extracting the features and grouping the features into different aspects are very important work to find whether a sentence is referred to an aspect. Therefore, Weakness Finder employs semantic methods to find explicit features, and statistical methods to find implicit features. Then an improved sentence level sentiment analysis has been proposed to determine the sentiment orientation of each aspect.

Finally, the weakness of the products could be found from two perspectives. One is to compare which aspect of the product is most unsatisfied by the customers. The other perspective is to know which aspect is poorer when compared with competitors' products. And in this rest of this section, we introduce the detailed approaches for the above work in turn.

3.2. Preprocess and candidate features extraction

Weakness Finder aims to find the aspects people like or dislike in a review about a product. So find out the feature words which can be used to describe an aspect in a sentence become an important work.

This work can be explained as following, our input review can be shown as follows:

"这个沐浴露保湿效果很好, 但是美白我倒没发现。 对于它的清洁力和杀菌力一直是比较信赖的。" ("This product has good performance on moisturizing, but the whitening function is not significant. The products' cleaning and degerm functions are reliable"). Then the feature extraction work will extract "*moisturizing*", "*whitening*", "*cleaning*", "*degerm*" features words from our sentences, therefore, we could know customers commented on which point.

3.2.1. Preprocess

Before extract the candidate features words from the review content, the review needs to be preprocessed at first, and our system takes following steps, illustrated in Fig. 3. (1) The customer may resubmit his review content because websites do not always response that quickly. The repeated reviews need to be removed from the database to correct the interference. (2) There are two types Chinese, simplified and traditional Chinese, because we did not know where the customer comes from, and use which type of Chinese. Therefore unify the review content into simplified Chinese is an essential preprocess work to help the computer to analyze the reviews automatically. (3) Chinese sentence has no space between the words, it needs to be segmented and does the POS tagging first, and our system applies ICTCLAS (Zhang, Yu, Xiong, & Liu, 2003) to do this work. According to our experience, we only keep the noun, verb and adjective words as the candidate features words. (4) Then remove the stopping words and correct the misspelling words in order to get candidate features. We also put the names of the body wash products in the stopping word list, because these words could not be a feature name, although it may have a high word frequency.

3.2.2. Extract feature candidates and aspects selection

The frequency of the features can help us find the candidate features set. And in our work, we define a word as a candidate feature if its word frequency is more than 10 in all reviews. For the Chinese, single character usually can not express the meaning clearly, such as "露" (dew), "好" (good), so we prune the single character features from the candidate feature set. And in our dataset, the top 10 frequent words are "*perfume*", "*not bad*", "*like*", "*bath*", "*wash*", "*moisturizing*", "*cleanness*", "*foams*", "*effects*", "*purchase*".

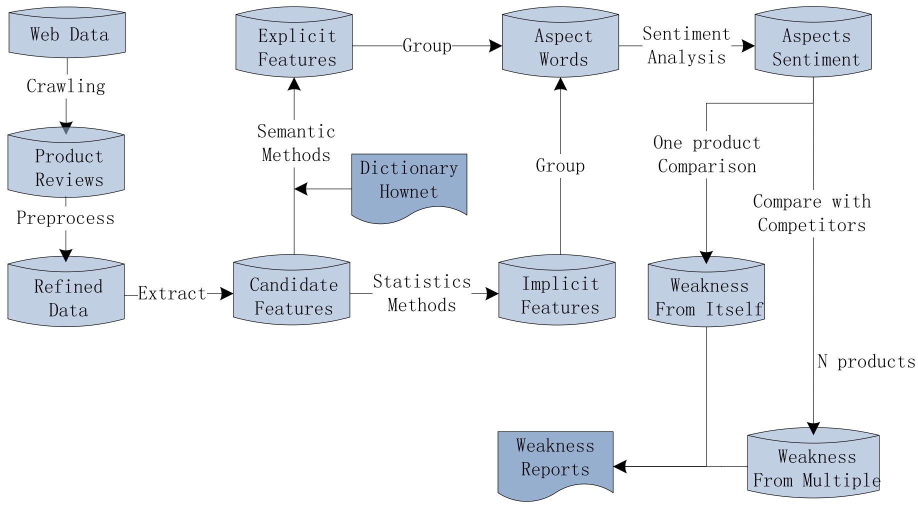

Fig. 2. The architectural overview of Weakness Finder system.

