Expert Systems With Applications 178 (2021) 115030

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

An integrated probabilistic graphic model and FMEA approach to identify product defects from social media data

Lu Zheng [1], Zhen He [*,2], Shuguang He

College of Management and Economics, Tianjin University, Tianjin 300072, China

A R T I C L E I N F O

Keywords:
Product defect discovery
Social media data
Probabilistic graphic model
FMEA
Text analysis

A B S T R A C T

Recently, the explosive increase in social media data enables manufacturers to collect product defect information promptly. Extant literature gathers defect information like defective components or defect symptoms without distinguishing defect-related (DR) texts from defect-unrelated (DUR) texts and thus makes defects discussed by few texts buried in enormous DUR texts. Moreover, existing studies do not consider the defect severity which is valuable and important for manufacturers to make remedial decisions. To bridge these research gaps, we propose a novel approach that integrates the probabilistic graphic model named Product Defect Identification and Analysis Model (PDIAM) with Failure Mode and Effect Analysis (FMEA) to derive product defect information from social media data. Comparing to extant studies, PDIAM identifies DR texts and then extracts defect information from these texts. And PDIAM provides more defect information than previous researches. Besides, we further analyze defect severity with the combination of FMEA and PDIAM which alleviates the inherent subjectivity brought by expert evaluation in the traditional FMEA. A case study in the automobile industry proves the predominant performance of our approach and great potential in defect management.

1. Introduction

Product defects are a widespread concern for manufacturers when conducting quality management and customer relationship management. When safety-related defects occur, manufacturers have to take remedial actions like product recalls and bear enormous economic costs. For those non-safety defects not leading to recalls, they can still reduce purchase intention and customer satisfaction, and cause within-warranty-repair cost increment (Alkahtani et al., 2018). And both safety and non-safety-related defects bring untold harm to the word-of-mouth of manufacturers (Rupp, 2004). Hence, manufacturers should identify product defects and take prompt remedial reactions before they widely spread.

However, for manufacturers, it is a tough and complex task to identify product defects especially in the stage of after-sales. Manufacturers may conduct strict tests to control product quality before products are available to customers, but customers use products under various conditions which are impossible to be considered thoroughly by manufacturers (Goldberg & Abrahams, 2018). This makes defect detection more difficult when products reach customers. To collect defect information and discover hidden defects, manufacturers have to utilize labor resources to dig up defect information and then transfer it into useful knowledge.

The traditional ways to collect defect information are to make questionnaires, review warranty claims or customer complaints from after-sale service centers. But this method has the deficiencies of hysteresis and incomprehensiveness (Liu et al., 2018). And the obtained information is generated in the warranty period. Defect information outside the warranty period will be omitted. Recently, social media have become popular platforms for customers to exchange their opinions freely. Therefore, data on social media are more trustworthy than the information collected in traditional ways (Farhadloo et al., 2016). For the advantages of availability and promptness, social media data have become an important information source for manufactures to gather defect information (Zheng et al., 2020). But how to collect defect information from the colossal volume of unstructured data puts a demanding challenge to manufacturers. To beat off this challenge, researchers have developed effective approaches. Part of the researchers

* Corresponding author.
E-mail address: zhhe@tju.edu.cn (Z. He).
[1] ORCID: 0000-0003-3939-3553.
[2] ORCID: 0000-0003-4149-1830.

https://doi.org/10.1016/j.eswa.2021.115030
Received 8 July 2020; Received in revised form 15 March 2021; Accepted 9 April 2021
Available online 16 April 2021
0957-4174/

view defect detection as text classification, they applied machine learning methods (Abrahams et al., 2013; Liu et al., 2018; Zhang et al., 2015) or a defect lexicon named "smoke words" to identify defect-related (DR) texts (For brevity, we use DR to denote defect-related and DUR to represent defect-unrelated) (Abrahams et al., 2012, 2015; Goldberg & Abrahams, 2018; Law et al., 2017; Winkler et al., 2016). But these researches only concentrate on DR text classification and cannot provide valuable defect information which is more meaningful for manufacturers.

Hence, to collect important defect information, some researchers utilize probabilistic graphic models (PGMs), a commonly used method to derive topics from corpora (Zhang et al., 2019, 2016). Through PGMs, defects discussed by customers can be derived from the texts. But the major deficiency of PGM-based studies is that these studies extract defect information without the discrimination of DR and DUR texts. Given that DUR texts are overwhelming in quantity, gathering defect information from both DR and DUR texts will make the defects discussed by few texts buried in DUR texts. And the extracted topics are not all relevant to product defects. Therefore, Gathering defect information from DUR texts will lead to biased and incomprehensive results. To enhance the accuracy of derived defect information, PGMs must be conducted on DR texts. Moreover, previous literature has not further estimated the defect severity which is valuable for manufacturers to make remedial decisions. In conclusion, the issues of the existing literature are listed as follows.

- Social media data become an important information source to identify product defects for manufacturers and researchers. But existing literature mainly focuses on DR text classification and overlooks the detailed defect information which provides valuable managerial insights to manufacturers.
- Extant studies derive defect information without the exclusion of DUR texts which influences the accuracy and comprehensiveness of defect information.
- Extant studies do not measure the severity of extracted defects.

Considering that PGMs have shown their effectiveness in deriving main topics from the large corpus in many studies (Chen et al., 2017; Zhang et al., 2019) and the value of social media data in defect management especially in the after-sale stage (Abrahams et al., 2012, 2015; Goldberg & Abrahams, 2018; Liu et al., 2018; Zheng et al., 2020), we collect defect information from based on social media data and PGMs. And the research problem in our study is how to develop a PGM that can identify defect-related texts, derive defect information from these texts, and measure the severity of defects. With the research problem, we propose a PGM named Product Defect Identification and Analysis Model (PDIAM) to discover DR texts and then collect defect information from these texts. Furthermore, based on PDIAM, we utilize FMEA method to evaluate the severity of identified defects. To summarize, our proposed method has the following three-fold contributions.

- Unlike the extant PGMs, PDIAM identifies DR texts and gathers defect information from these texts. The exclusion of DUR texts enhances the comprehensiveness and accuracy of derived defect information.
- Comparing with the state-of-the-art topic models, PDIAM shows its outperformance and helps manufacturers detect defects more accurately and automatically.
- Based on PDIAM, we use FMEA to further analyze the defect severity. The combination of FMEA and PDIAM not only alleviates the inherent subjectivity of traditional FMEA which is brought by expert evaluation but also provides more valuable defect information for manufacturers.

The remainder of this paper is organized as follows. In Section 2, we conduct a comprehensive literature review about probabilistic graphic models, social media data and their significant effects on product defect detection. In Section 3, we lay out the details of our approach. The case study is implemented to validate PDIAM's effectiveness and accuracy in Section 4. Section 5 concludes our study and provides an overview of its limitations and opportunities for future works.

2. Literature review

In this section, we review relevant literature about PGMs in text analysis, social media data, and product defect detection. Detailed discussions on the main points of prior work as well as limitations are elaborated after reviewing extant approaches.

