An Improved Association Rule Mining Approach to Identification of Implicit Product Aspects

Wei Jiang[1,2,*], Hao Pan[1] and Qing Ye[1]

[1]*College of Computer Science and Technology, Wuhan University of Technology, Wuhan, 430070, China;* [2]*Department of Information Engineering, City College of Wuhan University of Science and Technology, Wuhan, 430083, China*

Abstract: With the rapid development of Web 2.0, there has emerged a large number of product reviews written by users with their subjective views on online community, blog and e-commerce website. In product reviews, users are mostly concerned about the comments on a certain aspect or feature of the product, so the fine-grained opinion mining on product aspects is the current research focus. The early researches in this field focused primarily on explicit aspects but neglected implicit aspects. However, the implicit aspects expressed implicitly by a few words or phrases are of equally great importance and also can express the user's opinions and help us better understand user's comments. The identification of implicit aspect in product review text is a very challenging but significant subtask for opinion mining. This paper focus on identification of implicit aspects in product reviews. Aiming at deficiencies of previous works, this paper presents a novel association rule set mining approach that considers not only adjectives as indicator for the task, which integrates the improved collocation extraction algorithm with the topic model approach, and extracts reasonable and effective association rules from product review corpus as much as possible through the two complementary approaches, thereby enhancing the effect of identifying implicit aspects. Experiments show that the presented method is effective, and in the case that the precision does not reduce, the overall performance improves by about several percents compared to the baseline approach.

INTRODUCTION

With the rapid development of Web 2.0, there have emerged a large number of product reviews wrote by users with their subjective views on online community, blog and e-commerce website. However, with the rapid growth of reviews, the number of the reviews centering on a certain product may reach 1,000 or more, making it difficult for those potential buyers to quickly browse all the reviews and have a very comprehensive understanding of a product or even be misled by some non-objective reviews. Therefore, in recent years, the relevant researches on how to conduct opinion mining and generate a summary have attracted the attention of more and more researchers [1].

In product reviews, users are most concerned about the comments on a certain aspect or feature of the product, so the fine-grained opinion mining on product features is the current research focus [2, 3]. Ding *et al.* [4] gives the relevant definition, *i.e.*, using aspect to represent product part, attribute or function. An aspect is called explicit aspect if it occurs in the form of a noun or noun phrase, and a sentence containing an explicit aspect is called explicit sentence; if the aspect does not occur in the sentence, but is expressed implicitly by a few words or phrases, the aspect that is expressed implicitly is called implicit aspect, and the sentence that adopts this expression approach is called implicit sentence.

In general, the aspect-level opinion mining consists of three main tasks, namely:

(1) Extract product aspects from the review document, *i.e.*, identify the expression purpose of user's opinion. The content this paper studies focuses on this task;

(2) Identify corresponding opinions with product aspects, *i.e.*, mine the opinions related with each product aspect;

(3) Determine the sentiment polarity (positive or negative) of the corresponding opinions with product aspects, and this task is similar to sentiment classification task.

In view of the complexity of human language expression, the identification of implicit aspect in product review text is a very challenging difficulty, so the early researches in this field focused primarily on explicit aspects but neglected implicit aspects. However, the implicit aspects expressed implicitly by a few words or phrases are of equally great importance and also can express the user's opinions and help us better understand user's comments. For example, the comment sentence "This mobile phone looks very beautiful, but is a little expensive", where "a little expensive" implies the product aspect "price" Moreover, it is found from the statistics of a large number of online reviews that at least 30% of the comment sentences are implicit sentences, and that each such sentence contains at least one implicit aspect.

The identification of implicit feature in product review text is a very challenging difficulty. In the past ten years, many researchers have carried out researches on the extraction approach of implicit aspects. Su *et al.* [5] identified im-

plicit aspects by calculating point mutual information (PMI) based on semantic association analysis, but their work did not give a quantitative experiment result. Hai *et al.* [6] identified implicit aspects through the co-occurrence association rule mining technology, but they only considered the corresponding rules of sentiment words and product aspect words. However, in many cases especially in the context of the network, a large number of non-sentiment words are used to describe product aspects, so their approach will reduce the overall recall. Aiming at these deficiencies, this paper presents a novel association rule set mining approach, which integrates the improved collocation extraction algorithm with the topic model approach, and extracts reasonable and effective association rules from product review corpus as much as possible through the two complementary approaches, thereby enhancing the effect of identifying implicit aspects. Experiments show that the presented method is effective, and in the case that the precision does not reduce, the overall performance improves by about several percent compared to the baseline approach.

