A Context-Dependent Sentiment Analysis of Online Product Reviews based on Dependency Relationships

Submission Type: Completed Research Paper

Zhijun Yan
Beijing Institute of Technology
5 South Zhongguancun Street, Haidian District, Beijing, P.R.China
yanzhijun@bit.edu.cn

Meiming Xing
Beijing Institute of Technology
5 South Zhongguancun Street, Haidian District, Beijing, P.R.China
minghaizai@163.com

Dongsong Zhang
University of Maryland, Baltimore
1000 Hilltop Circle, Baltimore, MD21250, USA
zhangd@umbc.edu

Baizhang Ma
Beijing Institute of Technology
5 South Zhongguancun Street, Haidian District, Beijing, P.R.China
gufeiyan_295259083@qq.com

Tianmei Wang
Central University of Finance and Economics
39 South College Road, Haidian District, Beijing, P.R.China
wangtianmei@cufe.edu.cn

Abstract

Consumers often view online consumer product review as a main channel for obtaining product quality information. Existing studies on product review sentiment analysis usually focus on identifying sentiments of individual reviews as a whole, which may not be effective and helpful for consumers when purchase decisions depend on specific features of products. This study proposes a new feature-level sentiment analysis approach for online product reviews. The proposed method uses an extended PageRank algorithm to extract product features and construct expandable context-dependent sentiment lexicons. Moreover, consumers' sentiment inclinations toward product features expressed in each review can be derived based on term dependency relationships. The empirical evaluation using consumer reviews of two different products shows a higher level of effectiveness of the proposed method for sentiment

analysis in comparison to two existing methods. This study provides new research and practical insights on the analysis of online consumer product reviews.

Key words: *sentiment analysis, online product reviews, feature extraction, text mining, PageRank*

Introduction

In the past decade, as online shopping becomes increasingly popular, consumers tend to obtain product quality information through navigating online consumer product reviews before making their purchase decisions (Chevalier and Mayzlin 2006). However, different consumers may have different preferences in product features. For example, one consumer may consider appearance and weight as the most important features while purchasing a mobile phone, while another may be mainly concerned about the battery or functions of a phone. As the volume of online reviews continues soaring, it is time-consuming and practically impossible to navigate and filter reviews one by one in order to obtain useful information on product quality and product fitness at a product feature level.

There have been increasing studies on analysis of online product reviews, including product feature extraction (Chen et al. 2012; Huang et al. 2012), review sentiment analysis (Eirinaki et al. 2012; Taboada et al. 2011), and review helpfulness prediction (Pan and Zhang 2011; Zhang et al. 2012), etc. However, existing work has several limitations. First, most existing literature focuses on analyzing individual product reviews as a whole, instead of at a product-feature level. Thus consumers can hardly know the opinions of others on specific features of a product, which is often important to reduce product fit and quality uncertainty needed for making a purchase decision, let alone the formidable task of manually wading through all reviews (Hu et al. 2009). Different consumers are often interested in different product features. It will be difficult to improve the buy conversion rate if a consumer can't quickly identify individual reviews that have commented on certain product features of his/her interest (Somprasertsri and Lalitrojwong 2010). Second, due to linguistic and cultural differences, the approaches proposed for analyzing English reviews cannot be directly applied to Chinese reviews. For example, a sentence in Chinese should be parsed and segmented into different words before any further analysis, which is not necessary for English sentences (Fu et al. 2013; Zhang et al. 2009). Third, most previous studies analyze review sentiments based on a fixed sentiment term lexicon. Sentiment analyses of online product reviews can help consumers gain a comprehensive understanding of others' positive and negative opinions on a certain product (Ding et al. 2008). But lots of sentiment terms in consumer reviews are context-dependent and may have positive or negative polarities under different contexts. There lacks an effective way to incorporate context-dependent terms into a sentiment term lexicon that can improve the performance of sentiment analysis considerably.

A feature-level sentiment analysis aims to determine a consumer's sentiment toward a product feature (Quan and Ren 2014). Different from a review-level sentiment analysis, a feature-level sentiment analysis can provide a finer grained analysis of opinions on certain product features. On the one hand, it is able to generate a sentiment rating for a product feature, which will be helpful for making purchase decisions. On

the other hand, it enables e-commerce websites and product manufacturers to categorize consumer reviews into different groups based on features commented in reviews, which make the identification of product defects much easier.

