

EXPRS: An extended pagerank method for product feature extraction from online consumer reviews



Zhijun Yan^a, Meiming Xing^a, Dongsong Zhang^{a,b,*}, Baizhang Ma^a

^a School of Management & Economics, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China

^b Department of Information Systems, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA

ARTICLE INFO

Article history:

Received 27 August 2014

Received in revised form 23 January 2015

Accepted 14 February 2015

Available online 7 March 2015

Keywords:

Online product reviews

Feature extraction

Extended PageRank algorithm

Synonym expansion

Social media analytics

ABSTRACT

Online consumer product reviews are a main source for consumers to obtain product information and reduce product uncertainty before making a purchase decision. However, the great volume of product reviews makes it tedious and ineffective for consumers to peruse individual reviews one by one and search for comments on specific product features of their interest. This study proposes a novel method called EXPRS that integrates an extended PageRank algorithm, synonym expansion, and implicit feature inference to extract product features automatically. The empirical evaluation using consumer reviews on three different products shows that EXPRS is more effective than two baseline methods.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The rapid advance of Internet technology has been greatly propelling e-commerce. iResearch reports that the total online sales in China reached 0.3 trillion dollars in 2013 [21]. Many e-commerce websites and social media platforms, such as Amazon.com, Taobao.com, Eopinion.com, Yelp.com, and Dianping.com, provide an online information sharing channel that allows customers to post their comments on the products that they have bought or consumed and to browse such product reviews of other consumers in order to acquire more knowledge about product quality [7,45]. Research has shown that online consumer product reviews not only have significant impact on consumers' online purchase decisions [9,33], but also are helpful for manufacturers to improve product design and quality and for online retailers to improve their services [2,18,30].

There are two key problems that hinder consumers from fully taking advantage of online consumer product reviews. First, the explosive growth of online product reviews makes it

time-consuming and ineffective for a consumer to navigate individual reviews one by one in order to obtain useful information about a product, causing an information overload problem [25]. Previous studies have shown that information overload negatively influences purchase decisions considerably and may lead to fewer consumer purchases [5,8,15]. Therefore, how to deal with such an information overload problem becomes increasingly critical. Although text mining techniques have been applied to online review analysis, they are mainly used for identifying general trends or certain patterns of online consumer reviews instead of coping with information overload.

Second, in reality, individual consumers may have preferences for different product features. For example, some consumers may consider appearance and weight as the most important factors while purchasing a mobile phone, while others may be mainly concerned about its battery life or functions. As a result, reviews that are rated as 'helpful' by some consumers may not really be helpful to others if those reviews do not comment on specific product features of their interest. Most of current platforms of online consumer product reviews do not take needs and preferences of individual consumers into consideration. Few online review platforms organize and present reviews in a personalized, product-feature oriented manner. As a result, it is practically impossible for a consumer to wade through hundreds or even thousands of reviews in order to identify reviews that have commented on specific product features.

* Corresponding author at: Department of Information Systems, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. Tel.: +1 4104552851.

E-mail addresses: yanzhijun@bit.edu.cn (Z. Yan), minghaizai@163.com (M. Xing), zhangd@umbc.edu (D. Zhang), 295259083@qq.com (B. Ma).

Product feature extraction from online consumer reviews is a key to addressing the above two problems. As a fundamental and important step of consumer review analysis, feature extraction is aimed to identify product features commented in reviews automatically [11,36]. Existing approaches to feature extraction have several limitations. First, product features commented in reviews can be classified into two types: explicit features and implicit features [19]. Explicit features refer to the features mentioned directly in a review. Implicit features are those that are implicitly commented but do not appear in a review. Extraction of implicit features has been largely ignored in previous studies. Second, a product feature may be modified by several different sentiment terms, and one sentiment term can be used to modify different product features. Based on this phenomenon, we treat a term pair of a product feature and its sentiment term as a network node, and relations between different candidate features and sentiment terms can be treated as network edges. Such a network may be helpful for identifying product features. Although a few approaches to product feature extraction from online consumer reviews have explored the incorporation of such relations [20,35], none of them is based on a product feature-sentiment term relation network. Third, prior studies did not take synonyms of product features (e.g., the internal storage of a mobile phone, memory, and RAM) into consideration. As a result, some product features commented in reviews may not be properly identified.

Aiming for improving the accuracy of product feature extraction from online consumer reviews, this paper proposes a novel integrative approach by combining an extended PageRank algorithm that leverages relationships between product features and sentiment terms, feature expansion by adding relevant synonyms, and implicit feature inference. Specifically, the proposed method pre-processes online consumer reviews using a lexical analysis tool, then constructs a network based on term pairs of candidate product features and sentiment words. Next, an extended PageRank algorithm called NodeRank will be applied to rank all term pairs and derive a candidate feature set, which will be expanded by using a synonym lexicon and implicit feature inference to generate a final product feature set.

This study contributes to the literature in three important aspects. First, we develop an extended PageRank algorithm, namely NodeRank, to find possible product features discussed in online consumer reviews by ranking candidate terms. The evaluation shows that this approach significantly outperforms two current feature extraction methods. Second, consumers often use different terms to represent the same product features. They may also describe product features in an informal language, jargons, or even concocted terms [46]. The product features implicitly referred to by those terms are difficult to be identified by existing methods. To address this limitation, the proposed approach infers implicit product features based on the importance of feature-sentiment term pairs. Third, our proposed approach expands an initial feature set by incorporating synonyms of candidate feature terms based on a synonym lexicon. Results of empirical evaluation show the positive effect of this feature expansion on the performance of product feature extraction. The proposed feature extraction method enables personalized, feature-oriented summarization and presentation of online consumer reviews, as some online retailers such as Taobao.com and jd.com have already started to provide. It can significantly alleviate the information overload problem and fasten consumers' purchase decisions. Product manufacturers can also get valuable feedback from consumers on individual product features.

The rest of the paper is organized as follows. Section 2 will introduce the related research on product feature extraction. Then, the proposed product feature extraction method will be presented in Section 3, followed by the description of evaluation and results

in Section 4. Finally, we discuss major research findings and practical implications of this study in Section 5.