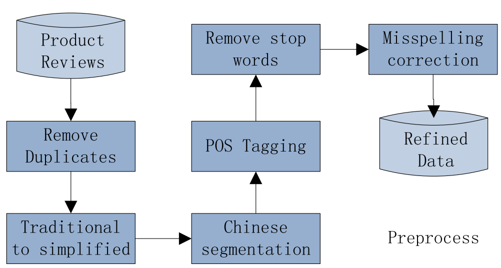

Fig. 3. The preprocess work of the Weakness Finder system.

In our project, the experts in cosmetic industry provide their most concerning aspects according to their professional experience, the aspects are namely "*degerm*", "*cleanness*", "*perfume*", "*moisturizing*", "*easy to wash*", "*dense*", "*dilute*". However, some other important aspects are also discussed widely by consumers according to our word frequency statistics, namely "*price*", "*effect*", "*comfort*", "*advertisement*". The above 11 aspects will be discussed in the rest of paper.

The 11 aspects could be classified into 3 categories. The first category is the aspects which can describe properties of body wash products, including "*degerm*", "*cleanness*", "*perfume*", "*moisturizing*", "*easy to wash*", "*dense*", "*dilute*", so the evaluation result of these aspects can help the manufacturer concentrate on the improvement of the weak property of the product. The second category is the comprehensive feedback when they are using the products, these aspects include "*effect*", "*comfort*". The other category of aspects can show the customers' evaluation on products' market strategy, including "*price*", "*advertisement*". Table 2 illustrated the categorization and its notations, and the category notations will be used in the rest of this paper.

Table 2
The aspects categorization and its notations.

Category name	Notation	Aspect name
Product property	C1	Perfume Moisturizing Easy to wash Cleanness Degerm Dilute Dense
Comprehensive feedback	C2	Effect Comfort
Marketing	C3	Price Advertisement

3.3. Grouping features

Grouping the features into different aspects is to cluster the words describe the same aspect together. From Fig. 1, we know that some words like "*price*" and "*price level*" can explicitly describe the aspect "*price*", while some words like "*cost-saving*", "*cheap*" and "*affordable*" can imply the aspect implicitly. In this part, we will introduce the algorithm to group two kinds of features respectively.

3.3.1. Grouping explicit features

Intuitively, the explicit features in the same group are usually the synonyms, antonyms, or the similar concept words. The traditional way to find synonyms and antonyms is to look up the Chinese synonyms and antonyms dictionary, it can provide precise synonyms and antonyms words, for example, the synonyms of "*cleanness*" are "*clean*" and "*cleanse*", and the antonym is "*dirty*". However, the results are always limited due to the development of vocabulary and some words may not be synonyms according to different domains. In the following, another two semantic based method similar concepts grouping methods were used in our system.

Chinese words are ideographic, so utilizing the word structure can help to find the similar concept words from the candidate features set. Although single Chinese character (morpheme) cannot express the exact meaning, but it is an important clue to find the similar concept words. This property could be used to find the explicit features from the corpus. It is a star structure. For example, the morpheme "价" in the word "价格" (price) can help us get more similar concept words, like "价位" (price-level), "价钱" (cost), etc. The result can be illustrated in Fig. 4, and naturally we could see the words with the sharing words might be similar concepts with a high probability. And we also include some important single character (morpheme) in our feature words list manually, such as "*wash*", because they are the indicators of the aspect. And the explicit features words results are shown in Table 3.

