A Domain Oriented LDA Model for Mining Product Defects from Online Customer Reviews

Zhilei Qiao
Virginia Tech
qzhilei@vt.edu

Xuan Zhang
Virginia Tech
xuancs@vt.edu

Mi Zhou
Virginia Tech
mizhou@vt.edu

Alan Wang
Virginia Tech
alanwang@vt.edu

Weiguo Fan
Virginia Tech
wfan@vt.edu

Abstract

Online reviews provide important demand-side knowledge for product manufacturers to improve product quality. However, discovering and quantifying potential products' defects from large amounts of online reviews is a nontrivial task. In this paper, we propose a Latent Product Defect Mining model that identifies critical product defects. We define domain-oriented key attributes, such as components and keywords used to describe a defect, and build a novel LDA model to identify and acquire integral information about product defects. We conduct comprehensive evaluations including quantitative and qualitative evaluations to ensure the quality of discovered information. Experimental results show that the proposed model outperforms the standard LDA model, and could find more valuable information. Our research contributes to the extant product quality analytics literature and has significant managerial implications for researchers, policy makers, customers, and practitioners.

1. Introduction

Online reviews provide important demand-side knowledge from customers to improve product quality [1], [2]. Many companies seek to collect data on customer satisfactions and needs via survey, email, and work logs [3]. Motivated by customers' extrinsic benefits and intrinsic demands, they would also like to share their experience on products in their feedbacks, which are helpful to other consumers to make purchase decisions and product managers for improving their service and product quality. Customer feedbacks on the Internet reflect customer requirements that can be implemented in future product innovations and upgrades. As a matter of fact, online customer reviews can be and have been the significant driving force to the products' evolution [4]. As online reviews surge explosively and spread virally at an unprecedented speed, many firms seek to create business opportunities by discovering hidden value from the reviews [5–7]. The content of online reviews is mostly a type of unstructured data that is oftentimes difficulty to understand and manage. Therefore, effectively and efficiently extracting valuable knowledge in terms of product defects from unstructured data has wide utility but is very challenging. Indeed, driven by the economic benefits of new business opportunities, a new wave of commercial data analytics companies appears, such as CarComplaints.com, Topsy.com, and GNIP.

To overcome the challenge, existing studies have proposed different methods to tackle the problem. In 2010, Li et al. proposed a CRF-based review summarization approach to recognize product features and customers' opinions in customer reviews [8]. Utilizing an unsupervised learning approach, [9] offered a joint inference model, and [10] presented a graph co-ranking method to identify opinion targets and their relevant opinion words [9], [10]. Although these methods demonstrate their advantages in extracting product features and customers' assessments from individual product reviews, they still have difficulty in identifying common opinions from massive data. Particularly, to the best of our knowledge, there's no existing method that is able to obtain integral defect information such as: which component of the product has the defect and what are the descriptions of the defect.

The Latent Dirichlet Allocation (LDA) model proposed by [11] is an effective tool for text summarization. But the original LDA only gives general keywords for each topic given a document set. It doesn't guarantee obtaining complete information (e.g. flawed component, and corresponding description) if applied to defect identification. Without complete defect information is hard to understand and manage, and then has little value for policy makers or practitioners.

Table 1. Two-facet Topics for A Defect

Component topic	brakes, disc, pad

URI: http://hdl.handle.net/10125/41376
ISBN: 978-0-9981331-0-2

Description topic	brake, wear, pad, premature, squeak

Inspired by the LDA model, we are trying to address the challenge of product defect discovery via a novel method to automatically generate a defect summary of online customer reviews. Since different domain has different attributes and keywords to describe a defect. To define a domain-oriented defect, we extend the standard LDA model and show the two-facet LDA idea in Table 1. For each aspect of a defect, we have a corresponding topic to show its keywords, even phrases. We may get these two-facet topics given a number of product reviews like Table 2. This method adapts text mining techniques to extract integral and valuable demand-side knowledge from online customer reviews. The method develops a two-facet topic model that utilizes the interdependency relationship between two-facet information of a defect, and summarizes the enormous amount of user reviews. For example, a defect of an automobile usually contains facts about defective components and description. The summary in two-facet will be more readable and representative for the defect. Meanwhile, we apply Part-of-Speech (POS) tagging to removing noisy data from raw review contents. We showcase the method by analyzing online reviews collected from NHTSA (National Highway Traffic Safety Administration) user reviews database. Distinguishing from previous work, our study not only finds the reprehensive sentence [8] for a defect, but also provides more concrete information about a defect across different manufacturers. The goal of the research is to help product producers to extract actual and accurate product defects, such as: which component have a defect and what are the descriptions of the defect, from massive online reviews at a lower effort level compared to manual or supervised work of processing massive online reviews.

