Sentiment Analysis and Mining of Product Review Text

Authors: Li Hanyu, Zhou Pengfei, Li Hanyu

Date: 2016-06-12T00:00:00+00:00

Abstract

With the rapid development of e-commerce, massive amounts of product review information have emerged on the Internet. Sentiment analysis and mining of product review texts are of significant value for researching product reputation and conducting product recommendations. This paper designs product attribute extraction and filtering algorithms and sentiment word discrimination algorithms to analyze product review information and automatically extract product attributes of user interest and users' evaluative opinions on corresponding attributes, further applying them to sentiment orientation analysis of product evaluation texts. Testing on a real-world dataset achieved an accuracy of 81.08% and a recall of 88.23%.

Full Text

Sentiment Analysis and Mining of Product Review Texts

HanYu Li, PengFei Zhou
(Information Systems Department, National Science Library, Chinese Academy of Sciences, Beijing 100190)

Abstract

With the rapid development of e-commerce, a massive amount of product review information has emerged on the Internet. Sentiment analysis and mining of product review texts hold significant value for studying product reputation and conducting product recommendations. This paper designs algorithms for product attribute extraction and filtering, as well as sentiment word discrimination. It analyzes product review information to automatically extract product attributes that users focus on and users' evaluation opinions on corresponding attributes, further applying these to sentiment orientation analysis of product review texts. Testing on a real dataset achieved an accuracy of 81.08% and a recall rate of 88.23%.

Keywords: Product Review Mining, Sentiment Orientation Analysis, Sentiment Word Polarity Determination, Data Mining
Chinese Library Classification: TP391

Funding Information: This work is supported by the National Natural Science Foundation of China (4123091).

Author Biographies:
Li Hanyu (born 1986), female, from Zhoukou, Henan, Master's degree, Assistant Librarian. Main research directions: information retrieval, natural language processing (lihy@mail.las.ac.cn).
Zhou Pengfei (born 1985), male, from Tangshan, Hebei, Master' s degree, Senior Engineer. Main research directions: data mining, sentiment orientation analysis (hb1gzpf@126.com).

1 Introduction

With the rapid development of e-commerce, an increasing amount of product review data has been generated on the Internet. Most of this review information contains users' subjective sentiments, harboring substantial value awaiting exploitation. Automated review text mining tools can save time on manual analysis of review data and offer strong practicality in improving efficiency [1][2]. The goal of sentiment analysis and mining for product review texts is to extract user-expressed opinions and sentiments from the text, identifying users' focus hotspots, which holds great practical value [3]. Manufacturers can discover product deficiencies and consumer needs based on user review information, thereby improving their products. Retailers can identify high-quality products and user preferences from user reviews, enabling them to recommend quality products to users to increase sales. Simultaneously, the extracted evaluation hotspots can be applied to product search ranking, allowing users to discover product strengths and weaknesses through summarized review information to find their most desired products.

Most current text sentiment orientation analyses do not consider the topic of the text. Document-level sentiment orientation analysis primarily focuses on analyzing the overall positive/negative tendency of a document [4][5], while word-level sentiment orientation analysis mainly studies the positive/negative tendency of individual words. Document-level sentiment orientation analysis mainly employs the following methods: (1) Dictionary-based methods, which use existing sentiment dictionaries to determine overall sentiment orientation based on the number of positive and negative sentiment words appearing in the document [6]; (2) Machine learning methods such as Support Vector Machines, Naive Bayes, Maximum Entropy, K-Nearest Neighbors algorithms, etc. [4]. Reference [4] treats sentiment orientation analysis as text classification, training classifiers with annotated training data and then using the classifiers to predict new samples, achieving good results. Word-level sentiment orientation research primarily includes the construction of sentiment dictionaries and the determination of sentiment word polarity [7][8][9]. Reference [10] provides a more detailed classification of sentiment words from a human perspective using aspects such as attitude, affect, and opinion. Most of this work is topic-independent and cannot meet user needs in many cases. For example, when users want to know which product functions consumers are most dissatisfied with, document-level sentiment analysis cannot provide answers.