2. Related work

Current approaches to product feature extraction from online consumer reviews can be classified into three categories: lexicon based, dependency relation based, and machine learning based approaches. We will introduce these three approaches in the rest of this section.

2.1. Lexicon based feature extraction approaches

This type of approach extracts product features based on a lexicon that includes manually predefined product-dependent features. For example, Poria et al. [34] proposed a rule-based approach that exploited common-sense knowledge and sentence dependency trees to identify product features. A predefined implicit feature lexicon was employed to extract product features and an opinion lexicon was used to analyze users' feature inclination. Li et al. [27] firstly constructed a lexicon that contained a list of feature words and opinion words and used it to assign a polarity tag to each feature. Then a natural language processing technique and a statistical method are applied to extract feature words and opinion words based on the defined lexicon.

A lexicon-based approach can be easily implemented. However, the effectiveness of this approach highly depends on predefined feature lexicons. With the rapid increase of online products, it is impractical to manually develop and update feature lexicons for every product.

2.2. Dependency relation based feature extraction approaches

A dependency relation based approach extracts product features based on dependency relations among terms appeared in review sentences. Researchers find that there often exist relations between feature terms and sentiment words in review sentences, which may help extract product features [35]. Such approaches firstly analyze term dependency relations in review sentences, then apply some rules and algorithms to extract product features from the identified dependency relations. For example, Qiu et al. [35] suggested that product features and sentiment words always co-exist with some particular dependency relations. They extracted dependency relations from review sentences using a syntactic parser, and applied the DP (Double Propagation) algorithm to extract product features and sentiment words from them. Zhang et al. [48] used the DP algorithm to extract features from term dependency relations between opinion words and features, and applied the HITS (Hyperlink Induced Topic Search) algorithm to rank features based on their relevance and occurrence frequency. The evaluation result showed that the HITS algorithm was effective for ranking candidate features in feature extraction. Huang et al. [20] treated feature extraction as a sequence labeling task and used the CRF (Conditional Random Fields) model to extract product features from dependency relations. They incorporated part-of-speech features and sentence structure features, which were analyzed by a syntactic dependency parser, into the CRF's learning process. Ahmad et al. [4] mainly analyzed dependency relations of short sentences in reviews, in which dependency relations may be simpler, to improve extraction performance.

Approaches to feature extraction based on dependency relations are product independent and can be applied to large-scale review data. However, because of the complexity of a language, it is difficult to identify all dependency relations in reviews. The omissions of some dependency relations will greatly affect

extraction performance. Therefore, after dependency relation analyses, some additional feature extraction steps will be necessary.

2.3. Machine learning approaches to feature extraction

Product features are normally nouns or noun phrases, so another general approach is to tag nouns and noun phrases in reviews and then extract candidate product features from them by using some machine learning algorithms. Eirinaki et al. [12] proposed the HAC (High Adjective Count) algorithm to extract potential product features. They assumed that words describing product features in reviews would always be nouns and sentiment words would always be adjectives. After part-of-speech (PoS) tagging, a process of generating a syntactic tag (e.g., noun, verb, adverb, and adjective) for each word in a sentence, they calculated the number of adjectives that were close to a noun as the score for that noun. They found that the nouns with higher scores are more likely to be important and distinguishing features. Hu et al. [19] applied association rule mining to extract product features from reviews. They used NLProcessor, a linguistic parser, to perform sentence segmentation and part-of-speech tagging. Nouns and noun phrases were extracted as transactions, and frequent itemsets were identified with the Apriori algorithm. Finally, product features were derived by trimming frequent itemsets. Wong and Lam [43] extracted product features and discovered hot item features from multiple auction Web sites based on an undirected graph model. They formulated the tasks of product feature extraction and hot item feature summarization as a single graph labeling problem using conditional random fields. The experiment result demonstrates that as a general model for capturing dependence among different variables, an undirected graph model has a good expandability for general data mining problems. Wei et al. [41] proposed a semantic-based product feature extraction technique. They applied an association rule mining algorithm to find all frequent item sets in a review corpus, and then used a list of positive and negative adjectives defined in the General Inquirer [22] as opinion words to refine the identified set of frequent features.

This type of approach is also product independent because it does not need any a priori product knowledge and can also be applied to large-scale review data [28]. However, the machine learning approach mainly relies on probabilities of nouns and noun phrases being product feature terms calculated based on their occurrence frequencies. It ignores the fact that high-frequency noun or noun phrases are not necessarily product features, or vice versa.

3. EXPRS: a new integrative feature extraction method

To address those problems of existing approaches, we propose an integrative method called EXPRS (An EXTended PageRank algorithm enhanced by a Synonym lexicon) to extract product features from a review corpus. With EXPRS, dependency relations between nouns/noun phrases and sentiment terms will be identified and analyzed, and filtering rules will be applied to generate candidate product features, which will be ranked by an extended PageRank algorithm called NodeRank. The ranked candidate feature set will be expanded by using a synonym lexicon, and implicit product feature inference will be performed to derive the final product feature set. Fig. 1 shows an overview of EXPRS, in which the input is text reviews of a specific product and the output is a collection of extracted product features.

The proposed method mainly includes the following steps: word segmentation, PoS, dependency relation analysis, rule-based

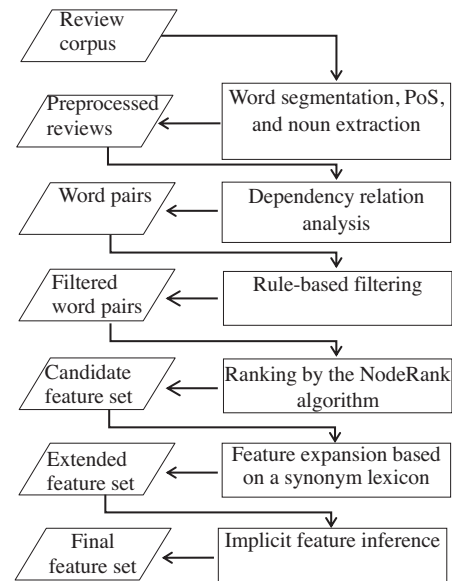


Fig. 1. An overview of the EXPRS method for feature extraction.

filtering, candidate feature ranking by NodeRank, feature expansion, and implicit feature inference.