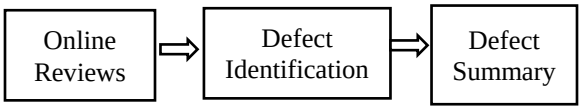

Figure 1. Overview of Our Method

Figure 1 provides an overview of our method to automatically retrieve topics for product defects. The method is able to extract major topics in online reviews and output defect summaries. From the example in Table 2, we can find the vehicle model, the flawed component, and the problem description. Another important observation in this research is that multiple defect sentences in a review represent the defect information. The sentence in regular font is the ownership sentence, which tells the vehicle model of this complaint corresponding to. The sentence in bold shows the details of the problem, which is more important. However, it's impossible to find the most complained defects of each product by just going through all the reviews manually, due to their huge amount.

Table 2. Data Example

Model	Accord 2008
Flawed Component	Service brakes / foundation components / disc / pads
Description	I am the original owner of this Honda Accord. **At 28,000 miles, the rear brake pads are prematurely worn out.**

The rest of paper is organized as follows. Section 2 reviews related work including the importance of conducting analysis on online customer reviews and related techniques. Section 3 presents the business value of product quality. The following section shows our probabilistic graphical method and presents our proposed method for the research problem. Section 5 reports our experimental evaluation. Section 6 discusses the results in relation to the existing literature and the contributions of the paper. The last section concludes the paper with a summary and discussion of possible future works.

2. Related Work

2.1 Business Value of Product Quality

This study is also related to research on the effects of product quality. As described by [3], the business value of product quality is associated with product competitive advantage, and then, product success and commercial success. Here, we focus on the effects of product quality issues from user reviews on customer relationship and defect management.

While firm managers have begun to pay attention to customer relationship management through social media, such as user reviews, very few studies focus on product quality management through analysis of unstructured text. More importantly, the quality issues reported in reviews are relatively credible in public communication channels [1], the word-of-mouth effects can quickly crash the product market [12]. In addition, with an exponential expansion in the number of reviews, it is hardly feasible to keeping pace to respond to individual user messages without the assistance of an automated tool. Fortunately, an automated product defect discovery method can

quickly and reliably identify specific product defects, and then practitioners can ensure quicker response to customer feedback. Consequently, fewer defective products will reach customers' hands and product quality issues can be controlled or addressed by firms in a timely manner; accordingly, firms can save costs for new products.

While firms are extrinsically motivated to design an automated product defect technique for customer relationship management, firms also are intrinsically motivated to develop such a technique for product quality management. Product quality is an important aspect of product competitiveness. Product defects damages product quality and the brand image. They are very costly to companies in some industries, e.g., particularly, in the automobile industry. If firms find defective units, they are mandated to report to the NHTSA and take timely remedy actions. For example, General Motors recently reported that the cost of repairing millions of vehicles reached $1.3 billion. In this case, the cost of the defect is huge, especially for a large volume of sales. Thus effective automated defect management techniques are crucial in helping firms find defects early and reduce the number of defective products, thereby reducing their potential financial loss.

2.2 Automatic Text Analytics

In the last decade, many studies have been working on sentiment analysis and opinion mining using massive text data [13]–[16]. In most of early research, the main goal of applying text analytics is to categorize a given text as positive and negative. Although distinguishing the sentiment of feedback can help customers make decisions, there is little value in assisting companies in making business decisions. To bridge this gap, recent aspect mining research focuses on extracting aspects (also called features, e.g. "restart button", "design", "ease of use", etc. for a mobile application) and estimating their ratings from feedback data [13]. While this type of research has provided more detailed information for customers and companies to make decisions, companies still need more integrated and actionable level of information, e.g., why customers dislike a specific aspect? How can companies improve that? While sentiment analysis can indicate whether customers are satisfied with certain products, companies still need more integral fine-grained information in improving the products. Integral information can be defined as the data that can be used to make feasible business decisions [16].

To extract valuable knowledge from online customer reviews, various data and text mining and information retrieval techniques have been proposed to collect and analyze online customer reviews. While many prior studies [17]–[20] are related to opinion mining, the research literature regarding the extraction of detailed information from feedback data to benefit the companies is emerging. In the following subsection, we review related studies to our work.

Latent aspect-based opinion mining conducts fine-grained analysis to discover sentimental ratings on aspects of items (e.g., "restart button", "resolution", "ease of use", etc. for a camera product) [13]. Most of early works on latent aspect-based opinion analysis are based on the frequency of noun phrases to identify aspects [16]. Later works are based on some dictionary methods or supervised learning techniques to learn aspects and their ratings [21]. However, most of current studies [22] are based on unsupervised topic models or LDA. Latent aspect-based opinion mining is mainly helpful for the customers to make decisions but not from product producers' perspective.