It is obvious that the morpheme based method relies on the candidate feature set, which only contains the frequent features. And another way to extend the word set with the explicit features

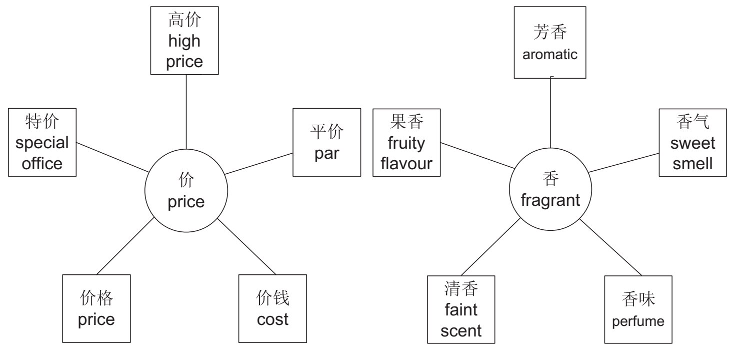

Fig. 4. The star structure of explicit features by using morpheme based method.

Table 3
Similar concepts discovery by utilizing the morpheme method.

Aspect	Price	Perfume	De-germ	Cleanness
Feature word 1	Price-level	Fruity flavor	Bacterial	Cleansing
Feature word 2	Parity	Sweet smell	Sterilize	Cleanse
Feature word 3	High-price	Aromatic	Disinfect	Cleanliness
Feature word 4	Cost	Faint scent	Antibacterial	Fresh and cool

is to calculate the similarity of the words based on the Chinese WordNet–Hownet (Dong & Dong, 2006), it can help us to find infrequent features from the dataset. The words with high similarity score indicate two words might be the near synonym words with a high probability, so the score can be used to help complete the aspect features set. The pseudo code is shown in Algorithm 1. Utilizing this method, we can extend the aspect "价格" (price) with the similar words like "售价" (sell price), "市场价" (market price) and so on.

Algorithm 1: Find near synonyms from Hownet

for each feature f_j in aspects A_i **do**
  for each features *cf* in feature list **do**
    if the similarity between f_j and *cf* > *threshold* **then**
      Add *cf* to f_j's set A_i
    end if
  end for
end for

3.3.2. Grouping implicit features

The implicit feature mining is an important but difficult technique. (Hu & Liu, 2004) simply mentioned they get the implicit features by using association rules. However, we find that the implicit features are always very near to the explicit features in Chinese, such as "这个沐浴露的价格很便宜" (The price of this body wash is very expensive), "它的价格很贵" (The cost is expensive), according to this phenomenon. We extend association rules method by applying collocation selection method to find the co-occurrence words, including frequency, PMI, frequency multiply PMI (Manning & Schütze, 1999) to find implicit features from the corpus.

PMI is a good measure of independence, and it is widely used in natural language processing problems such as discover collocations. In our case the occurrence of explicit feature f and potential implicit feature w is as follows:

$$PMI(f, w) = \log_2 \frac{P(f, w)}{P(f)P(w)} \tag{1}$$

where $P(f,w)$ is the joint probability of the co-occurrence of the feature f and the potential implicit feature w in a sentence, $P(f)$ is the probability of the feature f occurrence in sentences, and $P(w)$ is the probability of the potential implicit feature w. We know that PMI cannot handle the low frequency words well (Manning & Schütze, 1999), because the low frequency words always get a very high PMI value. However, the low frequency words have less confidence to ensure the word are really the implicit features of the aspect. An improved usual solution is to combine the frequency and PMI measure together, and the notation is the same as the Eq. (1):

$$frequency * PMI(f, w) = P(f, w) * \log_2 \frac{P(f, w)}{P(f)P(w)} \tag{2}$$

The collocation selection method helps us find the implicit feature words by using frequency multiply PMI methods, the threshold is 15.0 in our case, and the result is as following in Table 4.

The feature extraction and grouping methods help to find explicit and implicit feature words. However, the result needs to be checked manually, to delete the unrelated noise result, taking up about less than 30% of all the candidates feature words. So Weakness Finder can efficiently assist us to find feature words for the given aspects.