This paper proposes a method for automatically extracting product attributes and their corresponding sentiment words, conducting sentiment orientation analysis of product reviews based on the extracted attributes and sentiment words. Section 2 introduces the overall system architecture, Section 3 describes the product attribute and sentiment word extraction module, Section 4 presents the product review sentiment orientation analysis module, Section 5 discusses experiments and result analysis, and finally, we conclude and propose directions for future research.

2 Sentiment Analysis and Mining System for Review Texts

This paper extracts product attributes and corresponding evaluation sentiment words from product review texts to obtain the modification relationships between attributes and sentiment words, determining the sentiment orientation of sentiment words when modifying product attributes, which is further used for sentiment orientation analysis. The overall system flow is shown in Figure 1 [Figure 1: see original paper]. This method does not require a seed attribute set; it obtains product attributes and sentiment words through their co-occurrence relationships, extracting more reliable product attribute and sentiment word information. Apart from inputting product review texts, this system requires no domain-related knowledge and needs no modifications. As long as new domain evaluation data is provided, it can be applied to other domains, demonstrating good portability.

3 Product Attribute and Sentiment Word Discovery

Most product review data is relatively brief, such as "The phone screen is very large," which evaluates only one attribute. Some review data is longer and may contain evaluations of multiple attributes. We analyze such data to discover product attributes and sentiment words. This paper uses the ICT-CLAS Chinese word segmentation system for segmentation and part-of-speech tagging. Product attributes in review data are primarily in noun form, such as "screen," "battery," and "sound." Therefore, nouns from the part-of-speech tagged word sequence are used as candidate attribute words. Similarly, sentiment words are generally adjectives, so extracted adjectives are used as candidate sentiment words. Noise data is filtered based on noun frequency and co-occurrence with sentiment words. The candidate product attribute and sentiment word extraction algorithm is listed in Table 1 .

Table 1. Candidate Product Attribute and Sentiment Word Extraction Algorithm

Input: Review phrase collection Segments = {seg1, seg2, ···, segn}, (frequency threshold).

Output: Product attribute set F = {f1, f2, ···, fn}, evaluation sentiment word set S = {s1, s2, ···, sm} corresponding to each product.

1. Initialize candidate product attribute set F as empty and candidate sentiment word set S as empty.
2. For each review phrase, extract noun phrases and add them to attribute set F.
3. For each review phrase, extract adjective phrases and add them to sentiment word set S.
4. For each element fi in F, if its frequency in the review data is lower than , remove it from the set.
5. For each element sj in S, if its frequency in the review data is lower than , remove it from the set.
6. Take any pair of elements <fi, sj> from F and S; if both elements appear simultaneously in a review phrase, these two elements form an evaluation pair. All evaluation pairs combine to form the evaluation pair set PairSet.
7. For all elements in F, if an element does not appear in PairSet, remove it from F.
8. For all elements in S, if an element does not appear in PairSet, remove it from S.
9. After the above filtering, obtain product attribute set F and sentiment word set S.

Figure 2. Example of Product Attribute and Sentiment Word Extraction Results

The system input is user review information for a certain product, such as reviews for a particular phone model on a shopping website, including the user' s review text and the star rating given to the evaluation. The system converts user review information into shorter review phrases through preprocessing, sentence splitting, word segmentation, and part-of-speech tagging. Based on part-of-speech tagging results, nouns are used as candidate product sentiment words, and adjectives are used as candidate sentiment words. According to the co-occurrence principle of product attributes and sentiment words, product attributes and sentiment words corresponding to each attribute are extracted. Simultaneously, the positive/negative orientation of sentiment words is obtained based on the star rating of the original review text. Partial attribute extraction and sentiment word extraction results are shown in Figure 2 [Figure 2: see original paper], where the left side shows extracted product attributes such as phone attributes including "screen," "battery," "memory," "price," etc., and the right side shows the sentiment words corresponding to attributes. For example, the sentiment words extracted for the "screen" attribute include "large," "clear," "sensitive," "garbage," etc.