3.1. Word segmentation and part-of-speech (PoS) tagging

First, preprocessing of product reviews is done to remove useless characters, such as special symbols. Then, EXPRS performs word segmentation and PoS tagging on individual product reviews. Unlike in English sentences, two adjacent words in a Chinese sentence are not separated by a blank space. So word segmentation, which separates each word from its adjacent others in a sentence, is a basic operation. Considering product features are normally nouns or noun phrases, generating PoS tags and identifying nouns are essential to finding candidate product features. We use a Chinese lexical analysis tool called ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System, <http://www.ictclas.org/>) for review sentence segmentation and PoS tagging. The following example shows a parsed Chinese review sentence with PoS tags:

Original sentence: “????, 好?????” (i.e., “the price is affordable, and the appearance is also very nice”).

Result of PoS tagging by ICTCLAS: “??(price)/n ??(affordable)/an, /w 好?(appearance)/n ??(also)/d ??(very)/d ??(nice)/a”,

in which the tags ‘n’ represents a noun; ‘an’ represents an adnoun; ‘w’ represents a punctuation; ‘d’ represents an adverb; and ‘a’ represents an adjective.

Based on the result of ICTCLAS, an initial candidate product feature set that includes all identified nouns and noun phrases will be derived and sentiment words associated with them will be identified as well.

3.2. Analysis of dependency relations

Dependency syntax describes sentence structures with a set of dependency relations [37]. A dependency relation is an asymmetric binary relationship between a term called head or governor and another term called modifier or dependent [31]. An example of a sentence with dependency relations is shown in Fig. 2, in which a head ‘price’ and a modifier ‘OK’ form a dependency relation ‘nsubj’, while a head ‘very’ and a modifier ‘beautiful’ form another dependency relation ‘avdmod’.

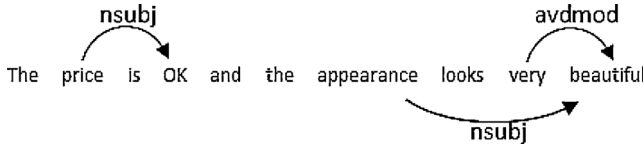


Fig. 2. The syntactic structure of a sentence consisting of dependencies.

A dependency syntax analysis can identify a term's role and dependency relations in a sentence. In this research, candidate product features and associated sentiment terms are identified and analyzed through a dependency relation analysis using the Stanford Parser (<http://nlp.stanford.edu>).

There can be different dependency relations in sentences, but only some of them are helpful for identifying product features. Similar to Wu et al.'s study [44], we consider four major types of dependency relations, including 'nsbj', 'amod', 'rcmod' and 'dobj', in the product feature extraction process. They refer to subject-predicate relations, adjectival modifying relations, relative clause modifying relations, and verb-object relations, respectively. In the example shown in Fig. 2, there are three dependency relations: 'price' and 'OK' form an 'nsbj' relation; 'appearance' and 'beautiful' create another 'nsbj' relation; and 'very' and 'beautiful' form an 'avdmod' relation (an adverbial clause modifier). We only consider two-term pairs with a dependency relation of 'nsbj', such as (price, OK) and (appearance, beautiful), as candidate feature-sentiment pairs.

Based on a set of dependency relations derived from the Stanford Parser, we extract those four dependency relations and generate a set of candidate feature-sentiment pairs from all product reviews. The method is illustrated in Fig. 3.

3.3. Rule-based filtering

Some nouns or noun phrases (e.g., brand names) are not product features. Thus we develop some filtering rules to exclude non-feature nouns or noun phrases by extending those used in [26]. In this study, if an initial candidate feature belongs to one of the following types of nouns or noun phrases listed in Table 1, it will be removed from the initial feature set. In the meantime, corresponding term pairs will be eliminated from candidate pairs generated by the previous steps.

3.4. NodeRank: an extended PageRank algorithm

PageRank is a well-known effective website ranking algorithm [6] used for determining key network nodes [13]. The intuition

INPUT:

$T_n = \{d_1(F_1, S_1), d_2(F_2, S_2), \dots\}$ a set of dependency relation pairs in a review

$V = \{T_1, T_2, \dots, T_n\}$ a set of T in an entire review corpus

$D = \{\text{nsbj}, \text{amod}, \text{rcmod}, \text{dobj}\}$ four dependency relations

OUTPUT:

$P = \{(F_1, S_1), (F_2, S_2), \dots, (F_n, S_n)\}$ a set of candidate feature-sentiment pairs

PROCESS (T_n, V, D):

begin

for each T in V :

for each $d_i(F, S)$ in T :

if d_i is in D then P.append(F, S)

// adding the current pair to the candidate feature-sentiment set

return P

Fig. 3. Process of dependency relation extraction.

Table 1

Examples of non-feature nouns or noun phrases.

Types	Examples
(1) Proper nouns (time, place, name)	September, Beijing, Tom
(2) Brand names	Canon, Samsung, Apple
(3) Verbal nouns	Feeling, something
(4) Personal nouns	Friend, father

behind the PageRank algorithm is that the importance of a website depends on other websites that link to it [39]. Treating hyperlinks' value as a ranking factor, PageRank distinguishes important websites from less important ones. The importance value of a website i should be equally distributed to those that i links to. If there are a total of N websites, the vector $P = (p(1), p(2), \dots, p(N))^T$ describes the importance values (i.e., PageRank values) of those websites. The coefficient matrix A , which describes the adjacency relations among different websites, is defined as follows:

$$A_{ij} = \begin{cases} \frac{1}{O_i}, & \text{if } O_i > 0, (i, j) \in E \\ \frac{1}{N}, & \text{if } O_i = 0 \end{cases} \quad (1)$$

where E is the set of directional links among websites; O_i is the number of websites to which the website i links. Then, the vector P can be described as follows:

$$P = \left((1 - \alpha) \frac{1}{N} + \alpha A^T \right) P \quad (2)$$

where I is a unit matrix; α is a damping coefficient. A power iteration method can be used to calculate the PageRank value of each website [16].

