AN LDA AND SYNONYM LEXICON BASED APPROACH TO PRODUCT FEATURE EXTRACTION FROM ONLINE CONSUMER PRODUCT REVIEWS

Baizhang Ma
School of Management and Economics, Beijing Institute of Technology
5 South Zhongguancun Street, Beijing, 100081, China
mabaizhang@126.com

Dongsong Zhang
Department of Information Systems, University of Maryland, Baltimore County
1000 Hilltop Circle, Baltimore, MD 21043, USA
School of Management and Economics, Beijing Institute of Technology
5 South Zhongguancun Street, Beijing, 100081, China
zhangd@umbc.edu

Zhijun Yan*
School of Management and Economics, Beijing Institute of Technology
5 South Zhongguancun Street, Beijing, 100081, China
yanzhijun@bit.edu.cn

Taeha Kim
College of Business & Economics, Chung-Ang University, Korea
84 Heukseok-Ro, Dongjak-Gu, Seoul, Korea
taehakim@gmail.com

ABSTRACT

Consumers are increasingly relying on other consumers' online reviews of features and quality of products while making their purchase decisions. However, the rapid growth of online consumer product reviews makes browsing a large number of reviews and identifying information of interest time consuming and cognitively demanding. Although there has been extensive research on text review mining to address this information overload problem in the past decade, the majority of existing research mainly focuses on the quality of reviews and the impact of reviews on sales and marketing. Relatively little emphasis has been placed on mining reviews to meet personal needs of individual consumers. As an essential first step toward achieving this goal, this study proposes a product feature-oriented approach to the analysis of online consumer product reviews in order to support feature-based inquiries and summaries of consumer reviews. The proposed method combines LDA (Latent Dirichlet Allocation) and a synonym lexicon to extract product features from online consumer product reviews. Our empirical evaluation using consumer reviews of four products shows higher effectiveness of the proposed method for feature extraction in comparison to association rule mining.

Keywords: Online product reviews; Feature extraction; Latent Dirichlet Allocation; Synonym lexicon; Data mining

1. **Introduction**

With the rapid advance of Internet technology, online shopping becomes increasingly popular. iResearch reports that the e-commerce market worldwide reached US$1.99 billion in Fall 2012, which was 21.9% more than 2011 [iResearch 2012]. However, online shopping makes it impossible for consumers to obtain the first-hand experience and knowledge about product quality through direct touching, seeing, and trying out products before purchase.

With the rise of Web 2.0 technology, users have become more and more comfortable with sharing their opinions on products or services over the Internet. Examples of these Web 2.0 web sites include e-commerce sites

* Corresponding author

such as Amazon.com, social network sites like Facebook.com, local service provider review sites like Yelp.com, and digital media distributors like Netflix.com. Empirical studies have shown that user reviews indeed have significant impact on sales, and users rely much more on reviews than on numeric ratings to decide if a product should be purchased [Chevalier & Mayzlin, 2006, Li et al., 2009]. Consumers often rely on other consumers' online reviews of features and quality of products while making purchase decisions because those reviews offer more useful information about products than the descriptive information available on a shopping website provided by manufacturers [Park and Kim 2008, Hu et al. 2009]. According to a survey, 70 percent of Americans regularly consult online consumer product reviews or ratings before making an important purchase [Ante 2009].

The rapid growth of the volume of online consumer product reviews, however, makes browsing a large number of reviews to identify information of interest time consuming and cognitively demanding. It poses a significant challenge for taking advantage of those reviews. Review mining that uses data mining and text mining techniques for analyzing a large quantity of online consumer product reviews [Popescu and Etzioni 2007] has been extensively explored as a means of addressing this information overload problem. The majority of existing research, however, focuses on prediction of helpfulness of reviews, determining the polarity of consumer opinions on products, and the impact of reviews on sales and marketing, etc. Relatively little emphasis has been placed on mining reviews to understand consumers' opinions toward individual features of specific products. Product features include product attributes, components, component attributes, etc. [Xi et al. 2011].

In reality, individual consumers may have different preferences for features of a product. For example, while purchasing a mattress, a consumer may consider the firmness of a mattress as the most and only important feature, while another may not care about the firmness but prefer a mattress with a pillow top. The third consumer may look for a mattress with a 10-year warranty. Although e-Commerce web sites often introduce some product features, there are several problems. First, most product features are described in technical terms that are not acquainted by regular consumers. Second, feature descriptions are usually very short. There is no justification or discussion on the quality and usefulness of features, which is very important to consumers. Third, consumer reviews may contain product features that are not introduced or mentioned on a product web page. Performing a feature-based review analysis can enable an online review system to categorize and visualize reviews based on features discussed in them. By doing this, a system can provide consumers with a group of relevant reviews, aiming to provide personalized feature-based summaries and presentation of reviews [Liu et al. 2005; Somprasertsri and Lalitrojwong, 2010]. The first step toward achieving the above goal is to identify and extract product features from online consumer product reviews so that reviews can be organized and indexed based on extracted features. Most existing online consumer review systems do not provide a 'keyword search' function to allow consumers to search for reviews that contain certain feature keywords. The keyword search function of an Internet browser can only search for a keyword page by page, which could be tedious and time consuming. Furthermore, because consumers may use different terminologies to represent the same features, the simple keyword match won't be effective. Therefore, developing effective methods for feature extraction from online consumer reviews is essential and important, which has been well recognized in the literature [Liu et al. 2005; Somprasertsri and Lalitrojwong, 2010; Li et al. 2012].