2.1. Probabilistic graphic models in text analysis

Based on graphs, PGMs represent the conditional independence relationships among various variables with complex distributions (Pernkopf et al., 2014). PGMs have been widely used in many research areas such as text analysis, computer vision, signal processing, and artificial intelligence (Koller & Friedman, 2009). In the area of text analysis, the most typical PGMs are topic models that extract main topics from texts. The most famous topic model is Latent Dirichlet Allocation (LDA) proposed by Blei et al. (2003). LDA transforms a large corpus into several topics and related words. Using LDA, many researchers publish findings with significant contributions in the areas of service quality (Guo et al., 2017), competitive analytics (Wang & Xu, 2018) and online platform research (Xiang et al., 2017).

Although LDA has good performance in topic extraction, the omit of relationships among paragraphs or sentences makes it less effective when analyzing complex corpora. To enhance the performance of LDA, researchers improve LDA from various aspects like better aspect extraction (Shams & Baraani-Dastjerdi, 2017), aspect and sentiment estimation (Chenghua Lin & Yulan He, 2009; Jo & Oh, 2011) or the consideration of paragraph relationships (Du et al., 2012). Other researchers choose to develop new PGMs based on their research problems. Amplayo et al. argued that product descriptions were helpful when extracting topics. Thus they constructed two topic models which took a thorough consideration of both online reviews and product description (Amplayo et al., 2018). By utilizing structural data, Roberts et.al built the topic model named Structural Topic Model (STM) to process texts and structural data jointly (Roberts et al., 2014, 2016). Given the relationships between words, Wang et al. and Zuo et al. proposed different topic models respectively (Wang et al., 2016; Zuo et al., 2016). Zoghbi et al. wanted to gather linking information between different usages of the same language and developed a novel topic model (Zoghbi et al., 2016). Previous literature has proven that PGM is the modeling framework with outstanding performance in text analysis and can be used across different application areas.

2.2. Social media data and product defect detection

Since the worldwide surge of social media has completely changed the way people exchange opinions, data generated in social media have been recognized as an essential indicator of customer opinions on products. Accordingly, social media data have drawn great interests from both manufacturers and researchers and have been applied in various research areas including competitive analysis (Gao et al., 2018; Liu et al., 2019), product design (Jin et al., 2015; Wang et al., 2018; Zhang et al., 2018; W. Zhang et al., 2012) and customer satisfaction measurement (Bi et al., 2019; Farhadloo et al., 2016; Guo et al., 2017; James et al., 2017; Korfiatis et al., 2019).

In the area of product defects identification, social media data are also an important information source to uncover buried defects. According to the used methods, we divide these studies into the machine-learning based, the smoke word based and the PGM based studies. The machine-learning based literature uses machine-learning methods like

Table 1
Summary of literature on product defect detection based on social media data.

Researches	Approaches	DR text discovery	Defect information	Derive topics from DR texts	Defect severity
Lo (2008)	ML	√			
Zhang et al. (2015)	ML	√			
Liu et al. (2018)	ML + PGM	√	√	√	
Abrahams et al. (2012)	SW	√			
Winkler et al. (2016)	SW				
Goldberg and Abrahams (2018)	SW	√			
Law et al. (2017)	SW	√			
Chen et al. (2017)	PGM		√		
Zhang et al. (2016)	PGM		√		
Zhang et al. (2019)	PGM		√		
Zheng et al. (2020)	PGM	√	√	√	
This study	PGM + FMEA	√	√	√	√

ML: Machine learning methods. SW: Smoke words.

Support Vector Machine (Lo, 2008; Zhang et al., 2015) and multi-view ensemble learning (Liu et al., 2018) to classify texts into DR and DUR texts. Machine-learning methods have shown their performance in DR text classification. But they cannot provide detailed defect information. Besides machine-learning methods, "smoke words" is the most outstanding and contributive approach to discovering defects (Abrahams et al., 2012). "Smoke words" is a lexicon which consists of words most related to product defects. When constructing the lexicon, experts need to tag which texts discuss defects and then decide the smoke words using term-prevalence metrics (like Correlation Coefficient score (Winkler et al., 2016), Relevance Correlation Value score, Information Gain score, etc.) (Abrahams et al., 2012). Although smoke words have been proven effective in different industries (Abrahams et al., 2012,; Abrahams, Fan, Wang, Zhang, & Jiao, 2015; Winkler et al., 2016), the heavy dependence on expertise hinders their extensive application in practice. And how to curate the best and proper smoke word lexicon may be demanding for manufacturers (Goldberg & Abrahams, 2018).

Given that PGMs are useful tools to extract main topics from texts, they have shown their potential to discover defects in many studies. Famous PGMs like LDA or STM have been applied and obtain good performance (Chen et al., 2017; Kuhn, 2018). But these PGMs can only derive main topics from corpora, they can not provide information on product defects detaily. To deal with this problem, Zhang et al proposed a PGM with the consideration of product attributes and offered wealthy defect information on product models, years of production, detective components, and symptoms (Zhang et al., 2016). Zhang et al developed a PGM named Product Defect Latent Dirichlet Allocation (PDLDA) which takes further considerations of defect resolutions (Zhang et al., 2019). For the advantages of less reliance on labor investments, the provision of abundant information and the excellent performance in text analysis, PGMs are effective tools to reveal defects from social media data. But extant studies of defect discovery conduct PGMs without the discrimination of DR and DUR texts. They derive DR topics from DUR texts which may lead to biased results (Zheng et al., 2020). In our previous research, we developed a PGM to discover DR texts and collect defect information. To enhance the accuracy of identified results, we used three filters (including sentiment, similarity and componeng-symptom filters) to filter out DUR texts. But the construction of filters needs pre-training (Zheng et al., 2020). In contrast, using PGMs to identify DUR texts will be more convenient and require less labor investment. Hence, how to construct PGMs that can distinguish DR texts from DUR texts and then ingest defect information from DR texts still needs further research.

2.3. Summary

Table 1 summarizes previous works on product defect detection using social media data. Three research gaps can be observed from this table. The first gap is that most studies concentrate on DR text discovery and overlook the defect information derivation. The second gap is that extant PGM-based studies derive topics from both DR and DUR texts which will induce inaccurate results. The third gap is that few extant studies provide information on defect severity. Given these research gaps, we propose an integrated PDIAM and FMEA approach to identify DR texts and collect detailed defect information.

3. Research methodology

Selecting a proper modeling approach is critical to construct an expert system. Considering that our aim is to deal with the colossal and unstructured social media data, manual processing modeling approaches like questionnaires (Nord et al., 2016) are difficult to gather defect information promptly and effectively. For prediction, statistics methods and machine-learning methods are the most used modeling approaches and show good performance (Topîrceanu & Precup, 2020). DR texts may be predicted using statistics or machine-learning methods. But the insights buried in texts are hardly derived by these methods because social media data usually are in the format of unstructured texts. Hence, we use PGMs to glean defect information from social media data. Motivated by the steps of expert system modeling approach proposed by Pozna and Precup (2014), to discover defects from social media data, we first construct PDIAM and then use it to extract main topics, which are the defect information, finally FMEA method is used for defect severity evaluation. In this section, we focus on online threads, a typical kind of social media data, and introduce the integrated PDIAM and FMEA approach.

3.1. Data preprocessing and lexicon construction

Data preprocessing is an essential step when analyzing threads. It includes stop-word removal, common word removal and special symbols filtering. Stop-words and common words are meaningless words with high frequencies. Special symbols filtering is to delete useless symbols like URLs.