A few studies are closely related to the problem in our study. [23] use a deep semantic analysis and Natural Language Processing techniques to extract opinions and suggestions to improve a recommendation system. [24] extend the study and manually formulated sematic rules (e.g. "a manufacturer entity which is a subject of a modal verb used in the past tense and perfective aspect") to extract suggestions for product improvement from customer reviews [24]. Using the pattern and rules-based methods, Ramanand et al. [25] design a method to discover "wishes" sentences in which customers make suggestions (especially for improvements) about a product or a service. On the other hand, there are many studies in mining mobile apps' reviews. [26] analyze Apple's App Store reviews and extract customer requirements by adapting a topic modeling technique [26]. However, the paper intends to highlight the superiority of automatic method to extract customer requirements compared to manual efforts while maintaining the accuracy of extracted information. Moghaddam [16] proposes a semi-supervised method to extract actionable defect and improvement information from online customer reviews. Yet, this study still simply uses LDA based on different feature sets and fails to extract readable and integral information. The standard LDA can only give general keywords for each topic given a document set. It doesn't promise to acquire integral information (e.g. flawed component, and corresponding description) if applied to defect identification. Extending the standard LDA model, we may design a two-dimension LDA model in term of data. For each aspect of a defect, we have a corresponding topic to show its keywords, even phrases. We may get these two- dimension topics given a number of product reviews like Table 2. In our

study, therefore, we propose a domain-oriented LDA model to summarize product defects from online customer reviews. The proposed latent product defect model overcomes the problems of unsupervised clustering by using many domain-specific attributes that contribute to defect identification.

3. A Domain-Oriented LDA Model for Mining Product Defects

3.1. Latent Dirichlet Allocation (LDA)

The well-known LDA model [27] shown in Figure 2 is a generative probabilistic model which is able to identify topics from documents in a corpus. Therefore, we take the original LDA as our baseline method. The basic idea is that each document can be represented as mixtures of latent topics, where a topic is characterized by a distribution over words [27].

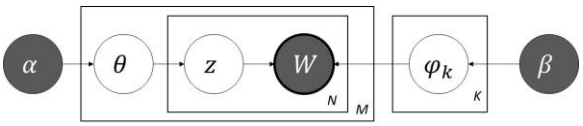

Figure 2. Graphical Model Representation of LDA

The document (review) generation process is shown in the algorithm below[28]. In the first step, a multinomial φ_k is drawn from the Dirichlet β as the word distribution for each topic *k*. Then, for each of the *M* documents, a multinomial θ is chosen. Here θ determines the probability of each topic given the document. After that, repeatedly sample the *N* words of this document. Specifically, a topic *z*, a latent variable, is sampled for each word *W* from θ. Finally, the word *W* is sampled from the corresponding word distribution φ_z.

Document generative process of LDA:

1. For each topic $k \in [1, K]$, sample a $\varphi_k \sim Dir(\beta)$, as its word distribution.
2. For each document $m \in [1, M]$:
    a. Sample $\theta \sim Dir(\alpha)$, as its topic distribution .
    b. For each of *N* words *W* in document *m*:
        i. Choose a topic $z \sim Multinomial(\theta)$.
        ii. Choose a word $W \sim Multinomial(\varphi_z)$

Two important assumptions for LDA are made. First, the dimensionality *K* of the Dirichlet distribution is known and fixed. Second, the word probabilities of topic *k* are parameterized by a multinomial distribution φ_z, which is sampled from the Dirichlet distribution β.

Given the parameters α and β, the joint probability distribution of a topic mixture θ, a set of *K* topics *z*, and a set of *N* words *W* is given by:

$$p(\theta, z, W|\alpha, \beta) = p(\theta|\alpha) \prod_{i=1}^{N} p(z|\theta) p(W|z, \beta) \quad (1)$$

Inferred by the Gibbs sampling algorithm, the Gibbs updating rule for the topic assignment is:

$$p\left(z_i = k \middle| \overrightarrow{Z^{\neg i}}, w = v, \overrightarrow{W^{\neg i}}, \alpha, \beta\right) \propto \frac{n(k,v)+\beta_v-1}{\sum_{v'=1}^{V}\left(n(k,v')+\beta_{v'}\right)-1} \cdot \frac{n(d,k)+\alpha_k-1}{\sum_{k'=1}^{K}\left(n(d,k')+\alpha_{k'}\right)-1} \quad (2)$$

Therefore, using the expectation of the Dirichlet distribution, the word distribution of topic *k* is calculated as:

$$\varphi_{k,v} = \frac{n(k,v)+\beta_v}{\sum_{v'=1}^{V}\left(n(k,v')+\beta_{v'}\right)} \quad (3)$$

The LDA model uses the latent variable θ to overcome the disadvantages of using a large set of individual parameters which are linked to training documents [27]. Nevertheless, based on the basic LDA model, there is only one dimension to characterize the topic and thus multiple dimensions of one topic and dependencies among dimensions are left out. Our experiment illustrates the disadvantage that the standard LDA has limitations in capturing multiple dimensional information and inter-dependence among dimensions.

3.2. The Latent Product Defect Mining Model

In this research, we introduce the Latent Product Defect Mining model (LPDM) which models the dependency between the product components and the corresponding problem description. It is a domain-oriented LDA model for mining product defects. As described in Table 2, different components of a product can have different quality and consequently different defect descriptions. As mentioned earlier, component and defect description are the key entities in terms of a product defect.