3.4. Sentiment analysis

3.4.1. Sentence level sentiment analysis

Sentiment analysis is very important for Weakness Finder, because it can tell us whether the customer is satisfied with the aspect of the product or not. Our dataset is in a single domain (Section 2.2), and consists few sentences, so we apply an easy but effective sentence level sentiment analysis method to identify the orientation of each sentence. Our method is extending (Ding et al., 2008; Liu, 2010)'s work, because we take the adverbs of degree into consideration and handle the neutral sentence with negation and adverb of degree words with a different way to improve sentiment analysis performance. The pseudo code of the sentiment analysis method is shown in Algorithm 2, the algorithm works as follows:

1. **Identify the opinion words from sentiment word list**: The opinion word list consists of positive, negative opinion words and adverbs of degree, which is sorted out manually by us. So given a sentence, we can identify the sentiment words, and give them weights according to the word's polarity, we give the positive word score of +1 while the negative score of −1. For example, for given sentence, "The effect of moisturizing is not significant, but the cool feeling can last quite long time", because "significant" and "great" are both in positive words list, so we assign "The effect of moisturizing is not significant (+1), but the cool feeling can last quite long time".

Table 4
Implicit feature words mining by utilizing the collocation method.

Aspect	Price	Perfume	Degerm	Cleanness
Feature word 1	Cheap	Pure and fresh	Elimination	Clean
Feature word 2	Expensive	Thick flavor	Against	Wash and clean
Feature word 3	Preferential	Lavender		
Feature word 4	Reasonable	Lemon		

2. **Negation judgement**: Negation words can reverse the polarity of the words, so we reverse the polarity of sentiment words found in step 1. The sentence turns to be "The effect of moisturizing is not significant (−1), but the cool feeling can last quite long time". Because of the negation word "not" modified the positive words, so the polarity changes from +1 to −1. However, we found the neutral sentence which contains negation are always expressing a negative sentiment in Chinese, so we give the neutral sentence score −1.
3. **But-clauses**: No matter in Chinese or English, the transition words "but", "however","except" will imply the change of the authors' attitudes, the polarity before "but" and after "but" are opposite. So the sentence turned to be "The effect of moisturizing is not significant (−1), but the cool feeling can last quite long time (+1)".
4. **Adverbs of degree**: The intensity of polarity can be expressed by the adverbs of degree, we give these words different weight, for example, assign "quite" weight 2.0, assign "a bit", "not very" 0.5 and so on, if no such words, the weight is 1.0, then we multiply the weight to the original score to get the final polarity score. So the polarity of the sentence can be shown as follows by applying this rule, "The effect of moisturizing is not significant (−1), but the cool feeling can last quite long time (+2)". Actually the adverbs of degree are the signs of sentiment, and this is mainly because the sentiment words in the sentence are not in our sentiment words list. Therefore, we give the neural sentence with adverbs of degree the same polarity score with the former clause.

Finally we sum the polarity of all the polarity words together, if the final score is more than 0, we judge the final sentiment of the review is positive, if less than 0, we judge the review as negative, otherwise, if the score equals 0, it's a neutral review. For the sentence above, the final score is +1, so we judge it as a positive sentence.

Algorithm 2: Sentence level sentiment analysis algorithm

for each word w_j in sentence *Sen* **do**
  initialize the sentiment score $P(w_j) \leftarrow 0$
end for
for each word w_j in sentence *Sen* **do**
  if the word $P(w_j)$ in sentiment words list *SL* **then**
    set $P(w_j) \leftarrow$ (positive: 1 and negative: -1)
  end if
  if the word w_j in negation words list *NL* **then**
    for w_i in the same clause with w_j **do**
      set $P(w_i) \leftarrow -1 \times P(w_i)$ **do**
    end for
    if all words w_i in the same clause with w_j and w_i not in *SL* **then**
      set $P(w_{|Clause|}) \leftarrow -1$
    end if
  end if
  if the word w_j in transition words list *TL* **then**
    for w_i in the same clause with w_j and w_i not in *SL* **do**
      set $P(w_{|Clause|}) \leftarrow$ reverse value of the former clause