RELATED WORKS

Identification of aspect words and sentiment words plays an important role in the fine-grained sentiment analysis. Domain-related aspect lexicon and sentiment lexicon can be constructed by making use of the extraction of aspect words and sentiment words. In addition, a visual product review summary can be generated based on attribute words and sentiment words if aspect words and sentiment words are extracted and the correspondence between them are correctly identified. In this paper, our work presents an approach to implicit aspect identification. The approach is a sub-task in opinion mining and related to aspect extraction as well as topic modeling technology. In next section, we will conduct a brief literature review on the two approaches used in this paper.

The basic idea of Hu *et al.* [3] is to identify product aspects by imposing constraints on the high-frequency nouns or noun phrases, find the commented product aspects through the association mining approach, and extract the most adjacent adjectives as the sentiment words of this aspect, thus generating a product attribute comment summary. Zhu *et al.* [7] proposed a statistical approach based on C-value measurement to conduct aspect extraction. Chong *et al.* [8] proposed an approach based on frequency and information distance to conduct aspect extraction. Thus it can be seen that many researchers have carried out various improvements on the approach proposed by Hu *et al.* [3] and achieved a number of opinion mining systems since 2005, which indicates that simplicity and effectiveness is the greatest advantage of the frequency-based approach in practical application, but the limitation of this approach lies in that it may generate too many non-frequent aspects and miss low-frequency aspects, thereby reducing the mining effect.

The basic idea of LDA [9] is to think that each review is consisted of several topics, and each topic is consisted of several words. In aspect-level opinion mining, it is thought that the commented aspect corresponds with the topic. Titov *et al.* [10] proposed MG-LDA topic model (Multi-Grain LDA), achieving the extraction of local topics, and on the basis of extending MG-LDA model framework, further pro-

posed MAS model [11] (Multi-Aspect Sentiment model), thus achieving the connection from the local topic to the corresponding sentiment text description. Jo *et al.* [12] proposed SLDA model (Sentence-LDA) and assumed that all the words in a sentence are from the same aspect. The aspects identified by SLDA can match the comment details. Finally, on the basis of SLDA, they further modeled the associated sentiment aiming at different aspects and proposed ASUM model (Aspect and Sentiment Unified Model) which can capture the most important part of review, namely, such kind of structural information as <aspect, emotional>.

Zhai [13] *et al.* studied how to add pre-existing domain knowledge to LDA model and how to use the knowledge of these domains to guide LDA model to generate more meaningful topic grouping. They adopted Must-Links and Cannot-Links to express domain knowledge and generate constraints on LDA model and then proposed Constrained-LDA model. Experiment results show that the semi-supervised topic model which takes constraints as prior knowledge is better than the unsupervised topic model in performance.

The approach presented in this paper, by introducing Constrained-LDA, further finds reasonable association rules from low-frequency aspect indicators so as to extend the basic rule set, thus further enhancing the effect of implicit aspect identification.

IMPROVED ASSOCIATION RULE MINING APPROACH

Framework

This paper studied a novel association rule set mining approach to identify implicit aspects. The association rule extraction process is mainly divided into two steps:

1) Make use of the improved matched and selected calculation method to extract basic rules which join in the rule set in the form of *aspect indicator→aspect* as Basic Rule Set (written as: Basic Rules, BR);

2) Use a semi-supervised learning LDA topic model to extract new rules, extend basic association rule set, and finally get a new rule set (written as: the Model Rules+Basic Rules, MBR).