In the context of feature extraction, a noun or noun phrase in an online consumer product review is more likely to be a feature term if it is modified by multiple adjectives in reviews [12] because people tend to use adjectives to express their sentiment inclinations toward specific product features [41]. Similarly, an adjective is more likely to be a sentiment word if it modifies multiple product feature terms. Thus a pair of a product feature and its sentiment word can be treated as a network node, and modifying relations between different candidate features and sentiment words can be treated as network edges. We call such a self-organized network as a feature-sentiment term relation network.

The PageRank algorithm measures the importance of network nodes based on the link structure of a network. For a feature-sentiment term relation network, the importance of a network node reflects the extent to which sentiment words modify the product feature in that network node, which is determined by links from other network nodes. The links from important nodes will carry more weights than links from less important nodes, and the importance of a node will be shared by all nodes that it links to. A network node with high importance is normally linked by many other important nodes, reflected by a large number of adjectives modifying the product feature that the node contains. Thus, the higher the importance of a network node, the higher the possibility of the node containing a product feature. Then the proposed NodeRank algorithm that extends the notion of PageRank is used to identify the importance of network nodes and derive product features.

Constructing a node-link graph is the first step of the NodeRank algorithm. Assuming that W represents a set of dependency pairs; F represents a set of all feature terms in W ; S is the set of all sentiment terms in W ; and the Cartesian product of F and S is V . A node in the node-link graph is a term pair, which is an element of V . A sentiment term s in a node v_1 and a feature term f in a node v_2

compose a pair (f, s). If (f, s) is an element of W, v_1 and v_2 are connected by a link with the direction from v_1 to v_2 , indicating that the sentiment term in v_1 modifies the feature term in v_2 . For example, if W has two elements (speed, fast) and (memory, large); V has four elements (memory, fast), (speed, fast), (speed, large), and (memory, large); and the edge set E in a graph has six elements (v_1, v_2), (v_1, v_3), (v_2, v_3), (v_3, v_1), (v_3, v_4) and (v_4, v_1), the node-link graph of this example is shown in Fig. 4.

Based on the above graph construction process, we can derive the entire network and rank feature-sentiment term pairs.

Considering that the more often a noun occurs in a review, the more likely it is a product feature, we define the importance $p(i)$ of network node i as follows:

$$p(i) = (1 - \alpha)H(i) + \alpha \sum_{(j,i) \in E} \frac{p(j)}{O_j} \quad (3)$$

where α is a damping coefficient; $p(j)$ is the importance value (i.e., NodeRank value) of node j ; O_j is the number of nodes to which the node j links; and E is the edge set in a node-link graph; $H(i)$ is the function of occurrence frequency of node i , which is defined as:

$$H(i) = \frac{N \ln f(i)}{\ln(\prod_{j=1}^N f(j))} \quad (4)$$

where N is the number of nodes in the graph; $f(i)$ is the occurrence frequency of node i in an entire review corpus. The log function can reduce the adverse impact of high occurrence frequency of some less relevant terms on node ranking. The formula (3) satisfies the three requirements of a Markov chain and can reach convergence [16]. The vector $P = (p(1), p(2), p(3), \dots, p(N))^T$ can be calculated by using a power iterative algorithm [16].

After dependency pairs are extracted, a node-link graph (or network) will be constructed. The NodeRank value of each dependency pair will be calculated by the NodeRank algorithm. Pairs with NodeRank values higher than a threshold will be included in the refined product feature candidate list.

3.5. Feature extension

The proposed EXPRS integrates synonym expansion and an implicit feature inference process with the NodeRank algorithm. Synonym expansion enriches a candidate feature set by adding synonymous nouns or noun phrases of candidate features. Implicit feature inference is aimed to identify product features to which are implicitly referred in reviews.

3.5.1. Synonym expansion

Although an initial product feature set is derived by the NodeRank algorithm, uncommon product features may not be extracted. At the same time, because of informal languages used in online reviews, a product feature may be often referred to by different terms [29]. Thus some nouns or noun phrases appeared in reviews but not included in an initial product feature set may also represent product features.

In this study, we integrate a synonym lexicon created by the Research Center of Social Computing and Information Retrieval

[17] with the NodeRank algorithm to expand an initial feature set by adding synonyms of each candidate feature term. For example, 'shot' has a synonym word 'lens'. If 'shot' is extracted as a feature term while 'lens' is not, the latter will be added to the feature set as well.

3.5.2. Implicit feature inference

Normally, each review should comment on one or more features of a product with some sentiment orientations [42,47]. Because of the diversity in online review expressions, some review sentences only include sentiment words without associated feature terms. The product features that do not appear in review sentences but are actually referred to are called implicit features [10]. In this study, we infer implicit product features based on the result of the NodeRank algorithm.

People like to use similar sentiment words to describe the same product features [38]. In this research, possible implicit features are inferred based on product features described by the same sentiment words in other reviews. This follows the philosophy of association rule mining [24] and Naive Bayesian classification [14]. Coping with reviews that do not include any feature words, the proposed EXPRS method searches for dependency pairs that include the same sentiment words and chooses a product feature with the highest NodeRank value, which indicates the highest probability of being an implicit product feature. For example, there are no explicit features mentioned in the following review sentence about a mobile phone "Very large, I like it". EXPRS finds that there are two term pairs (screen, large) and (memory, large) in the set of dependency pairs, with NodeRank values of 0.5 and 0.2, respectively. Then there will be a higher probability for 'screen' to be the implicit product feature associated with the sentiment term 'big' in this sentence than for 'memory'.

Based on the candidate feature set generated by the NodeRank algorithm, the synonyms of each candidate feature term will be added. Then implicit features will be inferred to finalize the product feature set.