This study proposes an integrative approach that combines LDA *(Latent Dirichlet Allocation)* and a synonym lexicon for extracting product features from Chinese online consumer product reviews. Our method identifies nouns and noun phrases in consumer reviews and creates a candidate feature set with LDA, extended by a synonym lexicon. After applying some filtering rules, the extended candidate feature set will be refined to generate the final feature set. Because the LDA model is a generic probability model, the proposed method is domain and product independent and can handle a large volume of reviews very well.

This study makes several research contributions. First, we use LDA in a different way in comparison to previous studies. Instead of using it to find topics described by bags of words, we use LDA to find features discussed in online consumer reviews. Second, consumers often use different terms to represent the same feature. Those terms may not appear in the product description provided by manufacturers. As a result, feature extraction methods that rely on the feature list in product description will result in low recall rates of infrequent feature words or phrases. To address this limitation, we couple LDA with a synonym lexicon to make the proposed approach capable of extracting product features that do not exist in product description. Third, we focused on Chinese online consumer reviews, to which the LDA model has never been applied. This research makes an initial yet crucial step toward building personalized online consumer product review systems that can enable filtering and summarization of reviews that address specific product features. Such systems can alleviate the information overload problem and improve consumers' purchase decisions.

The rest of the paper is organized as follows. We begin by discussing related work in Section 2. Then, the proposed method is introduced in Section 3, followed by the description of empirical evaluation and results in

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

2. Related Work

There are several limitations of existing methods for product feature extraction. First, many of them depend on manual intervention. For example, most prior studies needed experts to verify product feature nouns or noun phrases or feature expressions extracted, which requires significant time and manpower [Yi et al. 2003; Carenini et al. 2005; Kobayashi et al. 2005; Shi and Chang 2006; Zhuang et al. 2006]. Second, those feature extraction methods are usually only applicable to a particular product [Hu and Liu 2004b; Popescu and Etzioni 2007; Wang et al. 2010]. In other words, they are product-dependent. We summarize the related work in the rest of this section.

2.1. Terminology Identification

Existing approaches to terminology identification can be broadly classified into two categories, namely symbolic methods and statistical methods. Symbolic methods rely on syntactic description of terms (i.e., noun phrases). They were developed in early studies and needed experts to intervene and identify product features appeared in product reviews. For example, movie features are often composed of movie elements (e.g., plot, music, color) and movie related personnel (e.g., director, actor, and writer). Zhuang et al. [2006] decomposed a movie review and then performed Part-of-Speech (PoS) analysis to assign syntactic labels to individual words. Next, they extracted a set of nouns and noun phrases and found out the overlap between those noun/noun phrases and a set of standard feature words that they identified in advance. The overlapped terms were considered as movie features. Carenini et al. [2005] proposed an approach for organizing extracted features in a hierarchy. It incorporated users' prior knowledge about features and product information in a user-defined taxonomy of features. Then, it applied similarity matching techniques together with WordNet [Miller 1995], an online lexical database, to map learned features into the user-defined taxonomy. As shown by the above two examples, symbolic methods often need human involvement and domain knowledge in advance. In addition, they have difficulty in handling new features of a product.

Statistical methods use statistic models and data mining algorithms to identify terms of which occurrence frequencies are above a certain threshold. For example, Popescu and Etzioni [2007] proposed an OPINE framework to identify words or phrases within text reviews. The process first parses reviews using MINIPAR and applies a simple pronoun-resolution module to the parsed reviews to extract noun phrases and retain them with a frequency greater than an experimentally pre-defined threshold. Next, the OPINE's Feature Assessor evaluates noun phrases by computing the PMI (Point-wise Mutual Information) scores in order to select highly ranked noun phrases. The selected phrases will be sorted to create a product feature set. Hu and Liu [2004b] proposed to use association rule mining to identify feature words or phrases in product reviews. First, NLProcessor, a linguistic parser, parses each review by segmenting a review into sentences. Then Part-of-Speech analysis is performed to annotate words syntactically. Next, their method applies association rule mining to find frequent features (i.e., a set of nouns and noun phrases that frequently occur together in review sentences). Wang et al. [2010] proposed a boot strapping algorithm (i.e., aspect segmentation algorithm) to discover words or phrases from review text. The algorithm segments a review into sentences and determines initial aspect (i.e., feature) annotations by assigning each sentence to an aspect that shares maximum term overlapping. The aspect dependency is calculated using Chi-square statistics. Highly dependent words will be put under the corresponding aspect [Wang et al. 2010].