This approach first extracts candidate aspect indicators according to participle (sentence segmentation), POS tagging and aspect grouping, and then uses the improved collocation extraction algorithm to calculate the association degree between candidate aspect indicators and aspects. Each indicator and its corresponding aspect constitute a basic association rule set in the form of *aspect indicator→aspect*. Then, LDA topic model is adopted to further extract reasonable rules from low-frequency co-occurrence frequency aspect indicator words and non-indicator words so as to extend the basic rule set, and implicit aspects in implicit sentences in product reviews are identified through the finally obtained rule set.

Improved Matched and Selected Calculation Approach to Extract basic Rules

Firstly, frequent item set algorithm and some manual operation approaches are used to extract product aspect set. Considering the factor of synonyms, grouping is conducted on the synonyms in a set. Aspect extraction and synonym

Table 1.   Aspect word *f* and candidate indicator word *w*.

Aspect Word	Candidate Indicator		
	Present	**Absent**	**Total**
present	a	b	$a+b$
absent	c	d	$c+d$
total	$a+c$	$b+d$	N

grouping are not elaborated as they are not the focus of the research in this paper. Next, the sentences containing these aspect words or phrases in corpus are taken as an explicit sentence set, and candidate aspect indicators are extracted through POS tagging and the minimum occurrence frequency. In Ref. [6], only adjectives are selected as the candidate aspect indicators, on which this paper carries out extensions, *i.e.*, select POS tags including verb, noun, adjective, adverb, quantifier and pronoun as candidate aspect indicators. In addition, it is found that, through the observation of a lot of corpora, the aspect of a product is often expressed implicitly by two words at most. For example, in the sentence "This mobile phone responses quickly, but needs to be charged every day". "Day" and "charge" occur in the cell phone review sentence, and are often used to describe the battery capacity, so they can constitute a two-dimensional frequent item. Therefore, in this paper, the words constitute a two-dimensional frequent item if their speech is noun, adjective, adverb, quantifier and pronoun, and also are taken as the candidate indicators. Finally, the weight of each indicator is calculated based on the co-occurrence frequency between words or phrases and aspects.

In the natural language processing, identification and collocation extraction of implicit aspects are similar in a sense, and they are to find the co-occurrence frequency between words since the co-occurrence frequency measurement is the important basis to determine whether two words are matched. For convenience, the counts and symbols of the correlation between aspect word *f* and candidate indicator word *w* are given at first. As shown in Table **1** below, *a* refers to the number of times both aspect word and candidate indicator word occur, *b* and *c* respectively refer to the number of times that either aspect word or candidate indicator word occurs, *d* refers to the number of times that neither of aspect word or candidate indicator word occurs, and *N* refers to the times of total words in corpus. Therefore, obviously, *d=N-(a+b+c))*.

This paper first improves the PMI calculation approach and adds co-occurrence frequency into the calculation so as to reduce the dependence of this approach on corpus. The calculation formula is as follows:

$$\deg ree_{co-occurence} = \alpha FI(f,w) + \beta Co\sin e(f,w) + \gamma \chi^2(f,w) \tag{1}$$

$$FI(f,w) = frequence_{f\&w} PMI(f,w) = a \times \log_2 \frac{N \times a}{(a+b)(a+c)} \tag{2}$$

$$Co\sin e(f,w) = -\log_2 \frac{a}{\sqrt{(a+b) \times (a+c)}} \tag{3}$$

$$\chi^2(f,w) = \frac{N \times (a \times d - b \times c)^2}{(a+b)(c+d)(a+c)(b+d)} \tag{4}$$

Eq. (1) refers to the mixing co-occurrence frequency calculation formula, where α, β and γ refer to weight coefficients of three co-occurrence frequency calculation formulas respectively, and satisfy the condition α+β+γ=1. Eq. (2) refers to the improved PMI calculation formula into which co-occurrence frequency is added. Eq. (3) refers to the deformation of Cosine coefficient calculation formula. Eq. (4) refers to the χ^2 test. This paper concludes that, through many tests, a satisfactory result can be obtained when the weight coefficient values of α, β and γ are determined as 0.7, 0.1 and 0.2 respectively.