To facilitate the analysis, some text processing steps are necessary. An entity identification step is done before running the LPDM model. The non-relevant sentences (mostly ownership sentences) in the review description are removed using an extended stop word list, and the component words are recognized using a component lexicon. In addition, Part-of-

Speech tagging was applied to words of the review description, leaving only nouns, verbs, and adjectives, which are assumed to be the most informative words for product defect. Based on the $N_c$component words and N_d description words, $N_c * N_d$ word pairs are created by joining them.

We present the LPDM model in Figure 3. This model overcome the LDA weaknesses by jointly modeling latent product components and corresponding descriptions. The LPDM model can be considered as a generative process that first generates a product component and subsequently generates its issue description. The LPDM model generates the

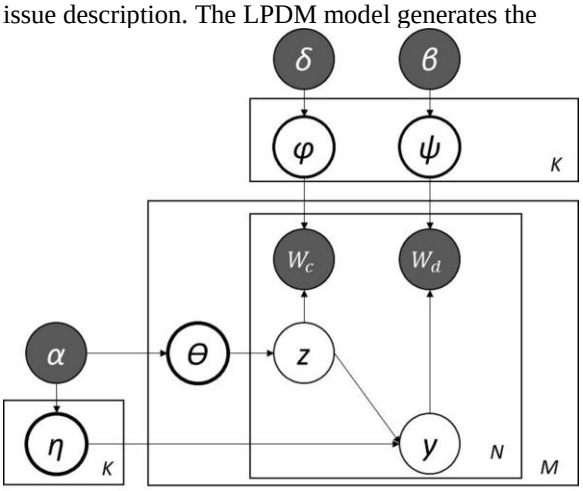

Figure 3. The LPDM Model of Reviews

word distributions for the component topics and description topics as the first step. Then a component topic distribution θ_m is sampled for each review *m*. For each word pair $<W_c, W_d>$ of review *m*, LPDM first generates a product component topic *z* from distribution θ_m. Next it draws a description topic y conditioned on the sampled component topic z. Here η is a k-dimension vector, therefore η_z is the *zth* row of a *k*k* topic dependence matrix. Finally, a product feature word W_c and a description word $W_d$are drawn based on φ_z and ψ_y , respectively. Formally, the LPDM model assumes the following generative process for a review:

1. For each component topic, sample a multinomial φ from $Dir\ (\varphi|\delta)$ as its word distribution; also sample a multinomial η from $Dir\ (\eta|\beta)$ as the word distribution of each description topic.
2. For each review *m*, sample $\theta_m \sim Dir(\theta_m|\alpha)$ as component topic distribution.
3. For each defect word pair $<W_c, W_d>$ of review *m*:
    a. Sample a component topic *z* from $multinomial(\theta_m)$.
    b. Sample a description topic ID *y* from multinomial $p(y|\eta_z)$.
    c. Sample $W_c \sim P(w_c|\varphi_z)$ and sample $W_d \sim P(W_d|\psi_y)$.

The collapsed Gibbs Sampling algorithm is used for the inference of this LPDM model. According to Figure 3, the joint distribution of the LPDM model is:

$$p(\vec{z}, \vec{y}, \vec{C}, \vec{D}|\vec{H}) = p(\vec{z}|\vec{\alpha}) \cdot p(\vec{C}|\vec{z}, \vec{\delta}) \cdot p(\vec{y}|\vec{z}, \vec{\alpha}) \cdot p(\vec{D}|\vec{y}, \vec{\beta}) \quad (4)$$

The Gibbs Updating Rule after derivation is:

$$P(z_i = k, y_i = l|\overrightarrow{z^{\neg i}}, \overrightarrow{y^{\neg i}}, \vec{C}, \vec{D}, \vec{H}) \propto \frac{n_{ct}(d,k)+\alpha_k-1}{\sum_{k'=1}^{K}(n_{ct}(d,k')+\alpha_{k'})-1} \cdot \frac{n_c(k,v_c)+\delta_{v_c}-1}{\sum_{v'=1}^{V_c}(n_c(k,v')+\delta_{v'})-1} \cdot \frac{n_{dt}(k,l)+\alpha_l-1}{\sum_{l'=1}^{K}(n_{dt}(k,l')+\alpha_{l'})-1} \cdot \frac{n_d(l,v_d)+\beta_{v_d}-1}{\sum_{v'=1}^{V_d}(n_d(l,v')+\beta_{v'})-1} \quad (5)$$

The word distribution for the *kth* component topic is:

$$\varphi_{k,v_c} = \frac{n_c(k,v_c)+\delta_{v_c}}{\sum_{v'=1}^{V_c}(n_c(k,v')+\delta_{v'})} \quad (6)$$

The word distribution for the *lth* description topic is:

$$\psi_{l,v_d} = \frac{n_d(l,v_d)+\beta_{v_d}}{\sum_{v'=1}^{V_d}(n_d(l,v')+\beta_{v'})} \quad (7)$$

In practice, the output words of component topics and description topics are decided by sorting the rows (vectors) in matrices φ and φ.