Although statistic methods can address the limitations of symbolic methods, they have their own inherent limitations. First of all, they put several constraints, such as compactness pruning and redundancy pruning [Hu and Liu 2004a], on using highly frequent noun phrases to identify product features. With such constraints, algorithms not only become more complex, but also produce many non-features. Furthermore, those statistical approaches require manual tuning of various parameters, which makes it difficult to apply those approaches to other types of text reviews.

2.2. Topic Model Based Summarization

In order to handle large text size for feature extraction, some topic model based summarization techniques have been developed. They identify latent topics and extract inherent aspects from a large collection of online consumer reviews. Topic models are a part of probabilistic modelling that calculates a joint probability distribution over both observed and hidden random variables.

An initial standard topic model is the Probabilistic Latent Semantic Analysis (PLSA) [Hofmann 1999], a probabilistic variant of Latent Semantic Analysis (LSA). PLSA assumes that a document is generated by a mixture of K topics, and the mixture coefficients are chosen individually for each document. However, a PLSA model does not make any assumptions about how the mixture weights are generated, making it difficult to test the generalizability of the model for new documents. In addition, the model used in PLSA has severe overfitting problems as the number of parameters grows with the size of a text corpus [Titov and McDonald 2008].

Latent Dirichlet Allocation (LDA) is one of the most commonly used topic models in recent years. In LDA, each document may be viewed as a mixture of various topics. This is similar to PLSA, except that in LDA, the topic distribution is assumed to have Dirichlet prior (i.e., a family of continuous multivariate probability distributions), which results in more reasonable mixtures of topics in a document [Blei et al. 2003]. Because LDA is a generic model and the distribution of its latent topics has Dirichlet prior, it addresses the overfitting problem of PLSA. To our best knowledge, LDA has not been used in feature extraction from Chinese online consumer product reviews. The detail of LDA will be introduced in Section 3.2.

2.3. Feature-Based Consumer Review Analysis

There has been an increasing amount of research on analyzing consumer product reviews and identifying consumer opinions on specific product features. Those studies used different methods for feature extraction. For example, Liu et al. [2005] emphasized the importance of feature-based consumer review analysis and visualization. They proposed an analysis system with a visual component to compare consumer opinions on products in different feature dimensions, such as picture, battery, zoom, size, and weight of digital cameras. They used association rule mining to identify frequent words or phrases as feature terms. Therefore, we used association rule mining as the benchmark in the evaluation of the proposed LDA-based approach in this study.

Aciar et al. [2007] proposed a recommender system based on user reviews of digital cameras. In their system, product features and customer expertise information were extracted by mapping user text reviews to a manually defined ontology. The problem with using an ontology in feature extraction is that building a quality ontology is non-trivial. Besides, different products may have different feature ontologies, making such an ontology-based approach to be product dependent and inflexible.

Somprasertsri and Lalitrojwong [2010] mined consumers’ opinions on product features using dependency analysis. Dependency grammars represent sentence structures by a set of dependency relationships. A dependency relationship is an asymmetric binary relationship between a word called head or governor and another word called modifier or dependent.

Major product features extracted from consumer reviews may let product manufacturers and retailers understand what features are mostly cared or concerned by consumers, and also help potential consumers make purchase decisions. Li et al. [2012] proposed a linear regression based approach for ranking product features according to their importance. They used a rule-based feature extraction method called double propagation, which consisted of eight rules. However, creating effective rules for feature extraction could be challenging.

3. An Integrative Approach with LDA and a synonym lexicon

Considering the limitations of existing feature extraction methods, in this research, we propose an integrative approach to feature extraction from online consumer product reviews by integrating LDA and a synonym lexicon. Figure 1 shows an overview of the proposed approach. The inputs are text reviews of a specific product, and the output is a collection of product features. In particularly, the proposed approach is aimed to extract product features on which consumers explicitly commented in their online reviews. For example:

*“The **pictures** are very clear.”*

*“The **resolution** of this camera is very good.”*

The above two review sentences include two product features of a digital camera, namely picture and resolution.

According to Li et al. [2009], Chinese consumer product reviews differ from English reviews in several aspects, causing the feature extraction and review analysis methods developed for the latter not directly applicable to the former. Those differences include 1) the variations in how opinions are expressed due to cultural differences; 2) the difference in language structure. For example, two adjacent words in an English sentence are separated by a blank space, but there is no blank space within Chinese sentences. Individual words have to be identified through word segmentation first; 3) the difference in grammars. Chinese grammar is more complex, causing it more difficult to process and annotate Chinese sentences. Also, we compared reviews at Amazon.com and at Taobao.com (A Chinese C2C website) and found that in general, the language used in Chinese consumer reviews is more casual and diversified than English reviews at Amazon.com. Some terms used in Chinese reviews even do not even exist in Chinese dictionaries (e.g., “给力”).