After data preprocessing, we analyze threads by PDIAM. In PDIAM, we assume a DR thread contains the general, component, symptom, or solution words. Component words are the words describing product components. Symptom words describe defect symptoms while solution words delineate the solutions that customers take to solve defects. Words that do not belong to these three types are general words. When PDIAM analyzing threads, we should decide the word type for each word. Thus, we use component, symptom and solution lexicons to provide information on word types beforehand. Words in these lexicons are extracted from maintenance reports respectively and lexicons have been validated by experts. If a word appears in the component, symptom or solution lexicon, then the word is a component, symptom or solution word. Otherwise, it is a general word.

3.2. PDIAM

PDAIM is a PGM concentrating on defect identification and defect

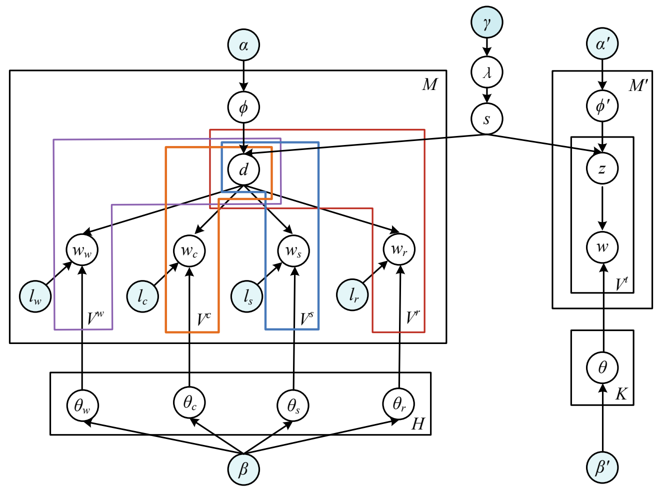


Fig. 1. Plate notation of PDIAM.

Table 2
Variable description.

Notation	Description
H	The number of defects
K	The number of topics
M, M'	The number of DR documents and DUR documents
V^t	The number of words in DUR documents
V^w, V^c, V^s, V^r	The number of general, component, symptom and solution words respectively
d_{ji}	Defect assignment to word i in document j
z_{ji}	Topic assignment to word i in document j
$d_{j,\neg i}$	Defect assignment to words in document j excluding word i
$z_{j,\neg i}$	Topic assignment to words in document j excluding word i
s_{ji}	Label assignment to word i in document j, indicating whether word i in document j is relevant ($s_{ji} = 1$) or irrelevant ($s_{ji} = 0$) to product defects
$s_{j,\neg i}$	Label assignment to words in document j excluding word i
w_w, w_c, w_s, w_r	General, component, symptom and solution words
l_w, l_c, l_s, l_r	The label of general, component, symptom and solution words
$\boldsymbol{\phi}$, $\boldsymbol{\phi}$'	Multinomial distribution over defects and topics
$\boldsymbol{\theta_w}$, $\boldsymbol{\theta_c}$, $\boldsymbol{\theta_s}$, $\boldsymbol{\theta_r}$	Multinomial distributions over general, component, symptom and solution words
$\boldsymbol{\theta}$	Multinomial distribution over words in DUR documents
α, α'	Dirichlet prior for hidden variable $\boldsymbol{\phi}$ and $\boldsymbol{\phi}$'
β, β'	Dirichlet prior for hidden variable $\boldsymbol{\theta_w}$, $\boldsymbol{\theta_c}$, $\boldsymbol{\theta_s}$, $\boldsymbol{\theta_s}$ and $\boldsymbol{\theta}$
$N_{1,\,j,\neg i}$	Number of words relevant to defects ($s_{ji} = 1$) in document j excluding word i
$N_{0,\,j,\neg i}$	Number of words irrelevant to defects ($s_{ji} = 0$) in document j excluding word i
$N_{jk,\neg i}$	Number of words assigned to topic k in document j excluding word i
$N^w_{ht,\neg i}$	Number of general words assigned to defect h in document j excluding word i
$N^c_{ht,\neg i}$	Number of component words assigned to defect h in document j excluding word i
$N^s_{ht,\neg i}$	Number of symptom words assigned to defect h in document j excluding word i
$N^r_{ht,\neg i}$	Number of solution words assigned to defect h in document j excluding word i

information collection. Fig. 1 displays the plate notation of PDIAM and Table 2 introduces the meanings of variables in PDIAM. In Fig. 1, l_w, l_c, l_s and l_r are labels indicating which word type the word w_{ji} is. These labels are valued 0 or 1. When $l_w = 1$, w_{ji} is the general word. Similarly for l_c, l_s and l_r. For brevity, we use $\mathbf{L}$ to represent the set of labels. $\mathbf{L} = \{\, l_w, l_c, l_s, l_r$ }.

In the document generative process of PDIAM, PDIAM first decides the label distributions. For each word w_{ji} in document j, the label s_{ji} is drawn from *Bernoulli* distribution λ. If s_{ji} equals 0, we then draw topic distribution ϕ_j' from *Dirichlet* distribution α' and word distribution θ_{zji} from distribution β'. If s_{ji} equals 1, we draw defect distribution ϕ_j from *Dirichlet* distribution α. The word distribution is drawn based on the word type of w_{ji}. If l_w equals 1, word distribution θ^w_{zji} is drawn from *Dirichlet* distribution β. Similarly for θ^c_{zji}, θ^s_{zji} and θ^r_{zji}. The detailed generative process of PDIAM is illustrated in Algorithm 1.

Algorithm 1: Generative process of PDIAM

for each document j **do**
  for each word $i \in j$ **do**
    draw $s_{ji} \sim$ *Bernoulli* (λ), where $\lambda \sim$ *Beta* (γ) which indicates the proportion between the DUR and DR probabilities of document j
    if $s_{ji} = 0$
      draw ϕ_j' ~ *Dirichlet* (α')
      draw a topic $z_{ji} \sim$ *Multi* (ϕ_j')
      draw $\theta_{zji} \sim$ *Dirichlet* (β')
      draw $w_{ji} \sim$ *Multi* (θ_{zji})
    elif $s_{ji} = 1$
      draw $\phi_j \sim$ *Dirichlet* (α)
      draw a defect $d_{ji} \sim$ *Multi* (ϕ_j)
      if $l_w = 1$
        draw $\theta^w_{zji} \sim$ *Dirichlet* (β)
        draw $w^w_{ji} \sim$ *Multi* (θ^w_{zji})
      elif $l_c = 1$
        draw $\theta^c_{zji} \sim$ *Dirichlet* (β)
        draw $w^c_{ji} \sim$ *Multi* (θ^c_{zji})
      elif $l_s = 1$
        draw $\theta^s_{zji} \sim$ *Dirichlet* (β)
        draw $w^s_{ji} \sim$ *Multi* (θ^s_{zji})
      elif $l_r = 1$
        draw $\theta^r_{zji} \sim$ *Dirichlet* (β)
        draw $w^r_{ji} \sim$ *Multi* (θ^r_{zji})
      end if
    end if
  end for
end for

3.3. Model inference

In the generative process of PDIAM, variables α, β, α', β', γ, H, K is preset while θ, θ_w, θ_c, θ_s, θ_r, ϕ, ϕ'and λ are unobserved. To estimate the hidden variables θ, θ_w, θ_c, θ_s, θ_r, ϕ, ϕ'and λ, we use the following posterior distribution for inference.