Table 3. Definition of Notations

$\vec{H}$	parameters of Dirichlet distributions, including $\alpha,\ \delta,\ \beta$
φ, ψ	parameters of multinomial distributions which denote the word distributions of component topics and description topics
$\Theta,\ \eta$	parameters of the multinomial distributions, which denote the component topic distribution over documents, and the description topic distribution over component topics respectively
z_i, y_i	component topic and description topic of the *ith* word pair in the corpus
$\overrightarrow{z^{\neg i}}, \overrightarrow{y^{\neg i}}$	topic assignment vector for all the word pairs in the corpus excluding the *ith* pair

$\vec{C}, \vec{D}$	component word vector and description vector for all the word pairs in the corpus (observations)
$n_c(k, v_c)$	the number of times of component word v_c assigned to component topic k
$n_{ct}(d, k)$	the number of words in document d assigned to topic k
$n_d(k, v_d)$	The number of times of description word v_d assigned to description topic k
$n_{dt}(k, l)$	the number of word pairs assigned with component topic k and description topic l
V_c, V_d	the size of component word vocabulary and description word vocabulary respectively
K	the number of topics
N	the number of word pairs in a review
M	the number of reviews in a corpus

Table 3 shows the notations used in LPDM.

The interdependency assumption of the LPDM model overcomes the lack of dependency correlation in the standard LDA model, which cannot distinguish component words and description words. The LPDM model captures the phenomenon that the flawed components determine the descriptions of their corresponding issues.

Representative complaints are useful to help people understand a defect. Compared to the component topics and description topics which are made of keywords, representative complaints are more straightforward. Since component topics are essential in this research, we retrieve representative complaints for each defect based on its component topic distribution. There're two steps for this retrieving process:

1) For each review m, decide the most relevant topic by searching for the maximum value in the vector θ_m
2) For a component topic i, pick the complaints which are marked as most relevant with i in step 1). Sort their component topic probability $\theta_{m,i}$ in descending order, then the top complaints in the list are the most representative complaints.

4. Experiments

Since none of the existing benchmark datasets for product defect study is publicly available, we had to create a new dataset. In this paper, we use the complaint database of NHTSA. We experimentally compared the two models discussed in this paper, i.e. LDA and LPDM. In the section, we first briefly describe our dataset and then present the evaluation of the proposed technique.

4.1 Datasets

We use the open database of NHTSA, which has 1.13 million vehicle complaints in total. The database includes complaints on various vehicle models. Each record has a number of attributes including a problem description. We take this field for our experiments. We choose a few vehicle models (e.g., CHEVROLET Cobalt 2006 and TOYOTA Camry 2007) with the largest number of complaints in the database for the experiments below. Both of the two vehicle models have around 2,400 complaints.

4.2 Evaluation

Both qualitative evaluation and quantitative evaluation have been done to ensure the quality of this research. First, we conduct a qualitative evaluation to extract the joint topics for critical product defects, which illustrates that the proposed method can capture the key defect information. Second, we evaluate the performance of the two models by measuring their topics with Precision-at-N (P@N) [29][30].

Qualitative Evaluation

Compared to the original LDA model, our proposed LPDM model can capture more coherent topics and accurate information. Table 4 and Table 5 show the joint topics generated by LPDM for the above 2 vehicle models. Looking at the component topic of the 1st defect of Cobalt 2006 in table 4, we can generally conclude this topic mainly about "*fuel system*" problem. The corresponding description topic further points out there was fuel leak caused by line crack which could be smelt. The 2nd defect of Cobalt 2006 is related with the "*power steering system*" according to its component topic. And its description topic further indicates the power steering failed and was difficult to steer the car. In contrast, the 1st defect of Camry 2007 is about the "*visibility system*" according to the component topic. Specifically, the sun visor on the driver side broke, which can be found in the description topic. The 2nd problem of Camry 2007 occurs to the "*accelerator*" as the component topic words include "*pedal*" and "*accelerator*". The corresponding description topic tells us that the

acceleration had some difficulty when the driver applied the accelerating pedal.

From these data samples, we can see the component topics reveal the defective components, while the description topics give more detailed information. Also these joint topics reflect good consistency and dependence.

In order to evaluate the accuracy of the proposed model and the standard LDA [7], we measure the P@N [30] of the key words extracted by them. P@N is to measure the precision of retrieved words in a defect gold standard set with N words [31].

The standard LDA is taken as the baseline method

Table 4. Joint Defect Topics for Cobalt 2006

Defect ID	Component Topic	Description Topic	Representative Reviews
1	fuel, pump, tank	fuel, leak, smell, odor, pump, gasoline, tank, strong, line, crack	1. Car began to leak fuel about one month after odor appeared, fuel leak was minimal at first, progressed to large puddle of fuel and frequency became consistent with a fuel leak. 2. Noticed a very strong fuel odor while driving, idling and parked.
2	power, steering, light	power, steering, drive, light, fail, warning, failure, difficult, lose, lock	1. Power steering failure while on way to work. Warning light came on prior to power steering failure. 2. Driving you lose power steering, just as before.
3	ignition, gear, shift	ignition, key, shift, gear, fail, park, lock, stick, lever, coil	1. The contact stated that the ignition key failed to release and the gear shift lever shifted out of park independently. 2. Couldn't get key out after starting, car would not go off. Shift lever would not go into park.