Although we focus on extracting product features from Chinese online consumer reviews in this study, the overall proposed methodology (i.e., combination of LDA and a synonym lexicon) is generic and should be applicable to feature extraction from online consumer reviews in other languages.

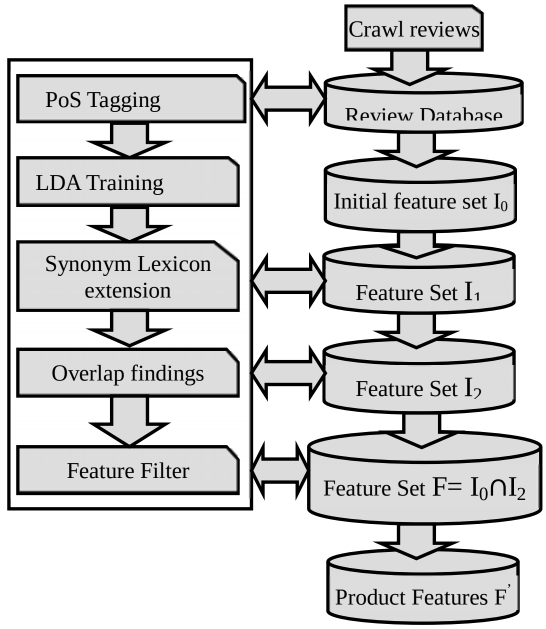

Figure 1: The Overview of the Proposed Approach to Feature Extraction

3.1. Part-of-Speech (PoS) Tagging

First of all, the proposed approach performs PoS tagging on individual Chinese consumer reviews. We use a Chinese lexical analysis system called ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System, http://www.ict.ac.cn/), a software package that parses each sentence in a review and generates a PoS tag (e.g., noun, verb, adverb, and adjective) for each word. The following example shows a parsed review sentence with PoS tags produced by ICTCLAS:

(Chinese)质量/n 没/v 的/ude1 说/v ，/wd 功能/n 强大/a。

After each review is parsed, the system extracts nouns and noun phrases and form an initial feature set $I_{0,}$ which will be used for training an LDA model.

3.2. The LDA Model

LDA posits that each document is a mixture of a small number of topics and that each word is attributable to one of the document's topics. The LDA process consists of two steps, namely generative process and inference [Blei et al. 2003].

The generative process starts by sampling a word distribution from a prior Dirichlet distribution for each latent aspect (i.e., feature). The generative process for each document i (i= 1, 2, … N) in a corpus D (a collection of documents) is:

(1) Choose $\theta_i \sim \mathrm{Dir}(\alpha)$, where θ_i is topic distribution of a document i and Dir (α) is the Dirichlet distribution of the parameter α.

(2) Choose $\varphi_k \sim \mathrm{Dir}(\beta)$, where φ_k is word distribution of a topic k.

(3) For each word w_{ij}:

3a. Choose a topic $z_{ij} \sim$ Multinomial (θ_i), where z_{ij} is the topic of the j_{th} word in the document i.

3b. Choose a specific word $w_{ij} \sim$ Multinomial ($\varphi_{z_{i,j}}$), where w_{ij} is a specific word (Observable variable).

Where α is the parameter of the Dirichlet prior on per-document topic distributions; β is the parameter of the Dirichlet prior on per-topic word distributions; θ_i is the topic distribution for review document i; φ_k is the word distribution for topic k.

Inference is performed to learn topic distributions per document, associated word probability distribution, and the particular topic mixture of each document distribution. It uses Gibbs Sampling for posterior inference of the parameters (α, β). The equation for Gibbs Sampling [Griffiths and Steyvers 2004] is given below:

$$p(z_i = k | w_i = v, z_{-i}, w_{-i}) \propto \left(\frac{n_{dk}^{DT} + \alpha}{\sum_{k'} n_{dk}^{DT} + K\alpha}\right)\left(\frac{n_{vk}^{WT} + \beta}{\sum_{v'} n_{vk}^{WT} + V\beta}\right) \quad (1)$$

where $z_i = k$ represents the assignment of the i^{th} term in a document to topic k, $w_i = v$ represents that an observed term w_i is the v^{th} term in the vocabulary of the text corpus, z_{-i} represents all the topic assignments excluding the i^{th} term, and w_{-i} represents all the terms in the vocabulary excluding the v^{th} term; n_{dk}^{DT} is the number of times that topic k occurs in document d, and n_{vk}^{WT} is the number of times that term v is assigned to topic k. Furthermore, K is the total number of topics; V is the size of the vocabulary; α and β are hyper-parameters for the document-topic and topic-word Dirichlet distributions, respectively.

The Gibbs sampling requires specification of values for the parameters. As we know, the content of online product reviews is often relevant to finite product features, which are the same as latent topics in an LDA model. In other words, we can view the distribution of product features or aspects as the distribution of topics in an LDA model. If the number of product features is K, then the probability of the i_{th} word or phrase w_i can be expressed by the following formula:

$$P(w_i) = \sum_{j=1}^{K} P(w_i|z=j)P(z=j) \qquad (2)$$

Where z represents a latent topic and $P(z=j)$ is the probability of $z = j$.