Through the extraction of these product attributes and their corresponding sentiment words, a knowledge base for sentiment analysis can be established for products. This knowledge base can be used for sentiment orientation analysis of new product reviews. Establishing a correspondence between evaluation attributes and sentiment words can solve the problem where sentiment words have different sentiment orientations when modifying different attributes. Simultaneously, deep mining can be performed on this basis, such as identifying product attributes that users pay more attention to, finding product attributes that users are most dissatisfied with, and finding product attributes with the most negative reviews [15].

4 Sentiment Orientation Analysis of Product Reviews

4.1 Benchmark Sentiment Word Extraction and Orientation Determination from Review Data

Through analysis of evaluation data, many sentiment words such as "not bad," "very good," "very poor," and "not good" are found to appear frequently with clear sentiment orientation. Such sentiment words are extracted here for use as benchmark sentiment words. Since the sentiment orientation of extracted product evaluation data can be determined based on the evaluation star rating, we utilize the sentiment orientation of existing evaluation data to determine the sentiment orientation of sentiment words. The chi-square test (CHI) method is used to calculate the degree of association between sentiment words and positive/negative categories. A larger chi-square estimate indicates more reliable sentiment orientation of the sentiment word. CHI values are calculated separately for each sentiment word of each product attribute. The CHI value of a sentiment word in a category is calculated as in Formula 1.

$$\chi^2(t,c) = \frac{N \times (A \times D - B \times C)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)} \quad \text{(Formula 1)}$$

where N is the total number of documents in the training corpus, c is a certain category, t is a sentiment word, A is the number of documents belonging to category c and containing sentiment word t, B is the number of documents not belonging to category c but containing sentiment word t, C is the number of documents belonging to category c but not containing sentiment word t, and D is the number of documents not belonging to category c and not containing sentiment word t. The association degree between sentiment word t and both positive and negative categories is calculated separately, and the category with the larger association degree is taken as the sentiment orientation of sentiment word t.

For each product attribute, the corresponding sentiment words are extracted and sorted by frequency. Sentiment words with frequency and CHI values greater than the thresholds are extracted as benchmark sentiment words. Here, the frequency threshold is 0.1 and the CHI threshold is 0.9. The extraction of benchmark sentiment words emphasizes accuracy; using the extracted benchmark sentiment word set for further expansion will improve recall.

4.2 Sentiment Orientation Determination of Evaluation Phrases

The extracted benchmark sentiment word set is used to further expand the scale of the sentiment word collection. The PMI-IR algorithm [17] is primarily used to calculate the distance between sentiment words and benchmark sentiment words. The PMI-IR algorithm calculation formula is shown in Formula 2.

$$PMI(word_1, word_2) = \log_2 \frac{P(word_1 \& word_2)}{P(word_1) \times P(word_2)} \quad \text{(Formula 2)}$$

where $P(word_1\&word_2)$ represents the probability of $word_1$ and $word_2$ appearing simultaneously, and $P(word_1)$ represents the probability of $word_1$ appearing.

When calculating the sentiment orientation of a sentiment word, the distance between the word and all words in the positive and negative benchmark sentiment word tables can be calculated [17], as shown in Formula (3).

$$SO(word) = \sum_{pword \in Pset} PMI(word, pword) - \sum_{nword \in Nset} PMI(word, nword) \quad \text{(Formula 3)}$$

where $word$ is the new word whose sentiment orientation is to be calculated, $Pset$ is the positive sentiment word set of benchmark sentiment words, and $Nset$ is the negative sentiment word set of benchmark sentiment words.

This method can be used to calculate the sentiment orientation of sentiment words with clear sentiment orientation in specific domains. However, the same sentiment word may exhibit opposite sentiment orientations when modifying different product attributes. For example: "The phone screen is very large" and "The phone noise is very large." Here, "large" serves as a sentiment word for both "screen" and "noise" attributes. When used as a sentiment word modifying "screen," it is a positive sentiment word; when used as a sentiment word modifying "noise," it is a negative sentiment word. If "large" is calculated as a single sentiment word for distance to benchmark words, the result will be inaccurate regardless of the calculation outcome. Here, the PMI method is improved by no longer calculating only the distance between sentiment words, but instead calculating the distance between the <attribute, sentiment word> pair and benchmark sentiment words to obtain the sentiment orientation of the sentiment word within the <attribute, sentiment word> pair. In the above example, instead of calculating the PMI value between "large" and benchmark sentiment words, the PMI value between the evaluation phrase <screen, large> and the benchmark sentiment word set is calculated. The improved method is shown in Formula (4).

$$SO(feature, word) = \sum_{pword \in Pset} PMI(< feature, word >, pword) - \sum_{nword \in Nset} PMI(< feature, word >, nword)$$

This calculates the sentiment orientation of *word* when used as a sentiment word modifying the $feature$ attribute. $Pset$ and $Nset$ have the same meaning as in Formula (3).