This study contributes to the literature in several aspects. First, this paper proposes a new feature-level approach to sentiment analysis of online product reviews centered on product features and dependency relationships between feature and sentiment terms. The proposed method first extracts feature-sentiment dependency relationships with the Stanford parser. Then, an extended PageRank algorithm is employed to derive initial candidate product features from those dependency relationships. Based on the result, the sentiment rating of each commented product feature in a review will be generated. Second, consumers may use some ambiguous or context-dependent sentiment terms to express their feelings about product features, which results in the low precision of automatic review sentiment analysis. The proposed research identifies the polarity of context-dependent sentiment terms based on linguistic rules. Third, in our approach, a sentiment term lexicon can be expanded automatically in the review analysis process, which makes the review analysis not subject to a pre-defined, fixed sentiment lexicon. Furthermore, this research provides a technical foundation for supporting personalized summarization and presentation of online product reviews according to certain product features of a consumer's interest to help them make better purchase decisions.

The rest of the paper will be organized as follows. At first, the related research on sentiment analysis of online product reviews will be introduced. The next section presents the proposed sentiment analysis method. Then, the experiment result will be presented and discussed, followed by the conclusion in the next section.

Related Research

Currently, there are two types of methods for sentiment analysis of online consumer reviews. The first type of methods relies on existing sentiment and polarity lexicons (Kim and Hovy 2004; Williams and Anand 2009). They search for sentiment words in a review, and then decide their polarities according to pre-defined lexicons. Another type of methods is based on supervised machine learning (Boiy and Moens 2009; Jiang et al. 2011), in which a model is learned with a training dataset for predicting sentiments (i.e., positive/neutral/negative).

A lexicon-based sentiment analysis method is usually carried out with word segmentation and PoS (part-of-speech) tagging first, then candidate product features and associated sentiment terms will be extracted with some rules, followed by the identification of the polarity of sentiment terms associated with specific features based on a sentiment lexicon. However, a traditional sentiment lexicon doesn't include all sentiment terms, especially context-dependent sentiment terms. Thus some extra steps are necessary to identify the polarity of those sentiment words.

Lexicon-based sentiment analysis methods can be classified into two types, including methods with iterations of a seed lexicon and methods with contextual analysis. Methods of the first type start with a standard lexicon as a seed lexicon. They expand the seed lexicon with some rules to obtain a complete sentiment word lexicon. For example, Kamps at al. (2004) construct a graph of sentiment terms, then

calculate the distance between an unlabeled term and 'good' as well as the distance between an unlabeled term and 'bad' to determine the term's polarity. Then the polarities of unlabeled sentiment terms can be derived based on a sentiment term graph. Godbole et al. (2007) shed light on sentiment analysis of news and blogs. They extract entities (i.e., people, places, things) and sentiment words from news and blogs, and a sentiment word path is constructed according to sentiment words' location in a corpus. Then a complete sentiment word lexicon will be built. Qiu et al. (2011) extract product features and sentiment words based on dependency grammar analysis and a DP (Double Propagation) algorithm, and judge polarities of sentiment words with a seed lexicon directly.

For a lexicon-based method with contextual analysis, polarities of sentiment terms are obtained based on the analysis of relationships among different words in the same sentences and different sentences in a document set. Wilson et al. (2009) propose a method to automatically distinguish between prior and contextual polarities of the words, with a focus on understanding which features are important for polarity identification. Liu (2010) determines the polarity of sentiment terms by analyzing the sentiment consistency within sentences and between sentences. The intra-sentence and inter-sentence consistencies are judged mainly through conjunction words, such as 'but', 'however', and 'and'. Moreover, rules are also created to determine the polarity of sentiment terms.

Lexicon-based sentiment analysis methods are product-feature level methods and have high precision. However, these methods highly depend on pre-defined lexicons and are context dependent. The quality of a pre-defined lexicon has great impact on the precision of sentiment analysis. With the rapid increase of products available on an e-commerce website, it is difficult to maintain a self-sustainable lexicon.

Many studies on sentiment analysis have adopted supervised machine learning methods. Go et al. (2009) introduce a novel approach for automatically classifying the sentiments of Twitter messages. They employ several machine learning algorithms, including Naïve Bayes, Maximum Entropy, and Support Vector Machine (SVM), to classify a Twitter message as either a positive or a negative post. Tan et al. (2009) propose an effective measure called Frequently Co-occurring Entropy (FCE) to handle the domain-transfer problem in sentiment analysis. They propose Adapted Naïve Bayes (ANB), a weighted transfer version of Naïve Bayes Classifier, to gain knowledge from data of a new domain. Boiy and Moens (2009) identify sentiment inclinations of blogs, online reviews, and online forum texts written in English, Dutch and French based on machine learning techniques. Support Vector Machine, Multinomial Naïve Bayes and Maximum Entropy are applied in the sentiment analysis process. Compared with a lexicon-based method, a method with supervised learning models has higher efficiency and lower cost. However, a method with supervised learning models is more appropriate for document-level sentiment analysis, with the domain-dependent problem still unsolved.