According to LDA, the distribution of words or phrases differs for different topics and the probability of words or phrases tend to be the same or close if they refer to the same features or topics. Take a digital camera as an example. Its product features include weight, quality, sharpness, flashlight, and picture, etc. Reviews may comment on any of these features. The distribution of words or phrases for a certain topic can be expressed by formula (3):

$$P(w|z=j) = \{p_{w_{j1}}, p_{w_{j2}}, p_{w_{j3}}, \cdots, p_{w_{jV}}\} \qquad (3)$$

Where $p_{w_{ji}}, i = 1, 2, 3, \cdots, V$ stands for the probability of word w_i under topic j. Then, we can get a feature set I_1 through the following rule:

$$I_1 = \left\{w_{ji} \middle| p_{w_{ji}} \geq \theta\right\},\ j = 1, 2, 3, \cdots, K; i = 1, 2, 3, \cdots, V \qquad (4)$$

Here θ is a parameter identified by experiments. Because the number of topics K is unknown, the optimal number of topics will be determined in a data-driven way. The suggested values of α and β are 50 and 0.1, respectively [Griffiths and Steyvers, 2004].

There have been some studies on variations of LDA. For example, Wallach [2006] developed a topic model that relaxed the bag of words assumption by assuming that topics generate words conditionally upon the previous word. Blei & Lafferty [2006] proposed a dynamic topic model that takes the ordering of documents into account. Titov and McDonald [2008] extended the basic LDA model by representing a document as a sliding window. Jo and Oh [2011] proposed a Sentence LDA (SLDA), in which they imposed an assumption that all words in a sentence are generated from one topic.

3.3 Synonym lexicon Extension

In this study, we integrate a synonym lexicon created by Mei et al.[1983] with LDA to expand the original feature set I_1 to set I_2 by adding the synonyms of each term in I_1, which is shown in Figure 2.

After extending I_1 to I_2, we generate a new feature set F, which is the intersection of I_0 and I_2 (i.e., $I_0 \cap I_2$). As suggested by Li et al. [2009], the words or phrases in the feature set F will be filtered by removing non-product feature terms, including product name, time, location terms, etc. At the end, the final product feature set F' will be generated.

```
Procedure FeatureWordExtendtion( )
begin
  for each word w_i in I_1
        for each syn or near-syn of w_i in Synonym lexicon
        put syn_i into D if it doesn't appear in D
  set I_2 = I_1 ∪ D
end
```

Figure 2: The Extension Process with a Synonym Lexicon

4. Evaluation

We have conducted an empirical evaluation of the proposed approach using more than 17,000 Chinese consumer reviews on three different digital camera products and one cell phone generated at http://www.jd.com between June 1 and August 1, 2012. Digital cameras are the most commonly studied product in previous studies on online consumer reviews [e.g., Liu et al., 2005; Aciar et al., 2007]. To assess the generalizability and robustness of the proposed approach across different types of products, we also selected Samsung Galaxy SII phone, which had fewer reviews but much more product features than the selected cameras. The descriptive information about those reviews is listed in Table 1.

Table 1: Data Description

Product Name	Number of Selected Reviews	Total Number of Reviews Available	Number of Manually Extracted Features
Nikon D3100	5,000	7,676	55
Canon 550D	5,000	5,557	60
Nikon D90	5,000	6,984	57
Samsung SII	2,049	2,049	103

The 5,000 reviews selected for each camera product were randomly chosen from all of the reviews generated for the corresponding product during the specified time span. Because the number of reviews available for Samsung Galaxy SII was much smaller, we used all the available reviews. All selected reviews satisfied two conditions: 1) the length of a review must be at least 10 words; and 2) each review must contain more than one noun or noun phrase. Three graduate students were asked to identify product features discussed in those reviews individually. For those disagreed features, they discussed them to see if a consensus can be reached. Eventually, only the features that all three reviewers had agreed upon at the end formed the correct benchmark product feature sets, which were then compared against the feature sets automatically extracted from Chinese consumer reviews by the proposed approach (LDA_SYN) and by the association rule mining (i.e., Apriori) approach. The numbers of finally agreed features of each product are shown in the last column of Table 1. We selected the association rule mining approach as the benchmark because it is about discovering 'frequent item sets' and has been used in prior studies on feature extraction [e.g., Hu and Liu, 2004b, Liu et al., 2005, Li et al., 2009].

We used precision (P), recall (R), and F-measure (F) as metrics to assess the extracted features of each product. Specifically, precision is the fraction of extracted features that are correct; recall is the fraction of correct features that are extracted among all correct features; F-measure is calculated as $2\frac{PR}{P+R}$. We randomly divided reviews into 30 groups. Then, precision, recall, and F measure of each group were calculated for either approach (i.e., association rule mining (ARM) versus LDA_SYN), which were then compared against each other. We further conducted paired T-tests to examine whether the improvements of the proposed approach over the ARM method were statistically significant. Means of precision, recall, and F-measures and T-test (df = 29) results are shown in Tables 2~4, respectively.