* (*continued*)

Algorithm 2: Sentence level sentiment analysis algorithm

    end for
  end if
  if the word w_j in adverbs of degree list *DL* **then**
    for w_i in the same clause with w_j and w_i in *SL* **do**
      set $P(w_i) \leftarrow degree(w_j) \times P(w_i)$
    end for
    if all words w_i in the same clause with w_j and w_i not in *SL* **then**
      set $P(w_{|Clause|}) \leftarrow$ former clause is positive: *1*, negative: -1
    end if
  end if
end for
for each word w_j in sentence *Sen* **do**
  score $\leftarrow \sum_{j=1}^{n} P(w_j)$
end for

3.5. Find weakness from product itself

The aspects of each review can be easily found by using the features grouping method, and the polarity of them can be calculated via sentence based sentiment analysis. Therefore, the positive, neural and negative sentences for each aspect can be sorted out when we evaluate the reviews of a brand product. Table 5 shows the result of Brand A body wash product. It is clear that the aspect "*perfume*", "*moisturizing*" and "*easy to wash*" were discussed most in reviews, and 70.59 % of "*perfume*" are discussed with a positive sentiment, while only 10.0% of the reviews are not satisfied with it. However, for the aspect "*easy to wash*", there are only 53.98% of reviews are positive, and 23.45% are negative. Obviously "*easy to wash*" is the weakness of the product according to the result among the three hottest issues.

In our system, we give thresholds 20% for the positive ratio and 60% for the negative ratio. So we underline the aspect whose negative ratio is more than 20.0% or the positive ratio is less than 60.0%, the conclusion is shown in Table 5, and the result is obviously shown that "*easy to wash*", "*dilute*" and "*dense*" are with both underlined ratio, it indicates the more customers are not satisfied with these aspects and less customers hold positive opinions about them, so they are the weaknesses of the products, the manufacturers could concentrate on. And the ratio of "*degerm*", "*effect*" are underlined, it tells us the "*degerm*" and "*effect*" are also the potential weaknesses of the product.

3.6. Find weakness when compared with competitors' products

In Section 3.5, we only compare the aspects of one product itself. It shows the weaknesses for a specific product. And another perspective is to compare the product with its competitors' product in order to know which aspect is weaker when compared with competitors' products. So in this part, we will compare the Brand

Table 5
The results for each aspect in Brand A body wash reviews.

Category	Aspect	Sentence number	Sentiment result			Sentiment ratio	
			Positive	Neutral	Negative	Positive ratio (%)	Negative ratio (%)
C1	Perfume	408	288	79	41	70.59	10.05
	Moisturizing	239	167	34	38	69.87	15.90
	Easy to wash	226	122	51	53	53.98	23.45
	Cleanness	188	123	30	35	65.43	18.62
	Degerm	34	23	2	9	67.65	26.47
	Dilute	16	7	4	5	43.75	31.25
	Dense	7	3	1	3	42.86	42.86
C2	Effect	103	62	19	22	60.19	21.36
	Comfort	89	72	8	9	80.90	10.11
C3	Price	47	37	7	3	78.72	6.38
	Advertisement	32	12	19	1	37.50	3.13

Table 6
The result comparison between Brand A's and Brand B's reviews.