$$\begin{aligned} &p(\theta,\theta_w,\theta_c,\theta_s,\theta_r,\phi,\phi^{'},\lambda,z|\boldsymbol{w},\boldsymbol{s},\boldsymbol{L},\alpha,\alpha^{'},\beta,\beta^{'},\gamma)\\ &=\frac{p(\theta,\theta_w,\theta_c,\theta_s,\theta_r,\phi,\phi^{'},\lambda,z,\boldsymbol{w},\boldsymbol{s},\boldsymbol{L}|\alpha,\alpha^{'},\beta,\beta^{'},\gamma)}{p(\boldsymbol{w},\boldsymbol{s},\boldsymbol{L}|\alpha,\alpha^{'},\beta,\beta^{'},\gamma)}. \end{aligned} \tag{1}$$

It is interactable to estimate unobserved variables via Eq. (1). Hence, we use Gibbs Sampling to infer hidden variables. When $s_{ji} = 0$, word w_{ji} in document j is DUR. And the probability of w_{ji} assigned with topic k equals to

$$p(z_{ji}=k,s_{ji}=0|z_{j,\neg i},\boldsymbol{s}_{j,\neg i},\boldsymbol{w},\alpha^{'},\beta^{'},\gamma), \tag{2}$$

where $\mathbf{w}$ is the vector of words containing in document j. Applying the Bayesian rule to Eq. (2), we then obtain Eq. (3) as follows.

$$\begin{aligned} &p(z_{ji}=k,s_{ji}=0|z_{j,\neg i},\boldsymbol{s}_{j,\neg i},\boldsymbol{w},\alpha^{'},\beta^{'},\gamma)\\ &\propto p(s_{ji}=0|\boldsymbol{s}_{j,\neg i},\gamma)\cdot p(z_{ji}=k|z_{j,\neg i},\alpha^{'})\cdot p(w_{ji}|z_{ji}=k,\boldsymbol{z}_{j,\neg i},\boldsymbol{w}_{j,\neg i},\beta^{'}). \end{aligned} \tag{3}$$

Given that *Bernoulli* distribution is conjugate to *Beta* distribution, the first term of Eq. (3) can be written as Eq. (4):

$$p(s_{ji}=0|\boldsymbol{s}_{j,\neg i},\gamma)=\frac{N_{0,j,\neg i}+\gamma}{N_{0,j,\neg i}+N_{1,j,\neg i}+2\gamma}. \tag{4}$$

Given that *Dirichlet* distribution is conjugate to *Multinomial* distribution, the second term of Eq. (3) then can be written as Eq. (5):

$$p(z_{ji}=k|z_{j,\neg i},\alpha^{'})=\frac{N_{jk,\neg i}+\alpha^{'}}{\sum_{k'=1}^{K}N_{jk',\neg i}+K\alpha^{'}}. \tag{5}$$

And the third term of Eq. (3) can be written as Eq. (6):

$$p(w_{ji}|z_{ji}=k,\boldsymbol{z}_{j,\neg i},\boldsymbol{w}_{j,\neg i},\beta^{'})=\frac{N_{kt,\neg i}+\beta^{'}}{\sum_{t'=1}^{V^{t}}N_{kt,\neg i}+V^{t}\beta^{'}}. \tag{6}$$

Hence, when $s_{ji}=0$, Eq. (2) is calculated by:

$$p(z_{ji}=k,s_{ji}=0|\boldsymbol{z}_{j,\neg i},\boldsymbol{s}_{j,\neg i},\boldsymbol{w},\alpha^{'},\beta^{'},\gamma)\propto$$
$$\frac{N_{0,j,\neg i}+\gamma}{N_{0,j,\neg i}+N_{1,j,\neg i}+2\gamma}\cdot\frac{N_{jk,\neg i}+\alpha^{'}}{\sum_{k'=1}^{K}N_{jk',\neg i}+K\alpha^{'}}\cdot\frac{N_{kt,\neg i}+\beta^{'}}{\sum_{t'=1}^{V^{t}}N_{kt,\neg i}+V^{t}\beta^{'}}. \tag{7}$$

When $s_{ji}=1$, word w_{ji} in document j is DR. In this circumstance, the probability that w_{ji} is assigned with defect h is calculated by

$$\begin{aligned}&p(d_{ji}=h,s_{ji}=1|\boldsymbol{d}_{j,\neg i},\boldsymbol{s}_{j,\neg i},\boldsymbol{w},\boldsymbol{L}_{ji},\alpha,\beta,\gamma)\\ \propto\, &p(s_{ji}=1|\boldsymbol{s}_{j,\neg i},\gamma)\cdot p(d_{ji}=h|\boldsymbol{d}_{j,\neg i},\alpha)\cdot p(w_{ji}|d_{ji}=h,\boldsymbol{d}_{j,\neg i},\boldsymbol{w}_{j,\neg i},\boldsymbol{L}_{ji},\beta).\end{aligned} \tag{8}$$

In Eq. (8), $\mathbf{w}=\{\mathbf{w}^{\mathbf{w}},\mathbf{w}^{\mathbf{c}},\mathbf{w}^{\mathbf{s}},\mathbf{w}^{\mathbf{r}}\}$. Similar to Eqs. (4)–(6), Eq. (8) is calculated by:

$$\begin{aligned}&p(d_{ji}=h,s_{j,\neg i}=1|\boldsymbol{d}_{j,\neg i},\boldsymbol{s}_{j,\neg i},\boldsymbol{w},\boldsymbol{L}_{ji},\alpha,\beta,\gamma)\propto\\ &\frac{N_{1,j,\neg i}+\gamma}{N_{0,j,\neg i}+N_{1,j,\neg i}+2\gamma}\cdot\frac{N_{jh,\neg i}+\alpha}{\sum_{h'=1}^{D}N_{jh',\neg i}+H\alpha}\cdot\\ &\left(l_{w}\cdot\frac{N_{ht,\neg i}^{w}+\beta}{\sum_{t'=1}^{V^{w}}N_{ht,\neg i}^{w}+V^{w}\beta}+l_{c}\cdot\frac{N_{ht,\neg i}^{c}+\beta}{\sum_{t'=1}^{V^{c}}N_{ht,\neg i}^{c}+V^{c}\beta}+l_{s}\cdot\frac{N_{ht,\neg i}^{s}+\beta}{\sum_{t'=1}^{V^{s}}N_{ht,\neg i}^{s}+V^{s}\beta}+l_{r}\cdot\frac{N_{ht,\neg i}^{r}+\beta}{\sum_{t'=1}^{V^{r}}N_{ht,\neg i}^{r}+V^{r}\beta}\right)\end{aligned} \tag{9}$$