Results reveal that the proposed approach significantly outperforms the ARM in all three measures across all four products at 0.01 level, except that the improvement of precision for Samsung SII was at 0.05 significance level (p = .018). The results consistently demonstrate the superiority of the proposed approach to association rule mining in extracting product features from online Chinese consumer reviews.

Table 2: Precision of Association Rule Mining Versus Precision of LDA_SYN (%)

	P1 (ARM)		P2 (LDA_SYN)		P1-P2 (MD)	t	Sig.
	Mean	SD	Mean	SD			
Nikon D3100	53.45	5.79	64.69	4.20	-11.24	-9.3061	.000
Canon 550D	59.81	14.39	69.74	4.46	-9.93	-3.5340	.001
Nikon D90	61.81	5.21	68.92	4.19	-7.11	-5.5344	.000
Samsung SII	70.44	9.31	75.17	6.33	-4.73	-2.5169	.018

SD: Standard Deviation, **MD**: Mean Difference

Table 3: Recall of Association Rule Mining Versus Recall of LDA_SYN (%)

	R1 (ARM)		**R2 (LDA_SYN)**		**R1-R2 (MD)**	**t**	**Sig.**
	Mean	**SD**	**Mean**	**SD**			
Nikon D3100	50.84	7.90	64.30	5.44	-13.46	-8.1746	.000
Canon 550D	42.93	13.25	69.48	4.42	-26.55	-10.1014	.000
Nikon D90	55.39	8.46	68.77	4.25	-13.38	-7.8545	.000
Samsung SII	56.54	13.02	73.25	9.33	-16.71	-6.5515	.000

Table 4: F-Measures of Association Rule Mining Versus F-Measures of LDA_SYN (%)

	F1 (ARM)		**F2 (LDA_SYN)**		**F1-F2 (MD)**	**t**	**Sig.**
	Mean	**SD**	**Mean**	**SD**			
Nikon D3100	51.89	6.01	64.45	4.54	-12.55	-9.5936	.000
Canon 550D	48.05	8.39	69.55	3.87	-21.49	-11.9591	.000
Nikon D90	58.10	5.78	68.80	3.77	-10.70	-8.5274	.000
Samsung SII	61.96	9.66	73.94	6.65	-11.98	-6.5203	.000

Furthermore, to gain more insights, we did a T-test to compare the performance of the LDA-only based approach versus LDA_SYN (i.e., LDA plus the synonym lexicon) to examine the potential contribution of incorporating the synonym Lexicon. The T-test results of precision, recall, and F-measure (df = 29) are shown in Tables 5, 6, and 7 respectively.

Table 5: Precision of LDA only Versus Precision of LDA_SYN (%)

	P1 (LDA-only)		**P2 (LDA_SYN)**		**P1-P2 (MD)**	**t**	**Sig.**
	Mean	**SD**	**Mean**	**SD**			
Nikon D3100	66.71	4.51	64.69	4.20	2.02	8.6380	.000
Canon 550D	71.76	4.65	69.74	4.46	2.02	6.7994	.000
Nikon D90	71.00	4.39	68.92	4.19	2.08	9.0398	.000
Samsung SII	78.01	7.10	75.17	6.33	2.85	6.2605	.000

Table 6: Recall of LDA only Versus Recall of LDA_SYN (%)

	R1 (LDA-only)		**R2 (LDA_SYN)**		**R1-R2 (MD)**	**t**	**Sig.**
	Mean	**SD**	**Mean**	**SD**			
Nikon D3100	58.84	5.01	64.30	5.44	-5.45	-14.7795	.000
Canon 550D	64.74	4.07	69.48	4.42	-4.74	-11.4807	.000
Nikon D90	63.53	4.02	68.77	4.25	-5.23	-13.3728	.000
Samsung SII	67.94	9.87	73.25	9.33	-5.31	-8.1361	.000

Table 7: F-Measures of LDA only Versus F-Measures of LDA_SYN (%)

	F1 (LDA-only)		**F2 (LDA_SYN)**		**F1-F2 (MD)**	**t**	**Sig.**
	Mean	**SD**	**Mean**	**SD**			
Nikon D3100	62.49	4.56	64.45	4.54	-1.96	-8.5886	.000
Canon 550D	68.02	3.94	69.55	3.87	-1.52	-6.8607	.000
Nikon D90	67.02	3.83	68.80	3.77	-1.78	-10.2731	.000
Samsung SII	72.25	7.15	73.94	6.65	-1.69	-5.1966	.000

The results reveal that the LDA-only based approach to feature extraction resulted in higher precision (P<.01), but significantly lower recall (P<.01) and F measure values (Most at 0.01 level), than LDA_SYN, implying that adding a synonym lexicon improves the performance of feature extraction by increasing recall and overall F-measure. Such phenomena are consistently observed across all products.