4.3 Sentiment Orientation Analysis of Product Reviews

The extracted product attributes and sentiment words corresponding to each attribute can be used as a knowledge base for sentiment orientation analysis to conduct sentiment orientation analysis of product reviews [18]. For example, by extracting information from phone reviews, we can obtain phone product attributes, evaluation sentiment words corresponding to each attribute, and the sentiment orientation when using these sentiment words to evaluate attributes. The processing method when encountering a review text in the same domain is shown in Figure 3 [Figure 3: see original paper].

Figure 3. Sentiment Orientation Analysis of Product Reviews

For each evaluation phrase, evaluation attributes are extracted based on the knowledge base. For example, if the knowledge base already contains extracted attributes such as "screen," "battery," and "system," and the attribute word "screen" appears in the evaluation phrase, then "screen" is taken as the extracted attribute word.

For each evaluation phrase, evaluation sentiment words are extracted based on the knowledge base. The sentiment orientation of the evaluation phrase is determined according to the sentiment orientation of the combination of product attributes and evaluation sentiment words in the knowledge base. The sentiment orientations of all evaluation phrases are aggregated to derive the overall evaluation orientation.

Product attribute review opinion summarization: Based on the extracted product attributes and their corresponding evaluation sentiment words, user evaluations for a particular product can be summarized. For example, evaluations corresponding to price include "affordable," "cheap," "expensive," "appropriate," "reasonable," "not bad," etc. Through the discovered product attributes and corresponding sentiment words, evaluation opinions for each attribute can be obtained. Analyzing the frequency of various evaluation opinions can reveal user evaluation hotspots for attributes.

5 Experiments and Result Analysis

5.1 Experimental Data

The experimental data in this paper was obtained by crawling review data for some popular mobile phone products from the e-commerce website JD.com (www.jd.com). The user review data on the website includes review text information for products and star ratings given by users to product evaluations. The distribution of star ratings from 1 (low) to 5 (high) is shown in Table 2 .

Table 2. Distribution of Review Star Ratings in Experimental Data

5.2 Evaluation Metrics

This paper adopts accuracy, recall, and F-value as evaluation metrics. Accuracy is shown in Formula (5).

$$P = \frac{tp}{tp + fp} \quad \text{(Formula 5)}$$

where tp is the number of correct analysis results, and fp is the number of incorrect analysis results. For example, when evaluating the accuracy of product attribute extraction, tp represents the number of correctly extracted product attributes, and fp represents the number of items that the system extraction results consider as product attributes but are actually not product attributes.

Recall is shown in Formula (6).

$$R = \frac{tp}{tp + fn} \quad \text{(Formula 6)}$$

where tp is the number of correct analysis results, and fn is the number of correct items that were not identified. For example, when evaluating the recall of product attribute extraction, tp represents the number of extracted product attributes that are indeed product attributes, and fn represents the number of product attributes that appear in reviews but were not extracted.

Harmonic mean (F-value) is shown in Formula (7).

$$F = \frac{2 \times P \times R}{P + R} \quad \text{(Formula 7)}$$

5.3 Analysis of Product Attribute and Evaluation Phrase Extraction Results

In the experimental data, user-focused attributes are mainly concentrated in aspects such as "quality," "screen," "battery," "price," etc. Some extracted product attributes and evaluation phrases from the experiment are shown in Table 3 .