For a clearer introduction of the parameter inference process, we display the detailed steps in Algorithm 2. After the burn-in period of Gibbs sampling, we estimate *Multinomial* distributions $\boldsymbol{\phi}$, $\boldsymbol{\phi}$'**,** $\boldsymbol{\theta_w}$**,** $\boldsymbol{\theta_c}$**,** $\boldsymbol{\theta_s}$**,** $\boldsymbol{\theta_r}$**,** $\boldsymbol{\theta}$ and $\boldsymbol{s_{ji}}$ according to the convergent sampling result of Eqs. (7) and (9). $\boldsymbol{\phi}$ and $\boldsymbol{\phi}$**'** indicate the defect and topic distribution of documents separately. $\boldsymbol{\theta_w}$**,** $\boldsymbol{\theta_c}$**,** $\boldsymbol{\theta_s}$ and $\boldsymbol{\theta_s}$ indicate the word distributions for defects while $\boldsymbol{\theta}$ denotes the word distributions for topics. The probability of document j referring to defect h is given by:

$$\phi_{jh}=\frac{N_{jh}+\alpha}{\sum_{h'=1}^{D}N_{jh'}+H\alpha}. \tag{10}$$

The probability of document j referring to topic k is given by:

$$\phi_{jk}^{'}=\frac{N_{jk}+\alpha^{'}}{\sum_{k'=1}^{K}N_{jk'}+K\alpha^{'}}. \tag{11}$$

For defect h, the defect-word distributions for general, component, symptom and solution words are given respectively by:

$$\begin{aligned}&\theta_{w}^{ht}=\frac{N_{ht}^{w}+\beta}{\sum_{t'=1}^{V^{w}}N_{ht'}^{w}+V^{w}\beta},\quad \theta_{c}^{ht}=\frac{N_{ht}^{c}+\beta}{\sum_{t'=1}^{V^{c}}N_{ht'}^{c}+V^{c}\beta},\\ &\theta_{s}^{ht}=\frac{N_{ht}^{s}+\beta}{\sum_{t'=1}^{V^{s}}N_{ht'}^{s}+V^{s}\beta},\quad \theta_{r}^{ht}=\frac{N_{ht}^{r}+\beta}{\sum_{t'=1}^{V^{r}}N_{ht'}^{r}+V^{r}\beta}.\end{aligned} \tag{12}$$

And for topic k, the topic-word distributions are estimated by:

$$\theta_{kt}=\frac{N_{kt}+\beta}{\sum_{t'=1}^{V^{t}}N_{kt'}+V^{t}\beta}. \tag{13}$$

$\boldsymbol{s_{ji}}$ is the label distribution of word w_{ji} and is estimated by Eq. (14),

$$s_{ji}^{1}=\frac{N_{1,j,\neg i}+\gamma}{N_{0,j,\neg i}+N_{1,j,\neg i}+2\gamma},\quad s_{ji}^{0}=\frac{N_{0,j,\neg i}+\gamma}{N_{0,j,\neg i}+N_{1,j,\neg i}+2\gamma}, \tag{14}$$

where s_{ji}^{1} denotes the DR probability and s_{ji}^{0} denotes the DUR probability.

Algorithm 2: Parameter inference of PDIAM via Gibbs sampling

Input: α, β, α', β', γ, H, K, the number of extracted words N, *iteration threshold*
Output: defect index assignment, defect-related words of H defects, topic index assignment, topic-related words of K topics,
Steps:

$n=1$
while $n<=$ *iteration threshold* **do**
  for each document j **do**
    for each word $i\in j$ **do**
      calculate s_{ji} based on the *Bernoulli* distribution with the probability $p(s_{ji}=0|s_{j,\neg i},\gamma)$, which is calculated by Eq.(4)
    if $s_{ji}=0$
      for each topic index k **do**
        calculate $p(z_{ji}=k,s_{ji}=0|z_{j,\neg i},s_{j,\neg i},\boldsymbol{w},\alpha',\beta',\gamma)$ using Eq.(7)
      end for

(*continued on next page*)

(*continued*)

Algorithm 2: Parameter inference of PDIAM via Gibbs sampling

      assigning topic index $k^{'}$ to z_{ji},
      $k^{'}=argmax(p(z_{ji}=k,s_{ji}=0|z_{j,\neg i},s_{j,\neg i},\boldsymbol{w},\alpha',\beta',\gamma))$
    elif $s_{ji}=1$
      for each defect index h **do**
        calculating $p(d_{ji}=h,s_{ji}=1|\boldsymbol{d}_{j,\neg i},s_{j,\neg i},\boldsymbol{w},\boldsymbol{L}_{ji},\alpha,\beta,\gamma)$ using Eq.(9)
      end for
      assigning defect index $h^{'}$ to d_{ji},
      $h^{'}=argmax(p(d_{ji}=h,s_{ji}=1|\boldsymbol{d}_{j,\neg i},s_{j,\neg i},\boldsymbol{w},\boldsymbol{L}_{ji},\alpha,\beta,\gamma))$
    end if
  end for
  $n=n+1$
end while
for each topic index k **do**
  extract top N topic-related words using Eq.(13)
end for
for each defect index h **do**
  extract top N DR general, component, symptom and solution words using Eq. (12)
end for
return results of defect index assignment, defect-related words of H defects, topic index assignment, topic-related words of K topics

3.4. Product defect detection via PDIAM

After inferring the hidden variables, we identify DR threads and collect detailed defect information based on the results of PDIAM. We use Eq. (15) to decide whether document j is related to defects. In Eq. (15), v_j is the number of words contained in j. If f_j is larger than the preset threshold ε, document j discusses defects. Otherwise, document j has no relationship with defects.

$$f_{j}=\sum_{i=1}^{V_{j}}s_{ji}^{1}. \tag{15}$$

To collect detailed defect information from DR documents, we first determine defect types for threads. Considering that a sentence usually

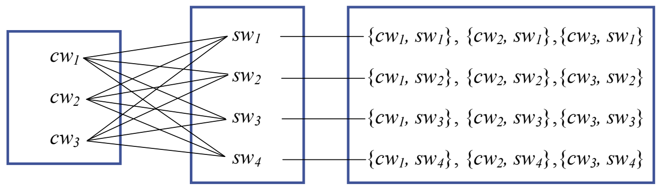


Fig. 2. An example of creating component-symptom pairs assigned to defect *h*.

has a discussed object, we first segment a thread into several sentences and leave the sentences containing DR words provided by PDIAM. For each sentence, its discussed defects are the defects of which DR words contain the component or symptom words appearing in this sentence. Then the thread's defect type is the one discussed by the most sentences. And detailed defect information is gathered based on its DR words.

3.5. Defect severity analysis via FMEA

We use the effective risk-management technique, FMEA, to identify the severity of defects discovered by PDIAM. We follow the steps of process FMEA which can be concluded into three aspects: (1) Identify product failure modes. (2) Evaluate failure modes in the aspects of the severity of failure modes (S), failure mode occurrence (O) and the detection probability of the failure (D). (3) Calculate the risk priority number (RPN) elements for failure mode assessment (Lo & Liou, 2018; Wang et al., 2018).

In our study, we do not change the FMEA procedure but use the social media data and proposed PDIAM to improve the identification of failure modes and the calculation of O and D. The traditional FMEA identifies failure modes via brainstorming and measure O and D by expert evaluation (Peeters et al., 2018). However, failure mode identification through brainstorming will lead to the incomprehensive consideration of defects. Moreover, because products are used under various conditions, defects that customers encounter will be more complicated and unpredictable in the after-sale stage. Thus, it will be incomprehensive that identifying failure modes only through brainstorming. In our study, the utilization of social media data and PDIAM can make the failure mode identification more comprehensive and alleviate the subjectivity. As for the calculation of O and D, the traditional FMEA estimates O and D by expert evaluation, which will bring inherent subjectivity and biased results (Bhattacharjee et al., 2020). In our research, we use PDIAM to measure O and D based on social media data. The leverage of PDIAM in the calculation avoids the subjectivity of expert evaluation and obtains more accurate results. Based on S, O and D, RPN is calculated by (Fattahi & Khalilzadeh, 2018):

$$RPN = S \times O \times D, \tag{16}$$

and the detailed evaluation process is introduced as follows.