5. Discussion

5.1. Major Findings

Online consumer product reviews are playing an increasingly important role in consumers' online purchase decisions. However, the rapid growth of the volume of reviews makes it challenging for consumers to browse reviews and identify information of interest. Due to the fact that different consumers may have different product feature preferences, feature-based review analysis and presentation would be very helpful. One of the essential steps toward achieving that goal is to extract product features from online consumer reviews automatically.

In this study, we propose a new feature extraction method based on the LDA model and a synonym lexicon to extract product features from online product reviews. The proposed method addresses the limitations of existing feature extraction approaches, namely the dependence on manual work and domain-specific nature. The results of empirical evaluation involving four products show that the proposed approach significantly outperforms the association rule mining method in precision, recall, and F-measure, and incorporating a synonym lexicon could help increase the recall and F-measure of feature extraction. The characteristics of LDA enable the proposed approach to be scalable and adaptive to large-scale reviews and different products. We speculate that the lower precision generated by the addition of a synonym lexicon in our study may be caused by the quality of the lexicon, which we would like to investigate in future research.

5.2. Practical Implications

In addition to research contributions discussed earlier, the findings of this research also provide some practical implications. For consumers, it would be beneficial to a consumer if an online consumer review system allows him to specify what specific features of a product are important to his purchase decision through a graphical user interface (GUI), and then automatically identifies, summarizes, and presents only those reviews that have discussed those features. By enabling such context-aware and personalized review selection and analysis, an online consumer product review system can significantly alleviate the information overload problem faced by consumers and help them make more informed decisions.

For most online retailers including Amazon.com, their current online review systems are a non-personalized, one-size-fits-all type of systems that cannot offer personalized review services to consumers. With the proposed feature identification and extraction method, online retailers can build and integrate feature-oriented review search and summarization functions into their review systems to provide more personalized and relevant support for individual consumers [Liang et al., 2012]. As a result, consumers' satisfaction and loyalty will be improved, which may lead to the increase of sales.

For product manufacturers, the extraction of product features makes identification of product defects much easier. A feature-oriented online consumer review system can enable categorization of consumer reviews into different groups based on features discussed in them, obtaining consumers' negative comments on specific product features through feature-based sentiment analysis, and synthesizing those negative comments to identify major defects and potential improvements of products.

5.3. Limitations and Future Research

There are several limitations of this study. First of all, like every other prior study in the literature, we analyzed consumer reviews of a few products only. E-Commerce research has categorized products into different types, such as perishable versus durable products, and search goods versus experience goods. Some products contain more complex or implicit features than others. Although we argue that the proposed approach is generic and product independent, differences in feature complexity and implicit features among different products may affect the performance of the proposed approach. Therefore, it would be helpful to apply it to different types of products in future research to examine if its superiority will remain.

We adopted one of the benchmarks used in previous studies, namely association rule mining, for evaluating the precision, recall, and F-measure of the proposed approach. As introduced in Section 2.3, there are other methods developed for feature extraction from online consumer reviews. To control the scope and complexity of this study, we could not include more methods as benchmarks. In the future, we plan to compare the feature extraction performance of the proposed approach with the performance of other methods.

Another limitation is that we focused on feature extraction from Chinese consumer reviews in this study. As discussed earlier in the paper, Chinese consumer reviews are different from consumer reviews in English. Also, Chinese online retailers or service providers adopted online consumer product review systems later than their western counterparts. Some unique mechanisms in Chinese review systems may influence the content of reviews. For example, at Taobao.com, a consumer (or an administrator) can change his/her previous reviews. It would be interesting to explore whether the proposed approach remains effective for consumer reviews in English (e.g., at Amazon.com).

In summary, exploring effective feature extraction methods for online consumer product reviews is an essential step for online consumer review systems to not only help consumers deal with the information overload problem, but also provide context-aware, personalized support for individual consumers, which aligns with the emerging trend of e-Commerce. This research is one of the initial efforts in achieving that goal and shows some promising results. With the ever increase of the volume of online consumer reviews, feature extraction will become increasingly important to fully taking advantage of such social media and improving consumers' decision making.

Acknowledgment

This research is supported by the National Natural Science Foundation of China (Award #: 71128003, 70972006, 71102111).

REFERENCES

Aciar, S., Zhang, D., Simoff, S., & Debenham, J. "Informed Recommender: Basing Recommendations on Consumer Product Reviews," *Recommender Systems,* Vol. 22, No. 3: 39-47, 2007

Ante, S. E. (2009). "Amazon: turning consumer opinions into gold". Bloomberg Business Week. Oct. 15, 2009. Available at http://www.businessweek.com/magazine/content/09_43/b4152047039565.htm.

Blei, M. and Lafferty, J. "Correlated topic models," In Proceedings of Advances in Neural Information Processing Systems.Vancouver, 147-155. British Columbia, Canada. December 5-8, 2006.

Blei, D. M., Ng, A. Y. and Jordan, M. I. "Latent dirichlet allocation," Journal of Machine Learning Research, 3: 993-1022, 2003.