Category	Aspect	Brand	Sentiment result			Sentiment percentage	
			Positive	Neutral	Negative	Positive ratio (%)	Negative ratio (%)
C1	Perfume	Brand A	288	79	41	**70.59**	**10.05**
		Brand B	458	108	44	74.92	7.38
	Easy to wash	Brand A	122	51	53	**53.98**	**23.45**
		Brand B	251	143	55	55.90	12.25
	Moisturizing	Brand A	167	34	38	**69.87**	**15.90**
		Brand B	192	20	24	81.36	10.17
	Degerm	Brand A	23	2	9	67.65	26.47
		Brand B	0	0	0	–	–
	Cleanness	Brand A	125	30	35	**65.43**	18.62
		Brand B	86	2	31	72.27	**26.05**
	Dense	Brand A	3	1	3	42.86	**42.86**
		Brand B	7	7	3	**41.18**	17.65
	Dilute	Brand A	7	4	5	43.75	31.25
		Brand B	0	0	1	**0.00**	**100.00**
C2	Effect	Brand A	62	19	22	60.19	21.36
		Brand B	52	28	24	**50.00**	**23.08**
	Comfort	Brand A	72	8	9	**80.90**	**10.11**
		Brand B	94	6	4	90.38	3.85
C3	Advertisement	Brand A	12	19	1	**37.50**	3.13
		Brand B	7	7	1	46.67	**6.67**
	Price	Brand A	37	7	3	**78.72**	6.38
		Brand B	80	7	7	85.11	**7.45**

Table 7
The evaluation result of features identification and grouping method.

Aspect name	Sentences number	Precision (%)	Recall (%)	F1-measure (%)
Perfume	1018	77.22	92.07	83.99
Easy to wash	675	86.82	100.00	92.95
Moisturizing	475	99.54	99.31	99.43
Comfort	325	97.71	99.61	98.65
Cleanness	307	100.00	98.91	99.45
Effect	207	99.31	100.00	99.65
Price	141	96.57	96.57	96.57
Advertisement	47	97.06	97.06	97.06
Degerm	34	100.00	100.00	100.00
Dense	24	94.29	100.00	97.06
Dilute	17	70.00	100.00	82.35
Total	3270	89.15	97.02	92.92

A's product result with another competitor's result. And their comparison result is shown in Table 6.

Each aspect of two brands' products can be compared from Table 6, so we mark the lower positive ratio and the higher negative ratio as bold. Obviously, the aspects of the brand with bold mean that the aspects are weaker than those of another brand. The result is clearly shown that the aspects "*perfume*", "*easy to wash*", "*moisturizing*" and "*comfort*" of Brand A are weaker than those of Brand B. Especially for the "*easy to wash*", the negative ratio (23.45%) is almost twice higher than Brand B's (12.25%), so we figure out that "*easy to wash*" is a serious weak point of Brand A products.Besides, "*moisturizing*" whose negative ratio is 15.90%, while Brand B's is

Table 8
The evaluation result of sentence level sentiment analysis.

Aspect name	Sentences number	Precision (%)	Recall (%)	F1-measure (%)
Perfume	1018	81.26	87.46	84.25
Easy to wash	675	80.43	84.00	82.17
Moisturizing	475	86.80	86.80	86.80
Comfort	325	90.40	92.24	91.31
Cleanness	307	78.29	77.39	77.84
Effect	207	83.86	83.86	83.86
Price	141	85.62	86.75	86.18
Advertisement	47	79.41	79.41	79.41
Degerm	34	80.00	80.00	80.00
Dense	24	80.00	80.00	80.00
Dilute	17	45.00	64.29	52.94
Total	3270	82.62	85.26	83.92

Chinese	English	Chinese	English
广告	advertisement	贵	expensive
买得起	affordable	清香	faint scent
抵抗	against	泡沫	foams
抑菌	antibacterial	清爽	fresh and cool
芳香	aromatic	果香	Fruity flavor
细菌	bacterial	高价	high-price
便宜	cheap	薰衣草	lavender
干净	clean	柠檬	lemon
洁肤	cleaning	喜欢	like
清洁度	cleanliness	滋润	moisturizing
清洁	cleanness	平价	parity
洁净	cleanse	香味	perfume
舒适度	comfort	实惠	preferential
价钱	cost	价位	price level
划算	cost-saving	价格	price
除菌	degerm	购买	purchase
稠	dense	清新	pure and fresh
稀	dilute	合理	reasonable
肮脏	dirty	杀菌	sterilize
去菌	disinfect	香气	sweet smell
易冲洗	easy to wash	太冲	thick flavor
效果	effect	冲洗	wash
去除	elimination	洗净	wash and clean

Fig. 5. Chinese English bilingual words translation.