Table 5. Joint Defect Topics for Camry 2007

Defect ID	Component Topic	Description Topic	Representative Reviews
1	sun, position, view	visor, driver, sun, side, fall, obstruct, view, break, block, fracture,	1. The contact stated that the driver's side sun visor failed. the sun visor detached from the plastic holder and obstructed the driver's view of the roadway. 2. While driving at different speeds, the driver side sun visor hinge fractured and caused the sun visor to drop and obstruct the contacts view.
2	pedal, accelerator, gas	pedal, accelerate, acceleration, accelerator, problem, hesitate, hesitation, control, depress, slow	1. He stated the brake pedal modification prevented his foot from reaching the brake pedal in a timely manner. 2. The problem is the design of the brake pedal and accelerator pedal.
3	oil, engine, light	oil, engine, light, leak, burn, quart, pressure, warning, rate, low	1. Car lost oil pressure from broken VVTI oil line. Oil was all over engine and completely all over the bottom of car, driveway and garage. 2. My car is consuming oil out of normal.

We believe that allowing the component words in the description topics improves their readability. Although excluding component words would make the description topics neater, it will hurt the readability significantly. Many of the words are not specific (e.g. “problem”, “fail”, “break”, etc.). Therefore, they are not quite meaningful without context.

Quantitative Evaluation

in our experiment. Since the standard LDA doesn’t produce the 2-dimensional topics as LPDM does, we have to do some extension to obtain the “description key words”. In order to do that, we follow the baseline method mentioned in [29]. First, all the reviews are clustered by the standard LDA. Then, two experts are employed to map the clusters to the defects identified by LPDM. Finally, for each defect we rank all the

words (stop words excluded) by their frequencies within corresponding review cluster. In this way, we obtain the top "description words" using standard LDA.

In order to measure the P@N of key words extracted by various models, we need a gold standard set for each defect. It is created using the "Pooling strategy" in information retrieval [30]. The "description words" extracted by two methods were pooled together. Two experts are employed to go through these words, and manually determine which words are really relevant to the defect. In this way, a gold standard "description word" set is prepared for each defect. Then, we compare the top 5, 10, and 20 words generated by 2 methods to the gold standard set respectively to calculate the P@N rate.

Tables 6 and 7 show the P@N of the description key words extracted by LDA and LPDM on the two datasets. It can be seen that LPDM clearly outperforms LDA in terms of P@5, P@10, and P@20 and the improvement percentages in parentheses, although the P@N value of LPDM keeps decreasing along with the increase of N. That indicates LPDM will include more noise if we display a large number of words for each description topic. However, it's not quite serious since the top words in a topic are more important than the low-probability words.

Table 6. P@N of Description Words by LDA and LPDM on Cobalt 2006

Method	Description Key Words		
	P@5	P@10	P@20
LPDM	**80.00% (60.00%)**	**77.50% (29.17%)**	**67.50% (20.00%)**
LDA	50.00%	60.00%	56.25%

Table 7. P@N of Description Words by LDA and LPDM on Camry 2007

Method	Description Key Words		
	P@5	P@10	P@20
LPDM	**85.00% (70.00%)**	**80.00% (45.45%)**	**68.75% (17.21%)**
LDA	50.00%	55.00%	58.75%

***: Numbers in parentheses indicate the degree of improvement over the baseline method**

5. Discussion

In this paper, we aim to discover potential product defects from massive unstructured online review data. Our analysis is consistent with existing studies of product defect discovery in that extracting product defects from online reviews has important business value [3], [32], [33]. Our results also indicate that the product defect information is not one-dimensional but draws on values associated with domain-oriented attributes (e.g. product model, defective symptoms, and others).

Previous studies [2], [22], [17] present different kinds of method of automatic content analysis for extracting valuable knowledge from large amounts of unstructured data. For example, dictionary methods are using key words or terms to summarize documents [2]. They are one of the simplest and intuitive ways to automatically analyze textual data. Because of its ease to use, dictionary methods are commonly used for measuring texts in social science. Taking product defect disclosure text analysis for example, researchers [32] create a distinctive key word list named as "smoke words" which appear significantly more frequent in vehicle defects than other postings. Supervised learning methods [21] provide another method for summarizing documents to predefined categories. For example, some key terms, product features, and semantic factors can help identify product defects, but stylistic, social, and sentiment features cannot [3]. Still, the important assumption of supervised machine learning methods is to have a set of predefined categories (product defect types), which is very tedious and not very flexible from the automatic perspective. Distinguishing from Dictionary and supervised machine learning methods, unsupervised learning methods don't need to predefine categories and any tagging labels. The application of unsupervised learning methods to analyze text in social science is still in at its infancy. In information system literature, [35] use unsupervised topic models to cluster the content of recommendation articles. However, they did not include any context-oriented information and use the standard LDA model to solve the challenging problem of unsupervised learning methods.