- S is evaluated by experts using 1–10 scales (Peeters et al., 2018). We decide on the detailed assessment scale following the steps in Liou and Lo's study (Lo & Liou, 2018). And the evaluation results are normalized into [0,1] when measuring RPN values.
- O is measured by frequencies of defects in the thread dataset. PDIAM provides component and symptom words for each defect. For a certain defect, we use the pairs of component and symptom words as the indicator of the defect. Taking the following example in Fig. 2 for illustration, for each component word cw related to defect h, the component-symptom pairs consist of cw and each symptom word sw in symptom words related to h, respectively. To prevent the overlap of component-symptom pairs related to different defects, we change the statements of overlapping component-symptom pairs and ensure the uniqueness of all component-symptom pairs assigned to different defects. After obtaining component-symptom pairs, we measure O by counting the occurrences of component-symptom pairs for each defect. We conduct this step on the sentence level. For each thread, we first split it into several sentences and find component-symptom pairs for each sentence. The number of occurrences for a certain defect equals the number of sentences referring to component-symptom pairs assigned to this defect. We normalize the number of occurrences into [0, 1] and use the normalized values as the probabilities of defect occurrences
- D is obtained through $\boldsymbol{\phi}$ provided by PDIAM. $\boldsymbol{\phi}$ is the defect distribution of threads. For thread j, its defect distribution ϕ_j reflects the probabilities that thread j is related to H defects. In other words, ϕ_{jh} can be regarded as the probability that defect h is detected by thread j. Therefore, we use the average probabilities of H defect probability distributions as the detection probabilities. For defect h, its D value is calculated using Eq.(17). The calculation results are normalized into [0,1] for the measurement of RPN.

$$D_h = 1 - \frac{1}{M} \cdot \sum_{j=1}^{M} \phi_{jh}. \tag{17}$$

When obtaining RPN values for each defect, manufacturers can prioritize defects. The larger the RPN value, the more severe the defect is. Manufacturers should pay more attention to these severe defects and take remedial decisions promptly.

Table 3
Data attributes of thread dataset.

Attribute	
Number of threads	10,480
Number of DR threads	829
Number of general words	19,593
Number of component words	266
Number of symptom words	546
Number of solution words	71
The average number of words per thread	50

Table 4
Confusion matrix.

	Prediction = 0	Prediction = 1
Ground truth = 0	*TN*	*FN*
Ground truth = 1	*FP*	*TP*

4. Case study: Defect detection in the automobile industry

Defects are major concerns for both manufacturers and customers in the automobile industry. In this case study, we focus on the automobile industry to testify the effectiveness of our approach.

4.1. Experimental setup

4.1.1. Dataset

We gather online threads from Autohome (www.autohome.com), a famous website of automobiles in China, and crawled 10,480 Chinese threads from the Sagitar forum. The detailed data description is shown in Table 3. We tag these threads manually by four undergraduate students majoring in Industrial Engineering at Tianjin University and each thread is tagged by two students. The inter-annotator agreement rate is 88%. We conduct experiments using Python 3.7. In the calculation process of PDIAM. Parameters α, β, α', β', γ are set as the most PGM-based studies do, which are $\alpha = 50/H$, $\alpha' = 50/K$, $\beta = \beta' = 0.01$, $\gamma = 0.3$ (Wang et al., 2016; Jo & Oh, 2011; Steyvers & Griffiths, 2007; Zuo et al., 2016; Zoghbi et al., 2016). H (K) ranges from 10 to 50 with the step of 10. ε ranges from 0.5 to 0.9 with the step of 0.1.

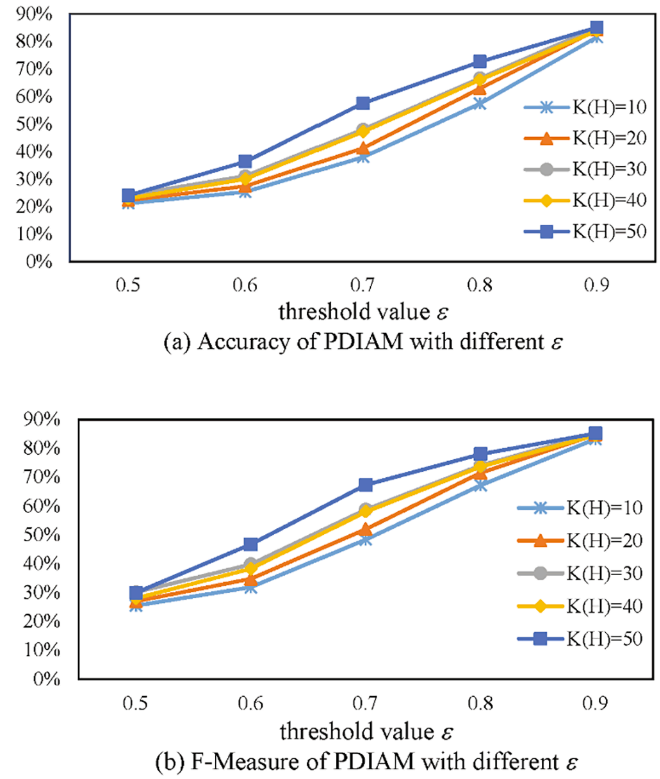


Fig. 3. Accuracy and F-Measure performance of PDIAM with various ε.

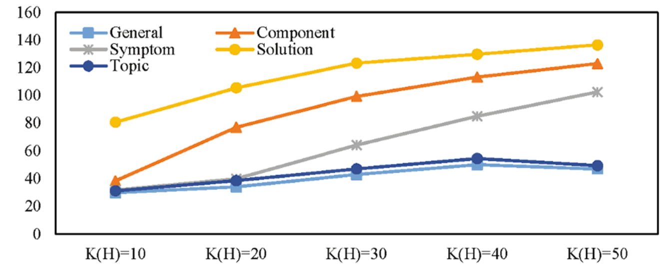


Fig. 4. Average PMI of PDIAM.

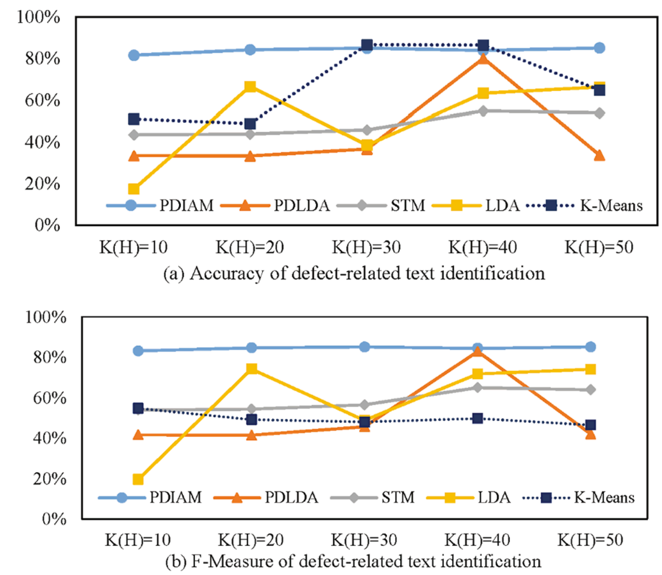


Fig. 5. Accuracy of topic models in DR text identification.