In sum, existing methods for sentiment analysis of online consumer product reviews suffer from several major limitations. Lexicon-based methods are highly dependent on pre-defined, fixed lexicons and it is time-consuming and ineffective to maintain a lexicon for different contexts. Supervised machine learning based methods are more appropriate for document-level sentiment analysis and their performance is domain-dependent. In this research, we apply a lexicon-based sentiment analysis method, with a sentiment lexicon that can be expanded by context-dependent sentiment terms dynamically and automatically.

Sentiment Analysis based on Dependency Relationships

The proposed method for feature-level sentiment analyses consists of several steps. The first step is to build an expandable sentiment lexicon based on a standard lexicon. The polarity of sentiment words in online product reviews will be identified. Second, syntactic analysis of review text, including word segmentation, PoS tagging, and noun or noun phrase extraction, will be performed. Third, dependency pairs of product features and sentiment words will be extracted, which will then be ranked by an extended PageRank algorithm. Finally, consumers' sentiment on different product features and reviews will be derived.

Sentiment Lexicon Building

A lexicon is the core of lexicon-based sentiment analyses. In the proposed approach, lexicons consist of a basic polarity lexicon, a polarity modifying lexicon, and a context-dependent lexicon. The basic polarity lexicon and polarity modifying lexicon are fixed, while the context-dependent lexicon can be expanded in the review analysis process.

A basic polarity lexicon is composed of sentiment terms that have relatively fixed sentiment polarities. We use HowNet (http://www.keenage.com/) as the basic polarity lexicon in this study, which consists of 219 degree-level terms, 3,116 negative evaluation terms, 1,254 negative sentiment terms, 3,730 positive evaluation terms, 836 positive sentiment terms, and 38 proposition terms.

Although the basic polarity lexicon provides the polarity of a large number of sentiment terms, it is not sufficient to deal with all possible sentiment terms in online consumer product reviews because a term's context (e.g. product type, sentence structure) also influences its polarity and needs to be considered. Therefore, we take negation, strengthening, and weakening modifiers of basic polarity terms into consideration by constructing a polarity modifying lexicon, which includes a negation lexicon and an emphasis lexicon. The emphasis lexicon is mainly originated from terms representing different degrees of sentiments in HowNet, and the negation lexicon is built by analyzing online product reviews. The emphasis terms will be classified into four levels, including extreme degree, high degree, medium degree, and low degree. Each degree will be assigned with a value. For example, the value of extreme degree emphasis terms is 2. Negation terms represent the inverted polarity of the modified sentiment terms.

In addition to the basic polarity lexicon and the polarity modifying lexicon, a context-dependent lexicon is constructed in the proposed method. Generally, the polarity of context-dependent terms is determined based on linguistic rules (Yan 2010) that include rules on conjunctions within sentences, conjunctions between sentences, and non-adversative conjunctions. The first two types of rules determine sentiment polarities through adversative and parity conjunctions, such as 'and' (和, 而且) and 'but' (但是). Rules for non-adversative conjunctions indicate that if there are no adversative conjunctions within a sentence or between sentences, the polarity of a context-dependent term stays the same as that of the adjacent sentiment term.

Review Preprocessing

Before any sentiment analysis, some review preprocessing needs to be done, including word segmentation,

PoS tagging, and noun or noun phrase extraction. First of all, the proposed approach performs word segmentation and PoS tagging on individual Chinese consumer review sentences using ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System, http://www.ict.ac.cn/). After that, each word and its syntactic tag will be identified. The following example shows a parsed Chinese review sentence with PoS tags produced by ICTCLAS: "价格(price)/n 实惠(affordable)/an ，/w 外观(appearance)/n 也(also)/d 很(very)/d 不错(nice)/a", where 'n' represents a noun; 'an' represents an adnoun; 'w' represents a punctuation; 'd' represents an adverb; and 'a' represents an adjective.

After word segmentation and PoS tagging, an approach to noun phrase extraction (Ma et al. 2013) will be applied. In this stage, rules for noun and noun phrase extraction and filtering will be used to get a list of nouns and noun phrases. After each review in a review corpus is parsed, those extracted nouns and noun phrases, excluding common non-feature terms, will form an initial feature set.

Extraction of Dependency Pairs

The polarities of consumers' sentiments on product features are determined by analyzing dependency relationships between feature terms and sentiment terms. Based on the result of ICTCLAS, the Stanford parser is used to perform feature-sentiment term dependency parsing. The result of dependency parsing is dependency trees and a set of feature-sentiment term dependency pairs. Each review will generate a dependency tree. Then dependencies that possibly contain both a product feature and a sentiment term will be identified.