4. Evaluation

We have conducted an empirical evaluation of the proposed EXPRS approach by using online consumer reviews of three different products collected from jd.com and using the methods proposed by Eirinaki et al. [12] and Zhang et al. [48] as benchmarks.

4.1. Data acquisition

We crawled online consumer reviews of three products from <http://www.jd.com> generated between December 2012 and August 2013. The three products included one digital camera, one mobile phone, and one DVD player, with 8901, 11,291, and 9,000 reviews, respectively. Those three product categories are the most commonly used ones in previous studies [3,36,41]. Digital camera and cell phone are search goods [32], while DVD player is experience goods [40]. The descriptive information about those reviews is listed in Table 2.

All selected reviews satisfied three criteria: (1) the length of a review was more than 10 words; (2) the number of non-duplicated characters was at least seven (including punctuation marks); and (3) default technical reviews provided by manufacturers were excluded from the analysis.

Three graduate students were asked to identify product features discussed in those reviews manually and individually. Then, they discussed disagreed features to see if a consensus can be reached. Eventually, only the features that all three reviewers had

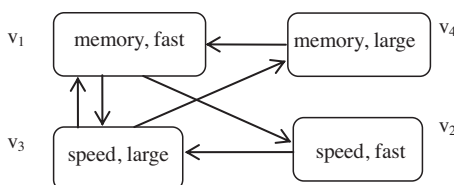


Fig. 4. A Node-Link graph.

Table 2
Data description.

Product name	Number of reviews	Average length (# of words)	Number of manually extracted features
Samsung GalaxyNote II	8901	74	68
Canon EOS 600D	11,291	72	55
Philips DVP3600	9000	52	59

agreed upon at the end were included in the correct benchmark product feature sets, which were then compared against the feature sets automatically extracted from consumer product reviews by the proposed approach and two baseline approaches. The numbers of finally agreed features of three products are shown in the last column of Table 2.

4.2. Metrics

We used precision (P), recall (R), and F-measure (F) as metrics to assess the effectiveness of the proposed EXPRS approach. Specifically, precision is the fraction of extracted features that are correct; recall is the fraction of correct features that are actually extracted among all correct features; F-measure is a harmonic average of precision and recall, which reflects the overall performance of precision and recall. Higher values of P, R, and F indicate better performance.

4.3. Benchmarks

To evaluate the effectiveness of the proposed approach, the HAC (High Adjective Count) algorithm proposed by Eirinaki et al. [12] and a feature extraction approach based on the HITS (Hyperlink-Induced Topic Search) algorithm proposed by Zhang et al. [48] are chosen as benchmarks. There are a few reasons why we selected those two approaches as the baseline methods. First, we could not find any previous work that adopted the PageRank algorithm to extract features from online reviews. Second, our proposed model identifies a final feature set by ranking candidate product features, so it makes sense to use methods that have also adopted such a ranking approach as a baseline. Third, we want to select methods that have been reported to perform very well in online review feature extraction. Fourth, those methods should be published recently. The two selected baseline methods satisfied all these criteria. The HAC algorithm assumes that a noun or noun phrase is more likely to be a feature word if it is modified by more adjectives in all reviews [12]. Thus, the number of modifying adjectives will be assigned as the score of a noun. Then, HAC ranks all nouns based on their scores and selects those nouns with a score higher than a threshold as product features. Similar to PageRank, HITS is also a link analysis method that analyzes the importance of websites [23]. The basic idea of HITS is that an authority website is pointed by many hub websites, and a hub website points to many authority websites. Zhang et al. [48] have applied the HITS algorithm to product feature extraction by ranking feature candidates based on feature importance, which was determined by feature relevance and feature frequency.

Both HAC and HITS have a feature ranking process after initial feature extraction, which is similar to the proposed approach. Meanwhile, HITS and PageRank are classic website importance ranking algorithms. Therefore, it is appropriate to use these two baseline methods in this study.

4.4. Data analysis and results

At first, we conducted experiments to determine the optimal threshold value for node ranking. Similar to Wang [39], we applied

the EXPRS method to 1000 randomly chosen reviews and assessed its performance with different node ranking threshold values between 0 and 1. At the end, we chose the optimal threshold value of 0.033 for the cell phone, 0.05 for the camera and 0.06 for the DVD player, which resulted in the highest F-measures for the chosen 1000 reviews.

We carry out three analyses on the collected data sets to examine the potential usefulness of synonym expansion and implicit feature inference by comparing the performance of the following approaches: the proposed EXPRS method without synonym expansion (denoted as M1), the proposed EXPRS method without implicit feature inference (denoted as M2), and the proposed integrated EXPRS method (i.e., containing both synonym expansion and implicit feature inference, denoted as M3).

We randomly divided reviews of each product into 30 groups with equal size. Precision of each group was calculated for every approach. Precision values of M3 were then compared against those of M1 and M2 to examine whether the improvements of M3 over M1 and M2 were statistically significant. Then, we conducted paired *T*-tests to examine whether the improvements of EXPRS over HAC and HITS based feature extraction methods were statistically significant. Some intermediate results of the proposed integrated EXPRS method are given in Table 3. Means and standard deviations (SD) of precision, recall, F-measure, and *T*-test (*df* = 29) results are shown in Tables 4–6, respectively.

As shown in Tables 4–6, the proposed integrative EXPRS method (M3) achieves better performance in precision, recall, and F-measure than M1 and M2 across all three products, which indicates that incorporating synonym expansion and implicit feature inference can help improve the performance of feature extraction significantly.

Results also reveal that the proposed NodeRank-based EXPRS approach enhanced by both synonym expansion and implicit feature inference (i.e., M3) achieves significantly better performance than the two baseline algorithms in precision, recall, and F-measures across all three products consistently. The proposed EXPRS method without synonym expansion (i.e., M1) or without implicit feature inference (i.e., M2) generally outperforms the two baseline algorithms, but the performance varies among different products and metrics. HAC and HITS methods result in inconsistent performance with different products. We examined the effect size using Cohen's *d*. The vast majority of calculated effect size values indicated medium to large effects. Table 7 presents effect sizes of different comparisons in F-measures as an example.