Table 3. Examples of Extracted Attributes and Evaluation Phrases

Positive Evaluation Phrases	Negative Evaluation Phrases
Screen large, screen clear, screen delicate, screen not bad, screen sensitive···	Screen small, screen dark, screen garbage···
Quality very good, quality reassuring, quality not bad, quality reliable···	Quality too poor, quality not good, quality no good···
Battery durable, battery sufficient, battery awesome, battery not bad···	Battery heating, battery not good, battery no good, battery very poor, battery not durable···
Price cheap, price affordable, price cost-effective, price reasonable···	Price expensive, price slightly high···
System old, system slow···	

For three products, each with 100 product review data entries, product attributes were extracted and manually evaluated. Using different   thresholds in the algorithm from Table (1), experimental results are represented using recall, accuracy, and harmonic mean, with results shown in Figure 4 [Figure 4: see original paper]. For three products, each with 100 product review data entries, evaluation phrases were extracted and manually evaluated. Using different   thresholds in the algorithm from Table (1), experimental results are represented using recall, accuracy, and harmonic mean, with results shown in Figure 5 [Figure 5: see original paper].

As can be seen from Figures 4 and 5, when the threshold   is small, more redundant data exists, resulting in lower accuracy and higher recall. When the threshold   is large, redundant data is reduced, but some correct results are also filtered out, leading to lower recall and higher accuracy. Good results are achieved when the threshold   value is 3. Through observation, it can be found that recall is relatively high because the same attribute information appears in many reviews. Although attributes in some reviews are not extracted, they can be compensated by extraction from other reviews, so the overall recall is relatively high. Accuracy is somewhat lower than recall, and further improvements to the algorithm can enhance accuracy.

5.4 Sentiment Orientation Analysis of Product Reviews

After extracting product attributes and sentiment words corresponding to each attribute, the extracted information can be used for sentiment orientation analysis. The analysis result for each review data is: which product attribute this review information evaluates and whether the sentiment orientation toward that attribute is positive or negative. Here, 1700 product review data entries are used to extract product attributes and sentiment words, determining the sentiment orientation of sentiment words when modifying product attributes. Then this knowledge base is used for sentiment orientation analysis. Experiments were conducted on review data for three products, with 100 reviews for each product undergoing sentiment orientation analysis. Manual evaluation results of the analysis are listed in Table 4 .

Table 4. Results of Sentiment Orientation Analysis of Product Reviews

Product	Accuracy	Recall	F-value
Phone Product 1	85.18%	79.31%	82.14%
Phone Product 2	85.71%	80.00%	82.76%
Phone Product 3	88.23%	81.08%	84.50%

As can be seen from the experimental results, the F-values of sentiment orientation analysis for different products are all above 80%, indicating that using this method for sentiment orientation analysis of product reviews is feasible. After conducting sentiment orientation analysis, different evaluation opinions are analyzed according to various product attributes, as shown in Figures 6 [Figure 6: see original paper], 7 [Figure 7: see original paper], and 8 [Figure 8: see original paper].

Phone 1 is a traditional Nokia feature phone. From the above, it can be seen that user evaluation hotspots reflect the product' s characteristics: the phone quality is very good, the price is cheap, the battery is sufficient, the screen is relatively small, functions are relatively few, and the system is relatively old. Users are very satisfied with quality and price, relatively satisfied with the battery, and not very satisfied with the screen, functions, and system.

Phone 2 is an entry-level smartphone. From the analysis results, it can be seen that users are very satisfied with price and functions, basically satisfied with quality and screen, and dissatisfied with battery and memory. Such analysis results basically reflect the characteristics of this phone.

Phone 3 is a large-screen smartphone. From the analysis results, it can be seen that users most frequently report that the screen is very good and the operating speed is very satisfactory; users provide the most feedback on these two aspects. Meanwhile, users point out that the battery and phone sound are not good. The analysis results well reflect the relative characteristics of each phone model and provide good reference value for users, merchants, and manufacturers.

6 Conclusion

This paper has implemented automated extraction of product attributes and evaluation sentiment words, achieving sentiment orientation analysis of product reviews with performance reaching over 80%. Experimental results demonstrate that the method achieves relatively high accuracy and recall. However, some aspects still require further improvement and refinement. Certain evaluation attributes and evaluation opinions are similar and need further clustering. For example, "price appropriate" and "price reasonable" should be merged into one. Through the work in this paper, comment mining regarding product attributes in product reviews has been completed, extracting review attributes and determining their positive/negative orientation. Future work can apply the sentiment orientation analysis results to recommend quality products to users and other applications.