In this study, we focus on term dependencies existing in subject-predicate relationships (nsubj), verb-object relationships (dobj), adjectival modifying relations (amod), and relative clause modifying relations (rcmod). The system will then search for feature terms in those identified relationships.

Feature Extraction

PageRank is a webpage ranking algorithm proposed by Larry Page and Sergey Brin in the late 1990s (Brin and Page 1998), which is mainly used to assess a webpage's importance. By considering the value of hyperlinks within a webpage as a ranking factor, PageRank is one of the most popular algorithms and the basis of the Google search engine.

The intuition behind the PageRank algorithm is that the importance of a Web page depends on the pages that link to it. Similarly, we predict that the possibility of a candidate feature term to be a real feature term depends on the sentiment terms that modify it. Such feature-sentiment term relations can form a graph, which may be helpful to improve product feature extraction. Considering product features are normally nouns or noun phrases, a noun or noun phrase extracted from an online consumer product review is more likely to be a feature term if it is modified by more adjectives in a review (Eirinaki et al. 2012). A term pair of a product feature and an associated sentiment term can be treated as a network node, and modifying relations between different candidate feature terms and associated sentiment terms can be treated as network edges. Thus, after extracting dependency pairs, we apply an extended PageRank algorithm to assess the importance of network nodes and derive correct product features (Ma 2014). Assuming the set of all term pairs is W; the set of all feature terms in W is F; and the set of all sentiment terms in W is S.

The Cartesian product of F and S is V. Then a node in the graph is a word pair and an element Of V. A sentiment term s in node v_1 and a feature term f in node v_2 compose a pair (f,s). If (f,s) is an element of W, we link v_1 and v_2 with a direction from v_1 to v_2. For example, if W has two elements (price, cheap) and (screen, large), F will be {price, screen} and S will be {cheap, large}. V will have four elements: (screen, cheap), (price, cheap), (screen, large), (price, large). Thus the edge set E in the graph has two elements, namely (v_1,v_2) and (v_4,v_3). A node-link graph is given in Figure 1.

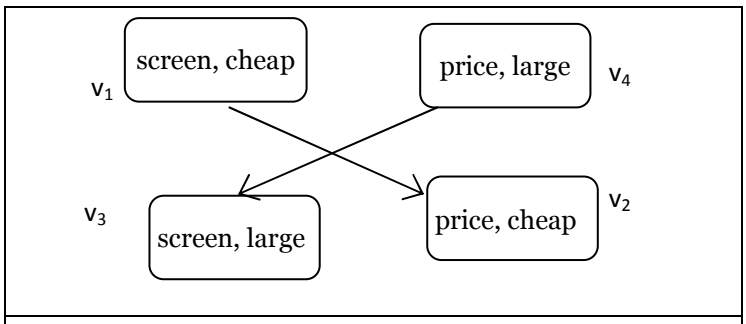

Figure 1. A Node-Link Graph

After the above graph construction step, we apply a node ranking algorithm that extends the PageRank algorithm to rank the feature-sentiment term pairs. Considering the more often a noun occurs in a review, the more likely it is a product feature, we define the rank of node i (P(i)) as follows:

$$P(i) = (1-\alpha)H(i) + \alpha \sum_{(j,i)\in E} \frac{P(j)}{O_j} \qquad (1)$$

where i is the current node; H(i) is a function of occurrence frequency of node i; α is a damping coefficient; O_j is the number of nodes to which the node j links; E is the edge set of the graph. It is easy to prove that the modified formula (1) satisfies the three requirements of Markov chain model and can reach the convergence. The vector $P=(p(1), p(2), p(3), \ldots, p(n))^T$ can be calculated by using the power iterative algorithm (Golub and Van Loan 2012). Specifically, H(i) in the formula (1) is defined as:

$$H(i) = \frac{n * \ln f(i)}{\ln(\prod_{j=1}^{n} f(j))} \qquad (2)$$

where n is the number of nodes in the graph; f(i) is the occurrence frequency of node i; and the log function can reduce the adverse impact of high occurrence frequencies of some terms on the ranking algorithm. At the end, the term pairs with ranking scores higher than a threshold will be identified, in which feature terms within those pairs will be selected as product features. The experiment results revealed that the extended PageRank algorithm outperformed the association rule mining algorithm in precision, recall, and F-measures of product feature extraction.

Sentiment Analysis

After product feature extraction, a sentiment analysis based on dependency relationships will be conducted. During this process, a sentiment score will be derived and assigned to each product feature term in a review, as well as each review as a whole. Additionally, the context-dependent sentiment lexicon

is expanded automatically in this step. The whole process of sentiment analysis includes clause segmentation, basic scoring, context-dependent scoring, lexicon expansion, sentence and review scoring, as shown in Figure 2.