10.17%, and positive ratio is 69.87% and Brand B's is 81.36%, so the "*moisturizing*" is also a weak point of Brand A's product.

From the category, we can see that the product properties (C1) of Brand A is a bit weaker than Brand B's, and from the comprehensive feedbacks (C2), the body wash of Brand A has a better effect but a bit worse comfort. From the marketing strategy (C3), the result implies that the customers are fed up with the advertisement, and less customers show their positive attitude while they are talking about Brand A products, and it can also imply that the Brand A's product is always moderate in marketing, there are less people complaining their price and advertisement, but also less people show their positive opinions on those aspects.

4. Evaluation

Since there are no existing benchmark datasets available, we evaluate Weakness Finder systems with the labeled dataset of two brands' products reviews which we introduced in Section 2.2. Weakness Finder employs aspect based sentiment analysis methods to find the weakness. Therefore, the evaluations could be divided into feature grouping evaluation and sentence level sentiment analysis evaluation. And we give out the precision, recall and F1 measure of the above methods for every aspect. So the precision of aspect A_i could be calculated by Eq. (3).

$$precision(A_i) = \frac{F(A_i)}{N(A_i)} \tag{3}$$

where $F(A_i)$ is the correct number about aspect A_i found by Weakness Finder, and $N(A_i)$ is total number of the words which have been found as aspect A_i by Weakness Finder.

$$recall(A_i) = \frac{F(A_i)}{M(A_i)} \tag{4}$$

where $M(A_i)$ is total number of aspect A_i labeled in our dataset in Eq. (4). And F1-measure combines precision and recall, which is shown in Eq. (5).

$$F1 - measure(A_i) = \frac{2 \times precision(A_i) \times recall(A_i)}{precision(A_i) + recall(A_i)} \quad (5)$$

4.1. Feature grouping evaluation

Table 7 shows the overall evaluation result of each aspect when applying our proposed features identification and grouping methods. The result shows that our proposed method has a good performance to find the explicit and implicit features, the average precision is 89.15% and the recall is 97.02%. The result of Degerm aspect has 100% precision and recalls, it's because "*germ*" in Chinese is a single character and has the only single meaning, so all the words contain "*germ*" are all referring to degerm. Instead, "*dilute*", "*perfume*" and "*wash*" are not a single-character words in Chinese, and they are always in compound words, so the precision is a bit lower than others. In general, the proposed feature identification and grouping methods have a good performance on the dataset.

4.2. Sentiment analysis evaluation

Table 8 depicts the evaluation result of sentence level sentiment analysis. The general precision is 82.62%, and the recall is 85.26%, F1-measure is about 83.92%. For most of the cases, the proposed methods have a good performance. However, it performs poorly on dilute aspect, and the reason is that the current algorithm is not taking the context information of "*dilute*" into consideration, and "*dilute*" will have different meanings in different contexts.

5. Conclusion and future work

In this paper, we propose an expert system Weakness Finder by analyzing the customers' reviews on the influential web communities with aspects based sentiment analysis, it can help the cosmetic manufacturers to find the weakness of the products in order to improve their products. The Weakness Finder system can help to identify the features, and group the features into different aspects by using explicit and implicit features grouping methods, then judge the polarity of each sentence by using sentence-level sentiment analysis. Methods of explicit features and implicit features identification for Chinese reviews are introduced, because the Chinese words are ideographic, so we use the sharing morpheme method and the similarity measure of Hownet to find the frequent and infrequent explicit features which describe the same aspect. Then use PMI $*$ frequency method helps to find implicit feature words, the final report show that the aspect "*easy to wash*" might be a big weakness point of the Brand A's body wash products, because more customers express their negative attitude while they are discussing the aspect "*easy to wash*", no matter when we compare from the single product's reviews alone or compare with their competitor's product reviews. In our current system, the aspect identification matches the words mechanically, so in the future we will take the words with multi-meanings identification and coreference resolution into consideration to improve the performance of the system.