Pruning

According to the above approach, the co-occurrence frequency of each candidate indicator is calculated as the weight between indicator and its corresponding aspect, and then candidate indicators are extracted by setting appropriate threshold value and put into candidate indicator set. There are some conflicts in the set obtained finally--some indicators appear several times repeatedly in the indicator set, which requires proper pruning operation on the candidate indicator set as well as removal of redundant indicators. According to our observation, we argue that there are two reasons for this case:

(1) In the online product review expression, the general sentiment words such as "good" or "nice" can describe many aspects and be frequently used;

(2) Many explicit sentences contain more than one aspect, but in the co-occurrence frequency calculation process, it is not possible to determine which aspect the indicator describes, thus repeatedly calculating the indicator weight. For the second case, some conflict indicators may be reasonable actually, so these conflicts indicators cannot be directly deleted.

This paper adopts a simple algorithm to conduct pruning operation on indicator set. First, if the statistical weight is greater than the number of occurrence times of candidate indicator of given threshold value (written as: CFI), the threshold value is equal to the value used to select the candidate basic rule set. If the number of occurrence times is greater than 2, it is likely that the indicator corresponds with

redundancy rule and can be directly deleted from the indicator set. Otherwise, the indicator can be retained if it occurs twice. We take into account two cases in pruning algorithm: one is the minimum weight in the current indicator set, the other is the weight proportion in the indicator set. If the minimum weight is greater than the given threshold value (set it as 5 times of the threshold), the constitution rule of indicator and aspect is unreasonable. If the proportion is smaller than the pruning threshold value, the indicator can describe two corresponding aspects. If the above case happens, the indicator is likely to be inappropriate, and should be deleted from the candidate indicator set. Otherwise, the greater one may be more appropriate, and the smaller one should be deleted. The deleted indicator is saved to the pruning set (written as: PR). The specific algorithm is described and shown in Algorithm 1 below:

Algorithm 1: Pruning

```
for each cfi in CFI do
    if the value of cfi > threshold then
        occurrence = occurrence+1
    end if
end for
for each cfi in CFI do
if occurrence >2 then
cfi save in PR
delete cfi from CFI
else if occurrence =2 then
crf_max = select the bigger value indicator from CFI
crf_min = select the another indicator from CFI
ratio = crf_max / crf_min
    if value of crf_min > threshold *5
    OR ratio < pruning_threshold then
delete crf_max and crf_min from CFI
cfi save in PR
else
delete crf_min from CFI
end if
end if
end for
```

Adopt the Extraction Rule of Extended LDA Model

The basic rules obtained only by the above approach have great disadvantages. First, it is an issue of data imbalance. The number of explicit sentences of each aspect is not the same, which may result in the distribution imbalance of rule set. Second, there are often a lot of brand words or abbreviations of a part in product reviews. For example, "MOTO", "LED", etc. may be ignored in the calculation process. Finally, considering the flexibility of the network language, the opinion words of some indicators whose weight is smaller than the given threshold value and non-indicators can constitute reasonable rules with the corresponding aspects. As some POS tags may result in redundant

indicators, a lot of noise rules may form if the threshold value is set too low. Therefore, the best approach is to further excavate reasonable and effective rule set by introducing other approaches on the basis of setting the appropriate threshold value, and to supplement and perfect the current rule set.

Aiming at this issue, topic model is a good supplementary approach. Topic model provides a mechanism which can find word co-occurrence mode so as to identify the relevant topics. In the formation process of LDA, if the conditional probability value $p(w=v\,|\,z=i)$ of the calculated topic i to term v is very high, any sentence s containing term v will have higher probability of being related to the topic. In other words, as the co-occurrence frequency of two words increases, all terms that co-occur with term v are more likely to have been generated by topic i. Therefore, the topic words generated by LDA are the words with quite high co-occurrence frequency in the document, which means that topic words can be taken as the corresponding candidate indicators if topic words and aspects belong to the same topic grouping. In recent years, many researchers have extended the basic LDA model. More meaningful grouping combination can be found from the review text by the extended LDA model.