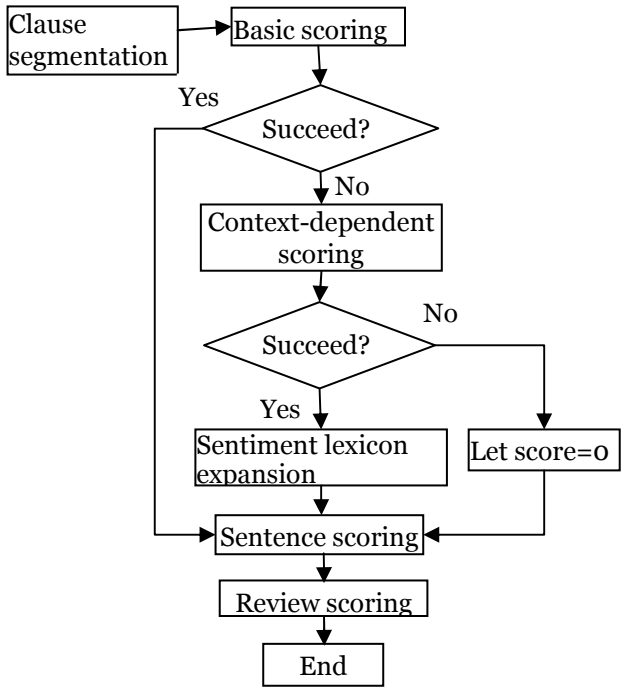

Figure 2. The Sentiment Analysis Process

Clause Segmentation

This step mainly deals with segmenting review sentences that have already gone through word segmentation and PoS tagging. Cues to segmentation include punctuation marks, such as period, comma, and ellipsis. For example, a review "屏幕(screen)/n 大(big)/a ，/w 看(watch)/v 电影(movie)/n 玩(play)/v 游戏(game)/n 很(very)/d 好(good)/a ，/w 电池(battery)/n 比较(pretty)/d 耐用(durable)/a 。/w" ("The screen is big, which is good for watching movies and playing games, and the battery is pretty durable") could be segmented into three clauses: "屏幕/n 大/a" ("the screen is big"), "看(watch)/v 电影(movie)/n 玩(play)/v 游戏(game)/n 很(pretty)/d 好(good)/a" ("is good for watching movies and playing games") and "电池(battery)/n 比较(pretty)/d 耐用(durable)/a"("the battery is pretty durable"). Based on the result of clause segmentation, the sentiment inclination of each clause can be determined firstly, and then the sentiment inclination of a whole sentence will be derived.

Basic Scoring

In this step, a basic scorer is used to detect the polarity of a sentiment term associated with a specific product feature. It will search the three sentiment lexicons for the terms and their modifiers and obtain a basic score after finding a product feature and its associated sentiment term. There may be negations and

degree modifiers that change the sentiment intensity or even the polarity of sentiment terms. In some special occasions, a sentiment term may potentially have more than one polarity. Therefore, the basic scorer needs to deal with these special circumstances.

(1) If there is a negation before a sentiment term, the basic scorer sets the polarity of the sentiment as the opposite of the original polarity of the sentiment term.

(2) If there is a modifier before a sentiment term, the basic scorer sets the polarity as the polarity of the sentiment term multiplied by the intensity degree of the modifier. For example, the degree of a modifier 'very'(非常) is 2, and the polarity of the term 'good'(好) is 1, so the polarity of "very good"(非常好) is 2*1=2.

(3) If there are both a negation and a modifier before a sentiment term, the basic scorer sets the polarity as the combined result of (1) and (2).

(4) If a feature is modified by several sentiment terms, the feature's sentiment polarity is the sum of all polarities of the sentiment terms modifying the feature.

If the basic scorer performs scoring successfully, the proposed approach continues to determine a sentence polarity. If a clause contains context-dependent sentiment terms, it indicates that context-dependent scoring is required. For example, the clause "屏幕(screen)/n 大(big)/a" has a nsubj dependency type. '屏幕(screen)' is a product feature. '大(big)' is a context-dependent sentiment. Therefore, the polarity of "大(big)" couldn't be determined by the basic scorer, and will be forwarded to the context-dependent scorer.