Although feature extraction is normally done offline, we still assessed the execution time of the proposed method and the two baseline methods using 1,000 randomly selected reviews. All tests were performed on a PC with 3.4 GHz Intel i7-4770 CPU and 4G RAM. The processing time of all reviews by HAC, HITS, and EXPRS were 545, 726, and 792 s, respectively.

5. Discussion

5.1. Major findings

Consumer reviews are valuable and helpful for others to make better and quicker purchase decisions. However, with the fast

Table 3
Intermediate results of EXPRS.

Product name	Number of identified dependency pairs	Number of network edges	Number of extracted features
Samsung GalaxyNote II	3509	119,305	61
Canon EOS 600D	2980	135,832	52
Philips DVP3600	3116	99,641	56

Table 4

Comparisons of Precisions (%) of HAC, HITS and EXPRS.

	P1(HAC)		P2(HITS)		P3(M1)		P4(M2)		P5(M3)		P1–P5	P2–P5	P1–P3	P1–P4	P2–P3	P2–P4	P3–P5	P4–P5
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)
Samsung GalaxyNote	71.1	6.54	75.0	7.47	79.7	7.12	76.3	8.42	81.3	6.85	–10.2**	–6.3**	–8.6**	–5.2*	–4.7*	–1.3	–1.6*	–5*
Canon EOS 600D	69.8	8.59	71.4	5.34	77.5	5.69	76.5	11.71	80.2	5.56	–10.4**	–8.8**	–7.7**	–6.7**	–6.1**	–5.1*	–2.7*	–3.7*
Philips DVP3600	71.9	7.74	71.0	9.29	77.8	5.98	75.9	4.13	80.0	7.26	–8.1*	–9**	–5.9**	–4*	–6.8**	–4.9*	–2.2*	–4.1*

* $p < 0.05$.** $p < 0.01$.**Table 5**

Comparisons of Recalls (%) of HAC, HITS and EXPRS.

	R1(HAC)		R2(HITS)		R3(M1)		R4(M2)		R5(M3)		R1–R5	R2–R5	R1–R3	R1–R4	R2–R3	R2–R4	R3–R5	R4–R5
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)
Samsung GalaxyNote	66.1	9.21	62.4	4.71	67.8	7.92	65.5	5.86	72.9	7.33	–6.8**	–10.5**	–1.7	0.6	–5.4**	–3.1*	–5.1**	–7.4*
Canon EOS 600D	62.1	4.86	61.2	5.36	63.7	5.55	64.1	5.41	68.2	8.28	–6.1**	–7.0**	–1.6	–2	–2.5	–2.9*	–4.5*	–4.1*
Philips DVP3600	63.5	7.65	67.0	7.29	65.2	9.83	66.0	6.94	69.1	5.71	–5.6**	–2.1*	–1.7	–2.5*	1.8	1	–3.9*	–3.1**

* $p < 0.05$.** $p < 0.01$.**Table 6**

Comparisons of F-measures (%) of HAC, HITS and EXPRS.

	F1(HAC)		F2(HITS)		F3(M1)		F4(M2)		F5(M3)		F1–F5	F2–F5	F1–F3	F1–F4	F2–F3	F2–F4	F3–F5	F4–F5
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)	(MD)
Samsung GalaxyNote	68.4	10.54	67.8	5.25	73.2	6.48	70.4	5.74	76.2	8.43	–7.8**	–8.4**	–4.8*	–2	–5.4**	–2.6*	–3*	–5.8*
Canon EOS 600D	65.3	6.58	65.6	7.34	69.9	6.76	69.7	4.8	73.0	6.98	–7.7**	–7.4**	–4.6*	–4.4*	–4.3*	–4.1*	–3.1*	–3.3*
Philips DVP3600	66.7	8.11	68.9	7.62	70.9	7.44	70.6	7.92	72.7	9.34	–6*	–3.8*	–4.2*	–3.9*	–2*	–1.7	–1.8*	–2.1*

SD: Standard Deviation, MD: Mean Difference.

* $p < 0.05$.** $p < 0.01$.

growth of the volume of online consumer reviews, it is impossible for consumers to browse every review of a product. Considering different consumers may have different interests or preferences in different product features, a product feature-based analysis of consumer reviews is very important and essential to enabling consumers to have a better understanding of products before making a purchase decision. To achieve that goal, product feature extraction from online consumer reviews will be the essential first step.

Pre-defined lexicon based approaches, dependency relation based approaches, and machine learning approaches are three dominant product feature extraction methods explored in the literature. However, existing approaches have rarely considered network structures in extracting feature-sentiment dependency pairs and implicit product features. In addition, the impact of incorporating synonyms of candidate feature terms on the performance of product feature extraction has not been well explored either.

In this study, we propose a new feature extraction method by extending the PageRank algorithm and incorporating synonym expansion and implicit feature inference to extract product features from online consumer product reviews. The proposed method firstly segments review sentences, generates PoS tags for individual words, and obtains dependency relations. Based on the ranking process of the proposed NodeRank algorithm, candidate product features will be identified. Then the synonym expansion and implicit feature inference are performed to generate a final feature set. The effectiveness of the proposed method has been empirically evaluated by using real-world data and two baseline methods. The results show that the proposed integrative method outperforms two baseline methods in

precision, recall, and F -measure across all three different products. Results also show that incorporating synonym expansion and implicit feature inference helps improve extraction performance significantly. The proposed method takes a little longer time to process reviews than the two baseline methods mainly because the proposed method has to construct a dependency relation network and calculate the importance of every node in that network. However, because feature extraction is done offline, the processing time of the proposed method is acceptable.

This research provides three-fold novel research contributions. First, we introduce a new dependency relation network for online reviews that can be constructed based on feature-sentiment dependency pairs and extend the PageRank algorithm to rank the importance of candidate dependency pairs to identify product features. Second, we develop an effective mechanism for implicit feature inference, which has never been done in prior feature extraction research. Third, we extend a candidate feature set with synonyms of candidate features, which has not been well explored in the literature. Although we used Chinese online consumer reviews in empirical evaluation in this study, the proposed approach is generic and should be applicable to feature extraction from online consumer reviews written in other languages.