Despite the applications of unsupervised learning methods in some research areas, there is limited studies that attempt to analyze online reviews for product defect discovery. Based on the prior supervised study [3], our work on product defect discovery incorporates contextual information to estimate topics. The proposed LPDM model overcomes the problems of unsupervised clustering by using many domain-specific attributes that contributes to defect identifications.

In summary, our analysis adds to the existing body of knowledge of product defect discovery by painting a more nuanced picture by representing each defect as a multi-dimensional concept. Researchers can use our study and our methodology as a source of aspiration when studying product defect identifications in online communities. Policy makers may draw on our evaluation results when discussing new safety laws that allow for further developing other

people's suggestions and intellectual property. Firms facing with hyper-competition, can find their product issues from customer feedbacks and learn from these feedbacks to further improve their product offerings. Hence, managers are urged to see the benefits in incorporating customers' feedbacks and transforming creative inputs into new business opportunities and practices. This is in line with existing literature involving users for value co-creation [36]. For example, [37]shows that semiconductor manufacturer's sales of user-designed chips had been up to $15 billion in 2000.

This study makes the contributions for defect discovery from online reviews. First, we propose a novel unsupervised Bayesian inference model, called latent product defect model (combined with the Gibbs Sampling algorithm), for identifying product defects and relevant details from unstructured textual data. To the best of our knowledge, this is the first method to incorporate interdepended relationships of different topics in an unsupervised learning method into the quality management field. Second, we conduct a comprehensive evaluation of our proposed LPDM model using both quantitative and qualitative evaluation methods. Experimental results show that our proposed model outperforms the competing LDA method and discovers more meaningful product defects. This model facilitates navigation of large amounts of textual data for our target users, including firm executives, product managers, or end users. Third, the proposed LPDM model (with the Gibbs Sampling algorithm) extends the automated defect discovery literature as well. It provides a comprehensive framework for incorporating contextual information (syntactic features and domain background) to uncover more meaningful topics. Last but not least, our study on defect identification from online reviews also contributes to the quality management literature. By responding to customer complaints for product quality issues, companies can efficiently limit the spread of defective products, as well as improve the quality of customer engagement.

Our paper also has several limitations. First, this research uses only one public data source and is constrained by the limitations of using only text analysis. An empirical study incorporating other sources of data from manufactures might yield more valuable and practical insights. Second, this study only studies product defect discovery problem of the automobile industry. Including product defect identification problems in other domains can present the generalizability of the novel model. In the future, we want to incorporate the other perspectives, e.g., product advantages and unmet user requirements, which can also help managers know customers' preference and demand to better position their products in the right customer segments. Although we evaluate the result based on the importance of key words, we do not investigate the topic coherence and readability. To test the robustness of our model, we plan to conduct sensitivity analysis on prior values and improve the readability of the identified defect information via sentences and n-grams in the future [16].

6. Conclusions

In this paper, we have profiled and described the product defect discovery using the open database of NHTSA, which has 1.13 million vehicle complaints on various vehicle models. We propose a novel unsupervised learning technique to automatically extract fine-grained information about defects from customer reviews. The method automatically identifies product defect information and its summary to assist companies to find more business opportunities and areas for future product improvements. The study adds to the existing body of knowledge of product defect discovery and confirms the efficiency and effectiveness of our proposed model.

7. Acknowledgements

This research was supported by both Center for Business Intelligence and Analytics (CIBA) at Virginia Tech and the Natural Science Foundation of China (Grant# 71531013).

References

[1] A. Abrahams, J. Jiao, G. Wang, and W. Fan, "Vehicle defect discovery from social media," *Decis. Support Syst.*, vol. 54, no. 1, pp. 87–97, 2012.

[2] X. Zhang, Z. Qiao, L. Tang, W. Fan, E. Fox, and G. A. Wang, "Identifying Product Defects from User Complaints Identifying Product Defects from User Complaints: A Probabilistic Defect Model," in *Proceedings of 22nd Americas Conference on Information Systems (AMCIS)*, 2016.

[3] A. S. Abrahams, W. Fan, G. A. Wang, Z. Zhang, and J. Jiao, "An Integrated Text Analytic Framework for Product Defect Discovery," *Prod. Oper. Manag.*, vol. 24.6, no. 6, pp. 975–990, 2015.

[4] D. Pagano and W. Maalej, "User Feedback in the Appstore: An Empirical Study," in *Requirements Engineering Conference (RE), 2013 21st IEEE International*, 2013, pp. 125–134.

[5] X. Luo, J. Zhang, and W. Duan, "Social media and firm equity value," *Inf. Syst. Res.*, vol. 24, no. 1, pp. 146–163, 2013.