Context-Dependent Scoring and Lexicon Expansion

The context-dependent scorer judges the polarity of a sentiment term based on review context, in which the term resides, after the basic scorer fails to score. Context mainly refers to relationships implied by conjunction terms, including adversatives and juxtapositions. Context-dependent sentiment terms may have different sentiment polarities in different context. The context-dependent scorer identifies the polarity of context-dependent sentiment terms according to linguistic rules (Yan 2010). It adds a term and its polarity to a self-expanding context-dependent lexicon if it successfully assigns a score to the term, and the lexicon in turn helps the basic scorer. If the polarity of a context-dependent term cannot be identified based on the linguistic rules, the proposed approach will set the terms' polarities to neutral as default.

Sentence and Review Scoring

Based on the result of basic and context-dependent scorers, the sentiment polarity of each product feature commented in each review will be derived, and the polarity of each clause can be calculated by adding the sentiment polarities of each product feature commented in a clause. Then, the polarity of a review as a whole will be determined by aggregating the polarities of all its clauses. The final polarity score of a review will be divided by the number of product features mentioned in the review for normalization. A review polarity score ranges from -2 to 2, while a positive score means a positive polarity, a negative score means a negative polarity, and a 0 represents a neutral polarity. For a whole review corpus, the sentiment of one product feature can be calculated by aggregating the polarities of sentiments on that feature in all reviews.

Thus the sentiment inclinations of each product feature and each review can be obtained and categorized. Consumers can search or browse reviews based on the polarity scores of product features of their interest.

Evaluation and Analysis of Results

Evaluation Indices and Data

Precision has been the most frequently used measure for evaluating approaches to sentiment analysis (Liu et al. 2005). The definition of precision is shown in formula (3).

$$Precision = \frac{number\ of\ reviews\ whose\ polarities\ are\ correctly\ labeled}{total\ number\ of\ reviews} \qquad (3)$$

A web crawler written in Java was utilized to collect online consumer reviews on a digital camera and a mobile phone generated at jd.com between December 2012 and August 2013. Digital cameras and mobile phones are the most commonly studied products in previous studies on online consumer reviews (Aciar et al. 2007).

Once obtaining the dataset, some pre-processing and filtering were done based on the following criteria:

(1) Review length: a review must contain more than 10 but fewer than 500 words. Reviews that were too short or too long were removed from the dataset.

(2) Duplicate characters: Some reviews consisted of duplicate characters or phrases. We restrict the number of non-duplicate characters in a review to be at least 7 (including punctuation marks).

(3) Useless reviews: Some consumers adopted the default review provided by the website. Such reviews were excluded from the analysis.

Finally, we randomly selected 3,000 reviews of each product from the qualified review collection. The average lengths of reviews of Canon EOS 600D and Samsung Note II phone were 72 and 74 words, respectively.

Three graduate students were asked to identify the sentiment inclination of each online review individually first. Then, they discussed disagreements to reach a consensus. Eventually, the sentiment inclination of each review was marked as positive, neutral or negative. The result forms the benchmark, which was then compared against the sentiment inclination result produced by the proposed approach.

Results

In this evaluation, Opinion Observer (Ding et al. 2008) and Naïve Bayes (Paltoglou et al. 2010) methods were selected as benchmarks. Opinion Observer is a lexicon-based sentiment analysis method and Naïve Bayes is a supervised machine learning method. Both have been used in many prior studies (Kang et al. 2012; Pang and Lee 2008).

In order to test the statistical significance of difference between the proposed method and benchmarks, we randomly divided reviews into 30 groups with equal size. Then, precision of each group was calculated for every approach (i.e., the proposed method, Opinion Observer, and Naïve Bayes). We further conducted paired T-tests to examine whether the improvements of the proposed approach over the

Opinion Observer and Naïve Bayes methods were statistically significant. Means and Standard Deviations (SD) of precision and T-test (df = 29) results are shown in Table 1 and Table 2, respectively.

	P1(Opinion Observer)		P2(The Proposed Method)		P1-P2 (MD)	t	Sig.
	Mean	SD	Mean	SD			
Samsung Note II	80.5	4.34	85.2	5.66	-4.7	-9.68	.003
Canon EOS 600D	79.8	6.91	88.5	5.74	-8.7	-14.06	.000

Table 1. Precision of the Proposed Method Versus Precision of Opinion Observer (%)

The T-test result illustrates that the proposed method achieves higher precision than the two existing methods across two products. The results also reveal that the performance of Naïve Bayes is better than Opinion Observer significantly for the reviews on the mobile phone ($P<0.05$).