Acknowledgments

This work is supported by the National Basic Research Program of China (973 Program, Grant No.: 2012CB316300) and the National Natural Science Foundation of China (Grant Nos.: 60875073, 61175110).

Appendix A

In this paper, most Chinese words are presented as its English translation with double quote and italic. And in this section, the Chinese English bilingual words translation is shown in Fig. 5.

References

Alvarez, M., & Lim, S. (2007). A graph modeling of semantic similarity between words. In *International conference on semantic computing*(pp. 355–362).
Bollegala, D., Matsuo, Y., & Ishizuka, M. (2007). Measuring semantic similarity between words using web search engines. In *Proceedings of WWW* (Vol. 7, pp. 757–786).
Ding, X., Liu, B., & Yu, P. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the international conference on web search and web data mining* (pp. 231–240). ACM.
Dong, Z., & Dong, Q. (2006). *HowNet and the computation of meaning*. World Scientific.
Fellbaum, C. (1998). *WordNet: An electronic lexical database*. The MIT Press.
Hughes, T., & Ramage, D. (2007). Lexical semantic relatedness with random graph walks. In *Proceedings of EMNLP-CoNLL* (pp. 581–589).
Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the national conference on artificial intelligence, 1999* (pp. 755–760). Menlo Park, CA, Cambridge, MA, London: AAAI Press, MIT Press.
Hu, M., & Liu, B. (2006). Opinion extraction and summarization on the web. *Proceedings of the national conference on artificial intelligence, 1999* (Vol. 21, pp. 1621). Menlo Park, CA, Cambridge, MA, London: AAAI Press; MIT Press.
Lin, D. (1998). Automatic retrieval and clustering of similar words. *Proceedings of the 17th international conference on computational linguistics* (Vol. 2, pp. 768–774). Association for Computational Linguistics.
Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing* 978-1420085921.
Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on world wide web* (pp. 342–351). ACM.
Manning, C., & Schütze, H. (1999). *MITCogNet. Foundations of statistical natural language processing* (Vol. 59). MIT Press.
Moghaddam, S., & Ester, M. (2010). Opinion digger: An unsupervised opinion miner from unstructured product reviews. In *Proceedings of the 19th ACM international conference on information and knowledge management* (pp. 1825–1828). ACM.
Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of EMNLP* (Vol. 4, pp. 412–418).
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.
Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on empirical methods in natural language processing* (Vol. 10, pp. 79–86). Association for Computational Linguistics.
Popescu, A., & Etzioni, O. (2005). Extracting product features and opinions from reviews. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 339–346). Association for Computational Linguistics.
Rushdi Saleh, M., Martín-Valdivia, M. T., Montejo-Ráez, A., & Ureña-López, L. A. (2011). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications, 38*(12), 14799–14804.
Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications, 36*(7), 10760–10773.
Yang, D., & Powers, D. (2005). Measuring semantic similarity in the taxonomy of wordnet. *Proceedings of the twenty-eighth australasian conference on computer science* (Vol. 38, pp. 315–322). Australian Computer Society, Inc..
Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. *Proceedings of the 2003 conference on empirical methods in natural language processing* (Vol. 10, pp. 129–136). Association for Computational Linguistics.
Zhai, Z., Liu, B., Xu, H., & Jia, P. (2011). Clustering product features for opinion mining. In *Proceedings of the fourth ACM international conference on web search and data mining* (pp. 347–354). ACM.
Zhan, J., Loh, H., & Liu, Y. (2009). Gather customer concerns from online product reviews-a text summarization approach. *Expert Systems with Applications, 36*(2), 2107–2115.
Zhang, H., Yu, H., Xiong, D., & Liu, Q. (2003). Hhmm-based chinese lexical analyzer ictclas. *Proceedings of the second SIGHAN workshop on Chinese language processing* (Vol. 17, pp. 184–187). Association for Computational Linguistics.
Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications, 38*(6), 7674–7682.