4.1.2. *Evaluation criteria*

We evaluate the performance of PDIAM from two perspectives: the accuracy of DR text identification and the topic quality (Liu et al., 2018; Zuo et al., 2016; Zhang et al., 2019).

- *Accuracy of DR text identification*: We use Accuracy and F-Measure for assessment. These metrics based on the confusion matrix in Table 4 are effective and popular to evaluate classification performance. They are calculated by Eq. (18)–(19).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{18}$$

$$F - Measure = \frac{2TP}{2TP + FP + FN}. \tag{19}$$

- *Topic quality*: We use topic coherence to evaluate the quality of topics extracted by different models. Topic coherence indicates the relevance between topics and their related words. The more coherent the words extracted by PGMs, the better the quality of PGMs. Following the process in the research of Newman et al. (Newman et al., 2010), we use the Pointwise Mutual Information (PMI) metric to measure topic coherence. For each topic model, we first measure the word PMI using Eq.(20):

$$PMI(w_1, w_2) = log_2\frac{p(w_1, w_2)}{p(w_1)p(w_2)}, \tag{20}$$

where $p(w_1, w_2)$ is the cooccurrence of w_1 and w_2. $p(w_1)$ and $p(w_2)$are the occurrence of w_1 and w_2 respectively. Then the average PMI for the topic model is obtained by Eq. (21),

$$AvgPMI = \frac{\sum_{k=1}^{K}\sum_{w_1,w_2\in \boldsymbol{KW}_k, w_1\neq w_2} PMI(w_1, w_2)}{K}, \tag{21}$$

where $\mathbf{KW}_k$ is the set of words related to the k^{th} topic and K is the number of topics. In our case study, the number of words in $\mathbf{KW}_k$ is set to 30. The larger the AvgPMI, the better the model performance.

4.1.3. *Benchmark methods*

To evaluate the performance of PDIAM, we select the following five effective models which are widely used in defect discovery.

- LDA (Blei et al., 2003): LDA is the classical topic model to extract topics from a large corpus. We use LDA to extract topics from our dataset and compare its performance to PDIAM.
- PDLDA (Zhang et al., 2019): PDLDA is an improved LDA designed to extract aspects of product components, defect symptoms, and resolutions.
- STM (Roberts et al., 2014, 2016): STM is an effective topic model and has been applied in the area of incident identification (Kuhn, 2018) and service quality measurement (Korfiatis et al., 2019). In our case study, we use STM as a baseline method to extract topics from our dataset.
- K-Means: K-Means is an effective cluster method and has been widely applied.
- Smoke words (Abrahams et al., 2012; Law et al., 2017; Winkler et al., 2016): Smoke word lexicon is developed for defect discovery. Smoke words have been proved to be effective in different areas.

The first four baseline models mentioned above are unable to identify DR texts directly. Hence, to compare the performance of DR text identification, we first use baseline models to derive topics and tag whether the topics are relevant to defects according to the topic-related

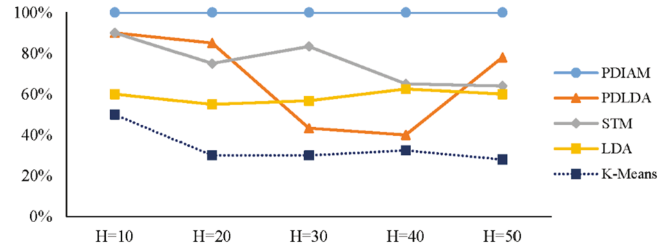


Fig. 6. Percentage of defect numbers identified by various models.

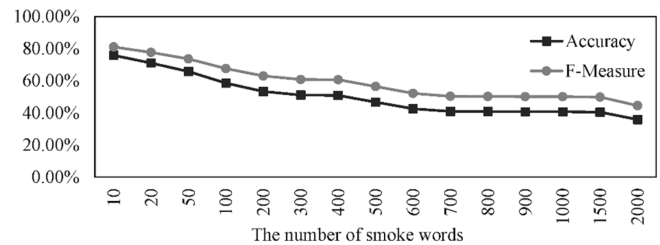


Fig. 7. Accuracy of smoke words.

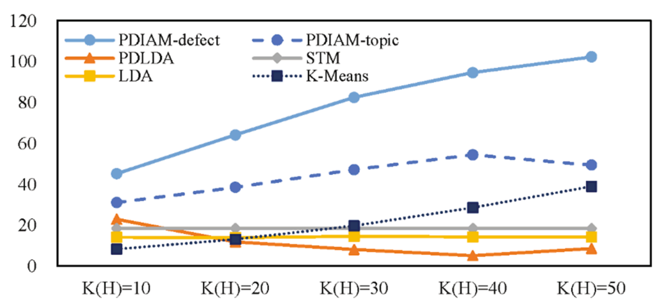


Fig. 8. Average PMI results.

words. Then we obtain topics assigned to each document. For a certain document, if its assigned topic is DR, this document is judged to be DR. Otherwise, this document is judged to be DUR.

4.2. Performance of PDIAM

We introduce the performance of PDIAM in Figs. 3 and 4. As Fig. 3 shows, both accuracy and F-Measure increase with the threshold value ε. And when the defect (topic) number is 50, PDIAM achieves the best performance in accuracy and F-Measure. When ε equals 0.5 or 0.9, both accuracy and F-Measure scores of PDIAM with various topic numbers are very close. But when ε equals 0.6, 0.7 or 0.8, PDIAM with 50 defect (topic) number is more effective.

Fig. 4 illustrates the average PMI performance of PDIAM. In Fig. 4, General, Component, Symptom, and Solution denote the general, component, symptom, and solution words derived from DR threads, respectively. And Topic denotes the topic words derived from DUR threads. From Fig. 4 we discover that all the AvgPMI values rise gradually with the number of defects (topics). And the performance of DR topics is better than DUR topics. Among the five kinds of words, solution words perform best, followed by component words. The performance of general words is very close to that of topic words. But the AvgPMI values of these two words are less comparable to the rest three kinds of words.

4.3. Experimental comparison with benchmark methods

Performance comparison of DR text identification. Fig. 5 depicts the accuracy of five models in DR text identification. Fig. 5(a) shows an

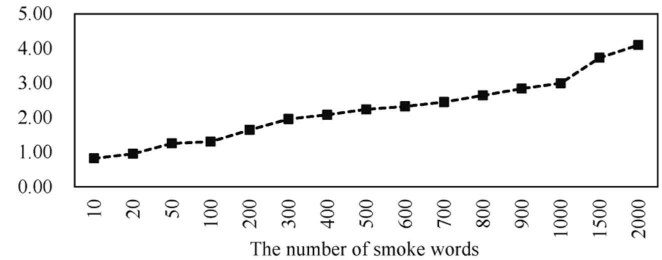


Fig. 9. PMI of smoke words.