	P1(Naïve Bayes)		P2(The Proposed Method)		P1-P2 (MD)	t	Sig.
	Mean	SD	Mean	SD			
Samsung Note II	83.0	6.21	85.2	5.66	-2.2	-5.52	.034
Canon EOS 600D	81.4	5.46	88.5	5.74	-7.1	-11.35	.001

Table 2. Precision of the Proposed Method Versus Precision of Naïve Bayes (%)

Conclusion

Online consumer product reviews are valuable and helpful for consumers to make better online purchase decisions. However, it is impossible for consumers to browse a large number of reviews one by one to identify those that have commented on certain product features of their interest. Because sentiment terms may have different polarities in different contexts and different consumers may have different product feature preferences, feature-oriented and context-dependent review sentiment analysis would be very useful to help consumers understand others' sentiments on specific features of a target product.

This paper proposes a new sentiment analysis method based on dependency relationships and product features. The proposed method uses a set of sentiment lexicons, including a basic polarity lexicon, a polarity modifying lexicon, and a dynamic context-dependent lexicon. We extend the PageRank algorithm to extract product features. The polarity of individual reviews is derived through the basic and context-dependent scoring processes. The evaluation results involving two products show that the proposed method significantly outperforms the Naïve Bayes and Opinion Observer methods.

In addition to research contributions discussed above, our study also provides some practical implications. For consumers, it would be beneficial for them to mainly browse reviews that include comments on specific product features of their interest. An online product review system that supports the analysis, filtering, and presentation of reviews based on commented product features can also help consumers overcome the information overload problem caused by the ever-increasing number of online reviews and help them make better and quicker purchase decisions. For e-commerce websites, the proposed method can enable them to build a feature-oriented review search and analysis system. Such a system can provide personalized review summarization for consumers, which can increase the buy conversion rate and

customer satisfaction greatly. For product manufacturers, the proposed method enables them to identify consumers' positive and negative comments on specific product features easily. It helps product manufacturers better understand what their customers truly think of their products from a product feature point of view, as well as help them identify the major flaws and potential improvement of products.

There are several limitations of this study. First, only reviews of two products collected from jd.com were analyzed in the evaluation. Considering a large variety of products on e-commerce websites, our dataset was limited in terms of product diversity. Although the proposed method is generic and should be applicable to different products, future work should be conducted to test the generalizability of the proposed method with reviews of different types of products. Second, this study focuses on sentiment analysis of Chinese product reviews. As discussed earlier, Chinese product reviews have different characteristics in comparison with product reviews in English. We used Chinese lexicons in this study to process Chinese online reviews. For a sentiment analysis of online reviews written in English, other sentiment lexicons will be necessary and the effectiveness of the proposed method should be assessed again. Third, we adopt two existing methods used in previous studies for evaluation of the proposed approach. There are many other methods developed for sentiment analysis of online product reviews. Finally, we don't take the computational complexity into consideration. The evaluation only focuses on the precision of sentiment analysis of three methods. Although computational complexity is less of a concern in this context because the entire review analysis is done offline, optimizing computation and reducing complexity of this whole process should be very beneficial for practical applications. Future research should address this issue.

In summary, a feature-level sentiment analysis of online product reviews not only helps consumers better handle the information overload problem, but also enables personalized review summarization and navigation for better purchase decisions. This study makes an initial effort toward achieving that goal. The evaluation results show that the proposed method is promising.

Acknowledge

This research is supported by the National Natural Science Foundation of China (Award #: 71128003, 70972006, 71272057) and Program for New Century Excellent Talents in University (Award #: NCET-11-0792).

References

Aciar, S., Zhang, D., Simoff, S., and Debenham, J. 2007. "Informed Recommender: Basing Recommendations on Consumer Product Reviews," *IEEE Intelligent Systems* (22:3), pp. 39-47.

Boiy, E., and Moens, M.-F. 2009. "A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts," *Information Retrieval* (12:5), pp. 526-558.

Brin, S., and Page, L. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems* (30:1), pp. 107-117.

Chen, L., Qi, L., and Wang, F. 2012. "Comparison of Feature-Level Learning Methods for Mining Online Consumer Reviews," *Expert Systems with Applications* (39:10), pp. 9588-9601.

Chevalier, J.A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.

Ding, X., Liu, B., and Yu, P.S. 2008. "A Holistic Lexicon-Based Approach to Opinion Mining," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp. 231-240.

Eirinaki, M., Pisal, S., and Singh, J. 2012. "Feature-Based Opinion Mining and Ranking," *Journal of Computer and System Sciences* (78:4), pp. 1175-1184.

Fu, X., Liu, G., Guo, Y., and Wang, Z. 2013. "Multi-Aspect Sentiment Analysis for Chinese Online Social Reviews Based on Topic Modeling and Hownet Lexicon," *Knowledge-Based Systems* (37), pp. 186-195.

Go, A., Bhayani, R., and Huang, L. 2009. "Twitter Sentiment Classification Using Distant Supervision," Stanford University, pp. 1-12.