This paper adopts Constrained-LDA model proposed by Zhai *et al.* [15], establishes Constrained-LDA model through a given number of topics (*topic_num*) and the number of topic words (*word_num*). In order to make the first topic word in each topic as aspect word, and make full use of the co-occurrence of words, this paper makes some constraints when introducing Constrained-LDA topic model: first, the representation of synonymous aspect words. In many cases, the same meaning can be represented by different words, so we make the aspect use the same representation. Second, remove redundant sentences and indicators. The so-called redundant indicators refer to the removed conflict indicators. Similarly, redundant sentences refer to the sentences containing only redundant indicators. The removed redundant indicators such as "good" and "nice" can be used to modify a lot of words. The sentences containing these redundant indicators words should be deleted directly as they may cause interference with the sampling process. Third, must-links constraints. Make use of the basic rule set to get must-links set, *i.e.*, an indicator and its corresponding aspect must be assigned to the same topic in the sampling process if they are in the basic rule set.

After obtaining the grouping result through topic model (written as: CLDA), divide the topic words in the same topic according to the first word as aspect and get the topic word set (as: TW) which corresponds with aspect. Finally, there are still some topic words related to at least one aspects, and these words can be directly deleted by adopting the previous similar pruning operation approach. The topic word in each aspect grouping can be taken as candidate aspect indicator. A topic word should be filtered out by virtue of threshold value if it occurred previously. Otherwise, only acronym and numeral should be selected (numeral usually implies the product aspects such as price and screen, so it can be taken as the appropriate rule of indicator aspects). The specific algorithm is described and shown in Algorithm 2 below:

Table 2.   Experimental data.

Aspect	#Explicit Sentence	#Implicit Sentence	#Total
screen	1356	204	1560
price	827	105	932
speed	629	301	930
battery	738	94	832
appearance	572	132	704
touch	427	72	499
software	204	43	247

Algorithm 2:Extracting rules *via* constrained LDA

$CLDA\leftarrow$ Constrained-LDA(*topic_num, word_num*)

$TW \leftarrow$ partition *CLDA via* the first topic word

ConflictPrune(*TW*)

for each *w* in *TW* **do**

    if *w* in *CFI* **then**

        if the value of *w* > *threshold* **then**

        $MR \leftarrow$ Rule(*w,f*)

        else if *w* is foreign word OR number word **then**

        $MR \leftarrow$ Rule(*w,f*)

        end if

    end if

end for

$MBR \leftarrow BR + MR$

EXPERIMENT EVALUATION

Data Preparation

The review corpus of this experiment comes from the e-commerce website (www.amazon.com). Web oriented search engine is used to carry out crawling on the review content of the web mobile phone sector and 4,545 comments are finally obtained, from which 1,816 pieces with more review content are randomly selected as experimental corpus. OpenNLP is used to perform sentence segmentation, Porter stemmer is adopted to conduct stemming operation, Stanford POS tagger tool is used to conduct speech tagging, the part-of-speech of candidate indicator is filtered out by POS tagging and the part of speech that can be selected only includes: noun, verb and adjective. After pre-treatment, 5,704 sentences are obtained finally, which are tagged manually by two groups of tagging people, and then explicit sentences and implicit sentences are selected. After comparison, the sentences with inconsistent tags are deleted, and the remaining is taken as corpus, from which 7 representative product aspects are selected as shown in Table **2**. The part-of-speech of two-dimensional frequent items is limited to the following: noun, verb, adjective, pronoun, quantifier and numeral.

Experimental Setting and Evaluation Approach

In order to evaluate our approach, the rule set extracted and obtained by only adopting the mixing co-occurrence frequency calculation method is first given, then the experiment is conducted on corpus, and finally the corresponding experimental result is given. Afterwards, the extended topic model is added on the basis of the basic rule set obtained previously, and the experiment is conducted on corpus according to the rule set extracted again. Finally, the two experiment results are compared, and the rule sets used in the former and latter experiment processes are written as BR (basic rules) and MBR (model rules+basic rules).