5.2. Practical implications

In addition to research contributions, this research also provides some practical implications. For consumers, it would be beneficial for them to browse reviews that comment on specific product features that they are interested in. Currently, very few

Table 7Effect size (Cohen's *d*) of *F*-measure comparisons.

	F1–F5	F2–F5	F1–F3	F1–F4	F2–F3	F2–F4	F3–F5	F4–F5
Samsung GalaxyNote II	–0.83	–1.22	–0.56	–0.24	–0.93	–0.48	–0.41	–0.82
Canon EOS 600D	–1.15	–1.05	–0.70	–0.78	–0.62	–0.67	–0.46	–0.56
Philips DVP3600	–0.70	–0.45	–0.55	–0.49	–0.27	–0.22	–0.22	–0.25

current e-commerce websites offer personalized filtering, summarization, and presentation of online consumer product reviews to consumers according to their product feature preferences. Therefore, consumers cannot fully take advantage of rich opinions shared in online reviews. This research proposes an effective method to enable an online retailer to provide an online consumer review platform that is product feature oriented. By automatically summarizing and presenting feature-specific reviews, consumers can compare different products along the same set of product features and make better purchase decisions. The proposed method can help online retailers build a feature-oriented personalized review search and summarization system. As a result, consumers' satisfaction and loyalty to an online retailer can be improved, and the buy conversion rate could be increased greatly.

For manufactures, the proposed method may enable them to cluster and analyze consumer reviews based on product features commented and obtain consumers' positive and negative feedback on specific product features timely. Through such a feature-oriented review analysis, manufacturers can identify major product flaws and potential improvement of products easily [1].

5.3. Limitations and future research

There are several limitations of this study, which provide opportunities for future research. First of all, like every other previous study in the literature, only reviews of a small number of products (i.e., three) are analyzed in this research. Although our study uses reviews on both experience and search goods and achieves consistent better performance than baseline methods, reviews of some other products may have different characteristics. Although theoretically the proposed method is generic and product independent, future work should be conducted to further test the generalizability of the proposed method using different products.

Second, due to the scope and complexity of this study, we adopt two alternative feature extraction methods used in previous studies as the baseline methods, namely HAC and HITS methods. There are many other methods developed for feature extraction from online product reviews. In the future, we may compare the proposed approach with other methods as well.

Finally, the evaluation mainly focuses on feature extraction performance, i.e. precision, recall, and *F*-measure, among three methods. Although the performance of the proposed method is better than two baseline methods, the initial result of computational cost evaluation results show that the running time of the proposed method can be improved. Though it is not a major concern given that such a feature extraction process is always performed offline, optimizing computational complexity of a feature extraction process can be useful and should be investigated in future research.

Acknowledgment

This research is supported by the National Natural Science Foundation of China (Award #: 71128003, 71272057) and National Key Technology Support Program (Award #: 2013BAH16F00).

References

- [1] A.S. Abrahams, W. Fan, G.A. Wang, Z.J. Zhang, An integrated text analytic framework for product defect discovery, *Product. Oper. Manag.* 2014 <http://dx.doi.org/10.1111/poms.12303>.
- [2] A.S. Abrahams, J. Jiao, W. Fan, G.A. Wang, Z. Zhang, What's buzzing in the blizzard of buzz? Automotive component isolation in social media postings *Decis. Support Syst.* 55 (4), 2013, pp. 871–882.
- [3] S. Aciar, D. Zhang, S. Simoff, J. Debenham, Informed recommender: basing recommendations on consumer product reviews, *IEEE Intell. Syst.* 22 (3), 2007, pp. 39–47.
- [4] T. Ahmad, M.N. Doja, Rule based system for enhancing recall for feature mining from short sentences in customer review documents, *Int. J. Comput. Sci. Eng.* 4 (6), 2012, pp. 1211–1219.
- [5] M. Aljukhadar, S. Senecal, C.E. Daoust, Using recommendation agents to cope with information overload, *Int. J. Electron. Commer.* 17 (2), 2012, pp. 41–70.
- [6] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, *Comput. Netw. ISDN Syst.* 30 (1), 1998, pp. 107–117.
- [7] H. Chen, R.H. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact, *MIS Q.* 36 (4), 2012, pp. 1165–1188.
- [8] Y.-C. Chen, R.-A. Shang, C.-Y. Kao, The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment, *Electron. Commer. Res. Appl.* 8 (1), 2009, pp. 48–58.
- [9] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: Online book reviews, *J. Market. Res.* 43 (3), 2006, pp. 345–354.
- [10] X. Ding, B. Liu, P.S. Yu, A Holistic Lexicon-based Approach to Opinion Mining, *The International Conference on Web Search and Web Data Mining*, Palo Alto, California, USA, 2009 pp. 231–239.
- [11] A. Duric, F. Song, Feature selection for sentiment analysis based on content and syntax models, *Decis. Support Syst.* 53 (4), 2012, pp. 704–711.
- [12] M. Eirinaki, S. Pisal, J. Singh, Feature-based opinion mining and ranking, *J. Comput. Syst. Sci.* 78 (4), 2012, pp. 1175–1184.
- [13] W.G. Fan, M.D. Gordon, The power of social media analytics, *Commun. ACM* 57 (6), 2014, pp. 74–81.
- [14] P.A. Flach, N. Lachiche, Naive bayesian classification of structured data, *Mach. Learn.* 57 (3), 2004, pp. 233–269.
- [15] A. Ghose, P.G. Ipeirotis, B.B. Li, Examining the impact of ranking on consumer behavior and search engine revenue, *Manag. Sci.* 60 (7), 2014, pp. 1632–1654.
- [16] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 2012.
- [17] HIT-SCIR, Tongyici cilin (extended), 2013–5, (<http://ir.hit.edu.cn/>).
- [18] Y. Hong, P.A. Pavlou, Product fit uncertainty in online markets: nature, effects, and antecedents, *Inform. Syst. Res.* 25 (2), 2014, pp. 328–344.
- [19] M. Hu, B. Liu, Mining opinion features in customer reviews, *The Nineteenth National Conference on Artificial Intelligence*, (San Jose, California, USA), 2004, pp. 755–760.
- [20] S. Huang, X. Liu, X. Peng, Z. Niu, Fine-grained product features extraction and categorization in reviews opinion mining, *Proceedings of the IEEE 12th International Conference on Data Mining Workshops*, (Brussels, Belgium), 2012, pp. 680–686.
- [21] IRESEARCH, China internet shopping market report in 2013, 2014–5, (<http://eciresearch.cn/shopping/20140114/224908.shtml>).
- [22] A. Kennedy, D. Inkpen, Sentiment classification of movie reviews using contextual valence shifters, *Comput. Intell.* 22 (2), 2006, pp. 110–125.
- [23] J. Kleinberg, Authoritative sources in a hyperlinked environment, *J. ACM* 46 (5), 1999, pp. 604–632.
- [24] S. Kotsiantis, D. Kanellopoulos, Association rules mining: a recent overview, *GESTS Int. Trans. Comput. Sci. Eng.* 32 (1), 2006, pp. 71–82.
- [25] R.Y. Lau, C. Li, S.S. Liao, Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis, *Decis. Support Syst.* 65, 2014, pp. 80–94.
- [26] S. Li, Q. Ye, Y. Li, R. Law, Mining features of products from chinese customer online reviews, *J. Manag. Sci. China* 12 (2), 2009, pp. 142–152.
- [27] Z.C. Li, M. Zhang, S.P. Ma, B. Zhou, Y. Sun, Automatic extraction for product feature words from comments on the web, in: G.G. Lee (Ed.), *Information Retrieval Technology*, Springer-Verlag Berlin, Berlin, 2009, pp. 112–123.
- [28] Y. Liang, R. Tan, J. Ma, Study on patent text classification for product innovative design, *Comput. Integr. Manuf. Syst.* 19 (2), 2013, pp. 382–390.
- [29] B. Liu, Sentiment analysis and subjectivity, in: N. Indurkha, F.J. Damerau (Eds.), *Handbook of Natural Language Processing*, CRC Press, Boca Raton, FL, 2010, pp. 627–666.
- [30] B. Ma, D. Zhang, Z. Yan, T. Kim, An LDA and synonym lexicon based approach to product feature extraction from online consumer product reviews, *J. Electron. Commer. Res.* 14 (4), 2013, pp. 304–314.
- [31] I.A. Mel'uk, *Dependency Syntax: Theory and Practice*, State University of New York Press, Albany, NY, 1988.
- [32] S.M. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on amazon.com *MIS Q.* 34 (1), 2010, pp. 185–200.