Carenini, G., Ng, R. T. and Zwart, E. "Extracting knowledge from evaluative text," In Proceedings of the 3rd international conference on knowledge capture, 11-18. Alberta, Canada. October 02 - 05, 2005.

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

Griffiths, T. L. and Steyvers, M. "Finding scientific topics," In Proceedings of the National Academy of Sciences of the United States of America 101, 5228-5235, , Irvine, CA, 2004.

Hofmann, T. "Probabilistic latent semantic indexing," In Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval, 50-57, Berkeley, CA, USA. August 15-19, 1999.

Hu, M. and Liu, B. "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, 168-177, Seattle, WA, USA. August 22 - 25, 2004.

Hu, M. and Liu, B. "Mining opinion features in customer reviews," in Proceedings of the National Conference on Artificial Intelligence, 755-760, San Jose, California, USA. July25-29, 2004.

Hu, N., Zhang, J. and Pavlou, P. A. "Overcoming the J-shaped distribution of product reviews," Communications of the ACM, 52(10): 144-147, 2009.

iResearch. "The core data of e-commerce market in the third quarter of 2012," Available at http://www.iresearch.com.cn/coredata/2012q3_1.shtml. Accessed on Nov. 2012.

Jo, Y. and Oh, A. H. "Aspect and sentiment unification model for online review analysis," in Proceedings of the fourth ACM international conference on Web search and data mining, 815-824, Hong Kong, February 9-12,

2011.
Kobayashi, N., Inui, K., Matsumoto, Y., Tateishi, K., and Fukushima, T. "Collecting evaluative expressions for opinion extraction," Natural Language Processing–IJCNLP, 596-605, Jeju Island, Korea, October 11-13, 2005.
Li, S., Guan, Z., Tang, L., Chen, Z. (2012). "Exploiting consumer reviews for product feature ranking." Journal of Computer Science and Technology, 27(3): 635-649, 2012.
Li, S., Ye, Q., Li, Y., and Law, R. "Mining features of products from Chinese customer online reviews," Journal of Management Sciences in China, 12(2): 142-152, 2009.
Liang, T., Chen, H., Du, T., Turban, E., and Li, Y. "Effect of personalization on the perceived usefulness of online customer services: a dual-core theory," Journal of Electronic Commerce Research, 13(4): 275-288, 2012.
Liu, B., Hu, M., and Cheng, J. "Opinion observer: analyzing and comparing opinions on the Web," in Proceedings of the 14th international conference on World Wide Web, 342-351, Chiba, Japan. 2005.
Mei, J., Lan, Y. and Gao, Y. "Synonyms lexicon," 1st Edition. Shanghai Dictionary Publishing House, Shanghai, 1983.
Miller, G. A. "WordNet: a lexical database for English," Communications of the ACM, 38(11): 39-41, 1995.
Park, D. H. and Kim, S. "The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews," in Proceedings of Natural Language Processing and Text Mining, Springer London. 9-28, 2008.
Popescu, A. M. and Etzioni, O. "Extracting product features and opinions from reviews," in Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 339-346, Vancouver, British Columbia, Canada. Oct. 6-8, 2005.
Rosen-Zvi, M., Griffiths, T., Steyvers, M., and Smyth, P. "The author-topic model for authors and documents," in Proceedings of the 20th conference on uncertainty in artificial intelligence, 487-494, Arlington, Virginia, USA, 2004.
Shi, B. and Chang, K. "Mining Chinese reviews," in Proceedings of the Sixth IEEE International Conference on Data Mining, 585-589, Hong Kong, December 18-22, 2006..
Somprasertsri, G. and Lalitrojwong, P. "Mining feature-opinion in online customer reviews for opinion summarization," Journal of Universal Computer Science, 16(6): 938-955, 2010.
Titov, I. and McDonald, R. "Modelling online reviews with multi-grain topic models," in Proceedings of the 17th international conference on World Wide Web, 111-120, Beijing, China. April 21-25, 2008.
Wallach, H. M. "Topic modelling: beyond bag-of-words," in Proceedings of the 23rd International Conference on Machine Learning, 977-984, Pittsburgh, Pennsylvania, USA. June 25-29, 2006.
Xi, Y. H., Zhang M., Yuan F., Wang Y. "A survey of product reviews mining," Journal of Shandong University (Natural Science), 46(5): 16-24, 2011.
Wang, H., Y. Lu and C. Zhai. "Latent aspect rating analysis on review text data: a rating regression approach," in Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 783-792, Washington DC, USA. July 25-28, 2010.
Yi, J., Nasukawa, T., Bunescu, R. and Niblack, W. "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques," in Proceedings of the Third IEEE International Conference on ICDM, 427-434, Melbourne, Florida, USA. Nov. 19-22, 2003.
Zhuang, L., Jing, F., Zhu, X., and Zhang, L. "Movie review mining and summarization," in Proceedings of the 15th ACM International Conference on Information and Knowledge Management, 43-50, Arlington, VA, USA. November 5-11, 2006.