Godbole, N., Srinivasaiah, M., and Skiena, S. 2007. "Large-Scale Sentiment Analysis for News and Blogs," in *Proceedings of International Conference on Weblogs and Social Media*, Boulder, Colorado, USA.

Golub, G.H., and Van Loan, C.F. 2012. *Matrix Computations*. JHU Press.

Hu, N., Zhang, J., and Pavlou, P.A. 2009. "Overcoming the J-Shaped Distribution of Product Reviews," *Communications of the ACM* (52:10), pp. 144-147.

Huang, S., Liu, X., Peng, X., and Niu, Z. 2012. "Fine-Grained Product Features Extraction and Categorization in Reviews Opinion Mining," in *Proceedings of IEEE 12th International Conference on Data Mining Workshops* Brussels, Belgium, pp. 680-686.

Jiang, L., Yu, M., Zhou, M., Liu, X., and Zhao, T. 2011. "Target-Dependent Twitter Sentiment Classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oregon, pp. 151–160.

Kamps, J., Marx, M., Mokken, R., and de Rijke, M. 2004. "Using Wordnet to Measure Semantic Orientations of Adjectives," in *Proceedings of Fourth International Conference on Language Resources and Evaluation*, pp. 1115-1118.

Kang, H., Yoo, S.J., and Han, D. 2012. "Senti-Lexicon and Improved Naïve Bayes Algorithms for Sentiment Analysis of Restaurant Reviews," *Expert Systems with Applications* (39:5), pp. 6000-6010.

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

Liu, B. 2010. "Sentiment Analysis: A Multi-Faceted Problem," *IEEE Intelligent Systems* (25:3), pp. 76-80.

Liu, B., Hu, M., and Cheng, J. 2005. "Opinion Observer: Analyzing and Comparing Opinions on the Web," in: *Proceedings of the 14th International Conference on World Wide Web*. ACM, pp. 342-351.

Ma, B. 2014. "E-Commerce Consumer Online Reviews Mining Research." Beijing: Beijing Institute of Technology.

Ma, B., Zhang, D., Yan, Z., and Kim, T. 2013. "An Lda and Synonym Lexicon Based Approach to Product Feature Extraction from Online Consumer Product Reviews," *Journal of Electronic Commerce Research* (14:4), pp. 304-314.

Paltoglou, G., Gobron, S., Skowron, M., Thelwall, M., and Thalmann, D. 2010. "Sentiment Analysis of

Informal Textual Communication in Cyberspace," in *Proceedings of ENGAGE 2010*, pp. 13-25.

Pan, Y., and Zhang, J.Q. 2011. "Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews," *Journal of Retailing* (87:4), pp. 598-612.

Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval* (2:1-2), pp. 1-135.

Qiu, G., Liu, B., Bu, J., and Chen, C. 2011. "Opinion Word Expansion and Target Extraction through Double Propagation," *Computational linguistics* (37:1), pp. 9-27.

Quan, C., and Ren, F. 2014. "Unsupervised Product Feature Extraction for Feature-Oriented Opinion Determination," *Information Sciences* (272), pp. 16-28.

Somprasertsri, G., and Lalitrojwong, P. 2010. "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization," *Journal of Universal Computer Science* (16:6), pp. 938-955.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. 2011. "Lexicon-Based Methods for Sentiment Analysis," *Computational linguistics* (37:2), pp. 267-307.

Tan, S., Cheng, X., Wang, Y., and Xu, H. 2009. "Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis," in *Advances in Information Retrieval,* M. Boughanem, C. Berrut, J. Mothe and C. SouleDupuy (eds.). Berlin: Springer-Verlag Berlin, pp. 337-349.

Williams, G.K., and Anand, S.S. 2009. "Predicting the Polarity Strength of Adjectives Using Wordnet," in *Proceedings of the Third International Conference on Weblogs and Social Media*, pp. 346-349.

Wilson, T., Wiebe, J., and Hoffmann, P. 2009. "Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis," *Computational Linguistics* (35:3), pp. 399-433.

Yan, S. 2010. "Research on Opinion Mining of Product Reviews in Chinese." Beijing: Beijing Jiaotong University.

Zhang, C., Zeng, D., Li, J., Wang, F.Y., and Zuo, W. 2009. "Sentiment Analysis of Chinese Documents: From Sentence to Document Level," *Journal of the American Society for Information Science and Technology* (60:12), pp. 2474-2487.

Zhang, R., Tran, T., and Mao, Y. 2012. "Opinion Helpfulness Prediction in the Presence Of "Words of Few Mouths"," *World Wide Web-Internet and Web Information Systems* (15:2), pp. 117-138.