- [33] C.W. Phang, C. Zhang, J. Sutanto, The influence of user interaction and participation in social media on the consumption intention of niche products, *Inform. Manag.* 50 (8), 2013, pp. 661–672.
- [34] S. Poria, E. Cambria, L.-W. Ku, C. Gui, A. Gelbukh, A Rule-based Approach to Aspect Extraction from Product Reviews, *The Second Workshop on Natural Language Processing for Social Media*, Dublin, Ireland, 2014 pp. 28–37.
- [35] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, *Comput. Linguist.* 37 (1), 2011, pp. 9–27.
- [36] C. Quan, F. Ren, Unsupervised product feature extraction for feature-oriented opinion determination, *Inform. Sci.* 272, 2014, pp. 16–28.
- [37] G. Somprasertsri, P. Lalitrojwong, Mining feature-opinion in online customer reviews for opinion summarization, *J. Univers. Comput. Sci.* 16 (6), 2010, pp. 938–955.
- [38] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, Hidden sentiment association in chinese web opinion mining, *Proceedings of the 17th International Conference on World Wide Web*, (Beijing, China), 2008, pp. 959–968.
- [39] G.A. Wang, J. Jiao, A.S. Abrahams, W. Fan, Z. Zhang, Expertrank: A topic-aware expert finding algorithm for online knowledge communities, *Decis. Support Syst.* 54 (3), 2013, pp. 1442–1451.
- [40] D. Weathers, S. Sharma, S.L. Wood, Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods, *J. Retail.* 83 (4), 2007, pp. 393–401.
- [41] C.-P. Wei, Y.-M. Chen, C.-S. Yang, C.C. Yang, Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews, *Inform. Syst. E-Bus. Manag.* 8 (2), 2010, pp. 149–167.
- [42] W. Wei, H. Liu, J. He, H. Yang, X. Du, Extracting feature and opinion words effectively from chinese product reviews, *Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, (Jinan, China), 2008, pp. 170–174.
- [43] T.-L. Wong, W. Lam, Learning to extract and summarize hot item features from multiple auction web sites, *Knowl. Inform. Syst.* 14 (2), 2008, pp. 143–160.
- [44] Y. Wu, Q. Zhang, X. Huang, L. Wu, Phrase dependency parsing for opinion mining, in: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, (Singapore), 2009, pp. 1533–1541.
- [45] D. Zeng, H. Chen, R. Lusch, S.-H. Li, Social media analytics and intelligence, *IEEE Intell. Syst.* 25 (6), 2010, pp. 13–16.
- [46] Z. Zhai, B. Liu, H. Xu, P. Jia, Clustering product features for opinion mining, in: *Proceedings of the fourth ACM international conference on Web search and data mining*, (Hongkong), 2011, pp. 347–354.
- [47] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, W. Zuo, Sentiment analysis of chinese documents: from sentence to document level, *J. Am. Soc. Inform. Sci. Technol.* 60 (12), 2009, pp. 2474–2487.
- [48] L. Zhang, B. Liu, S.H. Lim, E. O'Brien-Strain, Extracting and ranking product features in opinion documents, *Proceedings of the 23rd International Conference on Computational Linguistics*, (Beijing, China), 2010, pp. 1462–1470.