Experimental evaluation labeling still uses the traditional precision, recall and F-score. In the extraction process of rules, the corresponding rule that the weight of aspect indicator is greater than the given threshold value is selected. The selection of different threshold values will lead to different results, so the performance of rule set is the best when an appropriate threshold value is selected. In the experiment process, we set the threshold within the interval [2, 80] and observe the results. In this paper, GibbsLDA++toolkit developed by Stanford university is modified, and then the extended topic model (Constrained-LDA) based on Gibbs Sample [14] is obtained. Parameters are set as $\alpha=1$, $\beta=0.01$, and the number of iterations is 1,000. Considering the result has certain randomness, the model operates for 30 times in the experiment, and the average value is taken as the experiment result. The number of topics in the topic model in the experiment should be set as 7 at least since 7 typical product aspects are selected from corpus for analysis.

Experimental Result and Analysis

The experimental result of using BR rule set and MBR rule set to identify implicit aspects are respectively shown in Figs. (**2** and **3**) below. It can be seen from Fig. (**1**) that the precision is also improved with the increase of the threshold value, and the precision achieves the best effect as 88.47% when the threshold value is set to 35.

In the experiment, we set the number of topic words in each topic increases in turn from 3 so as to make the finally obtained MBR rule set achieve the best effect. According to this setting as shown in Fig. (**3**) below, the best F score obtained is 73.08% when the number of the topic words in each
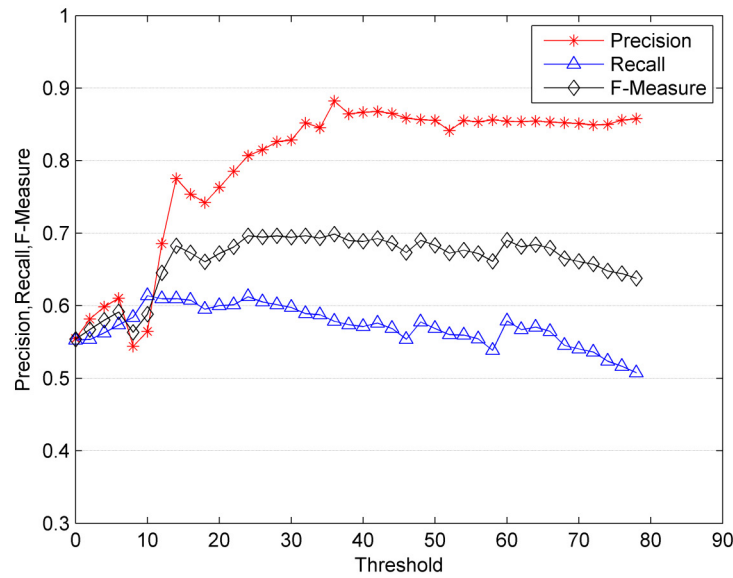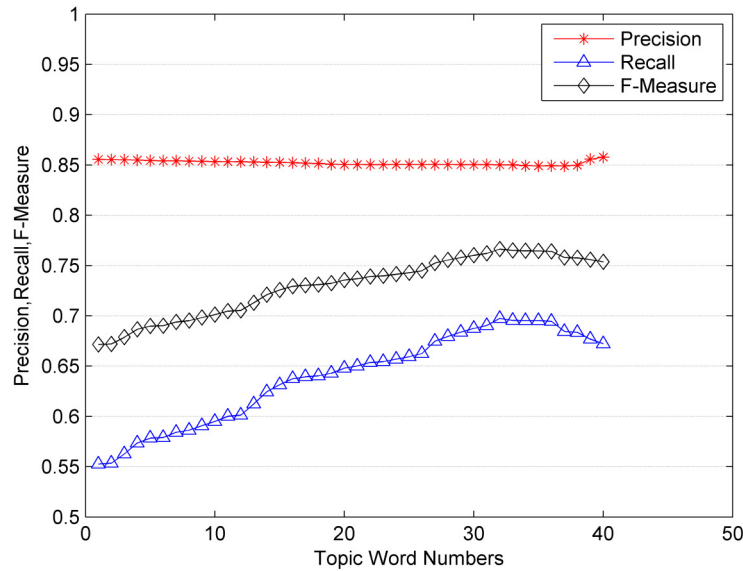
Fig. (1). Experimental result after using BR rule set.

Fig. (2). Experimental result after using MBR rule set.

Table 3. The comparison result.