References

[1] Hu Minqing, Liu Bing. Mining Opinion Features in Customer Reviews. Proceedings of the National Conference on Artificial Intelligence, 2004[C]. San Jose, USA: ACM Press, 2004: 755-760.

[2] Bing Liu, Minqing Hu, Junsheng Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. Proceedings of the 14th international conference on World Wide Web, 2005[C]. ACM, 2005: 342-351.

[3] Lisette García-Moya, Henry Anaya-Sánchez, Rafael Berlanga-Llavori. Combining Probabilistic Language Models for Aspect-Based Sentiment Retrieval. Proceedings of the 34th European Conference on Information Retrieval, 2012[C]. 2012: 561-564.

[4] B. Pang, L.L. Lee and S. Vaithyanathan. Thumbs up? Sentiment Classification Using Machine Learning Techniques. Proceedings of ACL-02 conference on Empirical methods in Natural Language Processing, 2002[C]. Volume 10, 2002: 79-86.

[5] Bo Pang, Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Proceedings of American Association for Artificial Intelligence Conference, 2004[C], 2004: 271–278.

[6] Zhang Chenggong, Liu Peiyu, Zhu Zhenfang, Fang Ming. A sentiment analysis method based on sentiment dictionary [J]. Journal of Shandong University, 2012, Vol. 47, No. 3.

[7] Yu, J, Zha, Z, Wang, M, Chua, T. Aspect ranking: Mining the peanut gallery: opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web, 2003[C]. ACM Press New York, 2003: 519-528.

[8] Vasileios Hatzivassiloglou, Kathleen R. McKeown. Predicting the Semantic Orientation of Adjectives. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, 1997[C]. 1997: 174-181.

[9] Zhu Yanlan, Min Jin, Zhou Yaqian, Huang Xuanjing, Wu Lide. Computation of lexical semantic orientation based on HowNet [J]. Journal of Chinese

Information Processing, 2006, 20(1): 14—20.

[10] Casey Whitelaw, Navendu Garg, Shlomo Argamon. Using Appraisal Groups for Sentiment Analysis. In Proceeding CIKM ' 05 Proceedings of the 14th Association for Computing Machinery international conference on Information and knowledge management, 2005 [C]. 2005: 625-631.

[11] Zhao Yanyan, Qin Bing, Liu Ting. Text sentiment analysis [J]. Journal of Software, 2010, 21(8): 1834-1848.

[12] Yao Tianfang, Cheng Xiwen, Xu Feiyu, Hans Uszkoreit, Wang Rui. A survey of text opinion mining [J]. Journal of Chinese Information Processing, 2008, 28(11): 2725—2728.

[13] Wu Xing, He Zhongshi, Huang Yongwen. Product feature extraction based on weakly supervised learning [J]. Computer Engineering, 2009, 13: 199-201.

[14] Hu Minqing, Liu Bing. Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004[C]. New York, NY, USA 2004: 168-177.

[15] Lee, D., Jeong, O., Lee, S. Opinion Mining of Customer Feedback Data on the Web. Proceedings of the 2nd international conference on Ubiquitous information management and communication, 2008[C]. ACM Press New York, 2008: 230-235.

[16] Dave, K., Lawrence, S., Pennock, D. M. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web, 2003[C]. ACM Press New York, 2003: 519-528.

[17] Zhang Meng, Peng Yifan, Fan Yang, Li Dan, Lin Xiaojun, Wu Xihong. Research on Chinese orientation analysis. Proceedings of the First Chinese Orientation Analysis Evaluation Workshop, [C], 2008, 38—45.

[18] Yue Lu, Malu Castellanos, Umeshwar Dayal, ChengXiang Zhai. Automatic Construction of a Context-Aware Sentiment Lexicon: An Optimization Approach. Proceeding of the 20th international conference on World Wide Web, 2011[C]. 2011: 347-356.

Author Contact Information: Li Hanyu, Information Systems Department, National Science Library, Chinese Academy of Sciences, 15201227359, lihy@mail.las.ac.cn

Note: Figure translations are in progress. See original paper for figures.