[6] Y. Yu, W. Duan, and Q. Cao, "The impact of social and conventional media on firm equity

value: A sentiment analysis approach," *Decis. Support Syst.*, vol. 55, no. 4, pp. 919–926, 2013.
[7] W. Fan and M. Gordon, "The power of social media analytics," *Commun. ACM*, vol. 57, no. 6, pp. 74–81, 2014.
[8] P. Li, J. Jiang, and Y. Wang, "Generating templates of entity summaries with an entity-aspect model and pattern mining," *Proc. 48th Annu. Meet. Assoc. Comput. Linguist.*, 2010.
[9] B. Yang and C. Cardie, "Joint Inference for Fine-grained Opinion Extraction.," *ACL*, 2013.
[10] K. Liu, L. Xu, and J. Zhao, "Extracting Opinion Targets and Opinion Words from Online Reviews with Graph Co-ranking.," *ACL*, 2014.
[11] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
[12] Y. Chen and J. Xie, "Online Consumer Review: Word-of-mouth As A New Element of Marketing Communication Mix," *Manage. Sci.*, vol. 54, no. 3, pp. 477–491, 2008.
[13] H. Wang, 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*, 2010, pp. 783–792.
[14] B. Liu, M. Hu, and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web," in *Proceedings of the 14th international conference on World Wide Web*, 2005, pp. 342–351.
[15] S. Stieglitz and L. Dang-Xuan, "Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior," *J. Manag. Inf. Syst.*, vol. 29, no. 4, pp. 217–248, 2013.
[16] S. Moghaddam, "Beyond Sentiment Analysis: Mining Defects and Improvements from Customer Feedback," in *Advances in Information Retrieval*, Springer, 2015, pp. 400–410.
[17] A. Stavrianou and C. Brun, "Expert Recommendations Based on Opinion Mining of User- Generated Product Reviews," *Comput. Intell.*, vol. 31, no. 1, pp. 165–183, 2015.
[18] X. Ding, B. Liu, and P. Yu, "A Holistic Lexicon-based Approach to Opinion Mining," *Proc. 2008 Int. Conf. Web Search Data Min.*, 2008.
[19] J. Jin, P. Ji, and C. K. Kwong, "What Makes Consumers Unsatisfied with Your Products: Review Analysis at a Fine-grained Level," *Eng. Appl. Artif. Intell.*, May 2015.
[20] Z. Z. Z. Zhang, "Weighing Stars: Aggregating Online Product Reviews for Intelligent E-commerce Applications," *IEEE Intell. Syst.*, vol. 23, no. 5, 2008.
[21] A. S. Abrahams, W. Fan, G. A. Wang, Z. Zhang, and J. Jiao, "An integrated text analytic framework for product defect discovery," *Prod. Oper. Manag.*, vol. 24.6, no. 6, pp. 975–990, 2015.
[22] Y. Bao and A. Datta, "Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures," *Manage. Sci.*, vol. 60, no. 6, pp. 1371–1391, 2014.
[23] A. Stavrianou and C. Brun, "Opinion and Suggestion Analysis for Expert Recommendations," in *Proceedings of the Workshop on Semantic Analysis in Social Media*, 2012, pp. 61–69.
[24] C. Brun and C. Hagege, "Suggestion Mining: Detecting Suggestions for Improvement in Users' Comments," *Res. Comput. Sci.*, vol. 70, pp. 171–181, 2013.
[25] J. Ramanand, K. Bhavsar, and N. Pedanekar, "Wishful Thinking: Finding Suggestions and 'Buy' Wishes from Product reviews," in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 2010, pp. 54–61.
[26] L. V Galvis Carreño and K. Winbladh, "Analysis of User Comments: An Approach for Software Requirements Evolution," in *Proceedings of the 2013 International Conference on Software Engineering*, 2013, pp. 582–591.
[27] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
[28] G. Heinrich, "Parameter Estimation for Text Analysis," *Univ. Leipzig, Tech. Rep*, 2008.
[29] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid," no. October, pp. 56–65, 2010.
[30] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*, vol. 1, no. 1. Cambridge university press Cambridge, 2008.
[31] C. D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval," *J. Am. Soc. Inf. Sci. Technol.*, vol. 1, p. 496, 2008.
[32] A. Abrahams, J. Jiao, G. Wang, and W. Fan, "Vehicle Defect Discovery from Social Media," *Decis. Support Syst.*, vol. 54, no. 1, pp. 87–97, 2012.
[33] A. Abrahams, J. Jiao, and W. Fan, "What's Buzzing in the Blizzard of Buzz? Automotive Component Isolation in Social Media Postings," *Decis. Support Syst.*, vol. 55, no. 4, pp. 871–882, 2013.
[34] A. Abrahams, J. Jiao, and W. Fan, "What's buzzing in the blizzard of buzz? Automotive component isolation in social media postings," *Decis. Support Syst.*, vol. 55, no. 4, pp. 871–882, 2013.
[35] S. Aral and D. Walker, "Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks," *Manage. Sci.*, vol. 57, no. 9, pp. 1623–1639, 2011.
[36] V. Ramaswamy and C. K. Prahalad, "Co-creation Experiences: The Next ractice in Value Creation," *J. Interact. Mark.*, vol. 18, no. 3, pp. 5–14, 2004.
[37] E. Von Hippel, S. Thomke, and M. Sonnack, "Creating breakthroughs at 3M," *Harv. Bus. Rev.*, vol. 77, pp. 47–57, 1999.