Approach	Recall	Precision	F-Score
Su-PMI	0.7429	0.7661	0.7543
CO-AR	0.7462	0.7755	0.7606
CO+MODEL	0.7365	0.8257	0.7786

topic increases to 32. When the result obtained by adding topic model is compared with the case as shown in Fig. (**2**), the precision does not decrease but the performance increases by about 3%, which shows the approach is effective and the rule added through topic model is reasonable.

Compare the approach presented in this paper (CO+MODEL) with the approaches presented in Ref. [5] and Ref. [6] (denoted as Su-PMI and CO-AR, respectively), and the result is shown in Table **3** below. It can be seen that the approach presented by us is better than the other two approaches, and F-score increases by 3.2% and 2.4% respectively. Furthermore, a lot of association rules that are not captured in the baseline approach are extracted in the approach proposed by us, thereby improving the overall performance.

CONCLUSION

This paper presents a novel association rule set mining approach, which integrates the improved collocation extraction algorithm with the topic model approach, and extracts reasonable and effective association rules as much as possible through these two complementary approaches used to identify implicit product aspects in product reviews. This approach uses an improved extraction algorithm to calculate the association degree between candidate aspect indicators and aspects, then uses LDA topic model to further extract reasonable rules to extend basic rule set from low-frequency co-occurrence frequency aspect indicator words and non-indicator words, and identifies implicit aspects in implicit sentence in product reviews by the rule set obtained finally. The final experimental results show that our proposed method is effective, and the identification performance has been greatly improved compared to other rule sets that are calculated and obtained by simply using co-occurrence frequency.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. M. Kim and E. Hovy, "Determining the Sentiment of Opinions", in Proceedings of the 20th International Conference on Computational Linguistics, 2004, pp.1367-1373.

[2] A. M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews", Natural Language Processing and Text Mining, Springer London, 2007, pp.9-28.

[3] M. Q. Hu and B. Liu, "Mining and Summarizing Customer Reviews", in Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168-177.

[4] X. W. Ding, B. Liu and P. S. Yu, "A Holistic Lexicon-based Approach to Opinion Mining", in Proceedings of the first ACM International Conference on Web Search and Data Mining, 2008, pp. 231-240.

[5] Q. Su, K. Xiang, H. Wang, B. Sun and S. Yu, "Using Pointwise Mutual Information to Identify Implicit Features in Customer Reviews", in Proceedings of the 21th Computer Processing of Oriental Languages, 2006, pp. 22-30.

[6] Z. Hai, K. Y. Chang and J. Kim, "Implicit Feature Identification *via* Co-occurrence Association Rule Mining", in Proceedings of 12th International Conference on Computational Linguistics and Intelligent Text Processing, 2011, pp. 393-404.

[7] J. B. Zhu, H. Wang, B. K. Tsou and M. Zhu, "Multi-aspect Opinion Polling from Textual Reviews", in Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009, pp. 1799-1802.

[8] C. Long, J. Zhang and X. Zhut, "A Review Selection Approach for Accurate Feature Rating Estimation", in Proceedings of the 23rd International Conference on Computational Linguistics, 2010, pp. 766-774.

[9] D. Blei, A. Ng and M. Jordan, "Latent Dirichlet Allocation", the Journal of machine Learning research, vol. 3, pp. 993-1022, 2003.

[10] I. Titov and R. McDonald, "Modeling Online Reviews with Multigrain Topic Models", in Proceedings of the 17th International Conference on World Wide Web, 2008, pp. 111-120.

[11] I. Titov and R. McDonald, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization", in Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, 2008, pp. 308-316.

[12] Y. Jo and A. Oh, "Aspect and Sentiment Unification Model for Online Review Analysis", in Proceedings of the 4th ACM International Conference on Web Search and Data Mining, 2011, pp. 815-824.

[13] Z. Zhai, B. Liu, H. Xu and P. Jia, "Constrained LDA for Grouping Product Features in Opinion Mining", in Proceedings of the 15th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, 2011, pp. 448-459.

[14] T. Griffiths and M. Steyvers, "Finding Scientific Topics", Proceedings of the National academy of Sciences of the United States of America, 2004, pp. 5228-5235.