Table 5
Time complexity of different topic models in an iteration of Gibbs Sampling.

Models	Time complexity
LDA	$O((M + M')\cdot K\cdot V)$
STM	$O((P + A)\cdot P\cdot(K\text{-}1)\cdot(M + M')^2\cdot V^2\cdot(K + A + A\cdot K))$
PDLDA	$O(2V^s + 2V^r+(M + M')\cdot K\cdot(V^c + V^s + V^r))$
PDIAM	$O((M + M')\cdot V + M\cdot H\cdot(V^w + V^c + V^s + V^r) + M'\cdot K\cdot V^t)$

P and *A* are numbers of topic covariates in STM. *V* is the number of words in $(M + M')$ documents.

obvious outperformance of PDIAM. No matter how the number of topics changes, PDIAM keeps stable accuracy values over 80%. Fig. 5(b) shows the F-Measure results of five models which indicate PDIAM predominates over the remaining four models. With the change of topic number, the F-Measure values of PDIAM are all larger than 80%. But concentrating on PDLDA and LDA, their F-Measure scores vary greatly with the topic number, and their best performance is still not comparable to that of PDIAM. Though STM has stable performance, its accuracy and F-Measure are less than 60%. K-Means achieves high accuracy when the topic number equals 30 or 40. But it has poor performance in the F-Measure score.

Besides, we compare the number of defects discovered by various models (displayed in Fig. 6). Fig. 6 shows that PDIAM is more effective in DR topic discovery than other models. The defect topics extracted by PDIAM are all related to product defects. PDLDA has a great fluctuation in the percentage of DR topics while the number of defects discovered by STM decreases gradually. Although LDA has stable performance, only about 60% of topics derived by LDA are relevant to defects. K-Means identifies the fewest defects than the other four models and thus it is ineffective to discover defects from the corpus.

The performance of smoke words keeps falling when the number of smoke words increases. That means fewer smoke words can obtain better identification results. However, as shown in Fig. 7, fewer smoke words will lead to smaller PMI values, which indicate less relevance these smoke words have. The accuracy and F-Measure scores of smoke words are closed to STM and LDA but are less comparable to PDIAM.

Comparison of topic quality. We display the Average PMI results in Fig. 8. In Fig. 8, PDIAM-defect denotes the average PMI trend of DR topics extracted by PDIAM while PDIAM-topic is the average PMI trend of DUR topics. PDIAM-defect is calculated based on DR words and PDIAM-topic is measured using topic words. Through Fig. 8 we find that both PDIAM-defect and PDIAM-topic achieve excellent performance. And PDIAM-defect performs better than PDIAM-topic, which means PDIAM is more effective in DR text analysis. As for PDIAM-defect, its average PMI shifts significantly with the increase of the topic number and reaches a peak when the number of topics is 50. While the average PMI of PDIAM-topic rises gradually and reaches the maximum when the topic number is 40. In contrast, PDLDA has good performance when the topic number is 10, but its average PMI drops gradually with the increase of the topic number. STM and LDA keep stable average PMI values when the topic number changes, but their performance is mediocre. K-Means's average PMI rises gradually but it cannot compete with PDIAM.

Table 6
20 defects extracted from the thread dataset.

Defects	Defect types	Defects	Defect types
D1	Abnormal sound in chassis	D11	Water leakage
D2	Window defects	D12	Electricity leakage
D3	Underpowered engine	D13	Car shuddering when shifting
D4	Abnormal sound when turning	D14	Abnormal sound when braking
D5	Engine shuddering	D15	Dashboard defects
D6	Peculiar smell	D16	Fuel leakage
D7	Sunroof defects	D17	Abnormal sound when accelerating
D8	Fuel cap cannot be closed	D18	Abnormal sound when reversing
D9	Air conditioning defects	D19	Abnormal sound in tire
D10	Clutch is not responsive	D20	Difficult to turn the steering wheel

Fig. 9 displays PMI results of smoke words. From Fig. 9 we observe that the maximum of smoke words' PMI score is small than 5, which indicates that the smoke word lexicon performs worst among six methods.

Comparison of analysis time. We further compare the time complexity of different models in an iteration of Gibbs Sampling and display results in Table 5. From Table 5 we see that LDA performs best in terms of time complexity. STM spends more time than LDA for the consideration of structural variables. PDLDA needs a longer time than LDA because it decides whether symptom and solution words are background words. PDIAM extracts topics from DR and DUR texts respectively. And when analyzing DR texts, PDIAM process texts with the consideration of different word types, which makes PDIAM conduct more calculations than LDA. But when H equals K, we find that the term of $M{\cdot}H{\cdot}(V^w + V^c + V^s + V^r) + M'{\cdot}K{\cdot}V^t$ is smaller than $(M + M'){\cdot}K{\cdot}V$ and PDIAM only consumes more time on deciding which documents are DR. Given that PDIAM gathers more information than baseline models and only spends more time on defect-related text determination, the time consumption of PDIAM in the text analysis is acceptable.

4.4. Defect and topic analysis

Defect analysis. We select 20 defects ($H = 20$) and describe detailed defects in Table 6. Detailed defect information for each defect is gathered from its related component, symptom and solution words. We take D1 for example and present D1′s related words in Table 7. For D1, its defective component is the chassis and the chassis often occurs the problems of the bulge, abnormal sound, cracks and rust. And the cars with defect D1 usually suffer from the problems of connecting rods and transmissions. As for solutions, customers choose to solve this defect problem by warranties or cleaning the chassis. General words of D1 convey the information that vehicles encounter defect D1 are usually refitted or equipped with dry dual-clutch transmissions.

Topic analysis. Topics extracted from online threads reflect customers' interests or concerns on vehicles but these interests or concerns are not relevant to defects. We also derive 20 topics and display them in Table 8. As Table 8 shows, customers like to discuss the price of cars, ways to drive and ask various questions regarding the purchase. They also like to share their lives. Through the topic words, we can collect more information about what customers care about. For instance, T1 is about insurance. T1′s related topic words in Table 7 show that what customers ask most is the insurance fee and whether the tire damage can be reimbursed with the insurance. And customers also concern about whether the insurance fee provided by 4S is more preferential.

Detailed defect information derived from threads. We take the following Thread A in Table 9 as an example to illustrate how to get detailed defect information from threads.

We first split Thread A into sentences and select the sentences containing DR words. DR words in these sentences are all underlined and annotated. We use **[C]** (**[S]** or **[R]**) to denote component (symptom or solution) words. For each sentence, we first decide the discussed defect type by searching which defect mentions the DR words. D8 (DR words are shown in Table 10) contains the DR words of Sentence (1) - (6) while D9 contains the DR word (*compressor*) of Sentence (6). Given that D8 is

Table 8
20 topics extracted from the thread dataset.

Topics	Defect types	Topics	Defect types
T1	Insurance	T11	Refit
T2	Price of used car	T12	Car version
T3	Travel	T13	Ways to warm cars
T4	Scraping	T14	Car paint
T5	Price	T15	Ways to drive
T6	Service of 4S	T16	Websites
T7	Package	T17	Cost performance
T8	Order	T18	Fuel consumption
T9	Wash cars	T19	Service Charges
T10	Exhibition	T20	Gifts

Table 9
Exampled thread for defect information derivation.

Thread A
Post: There seem to be fuel stains here.
Reply 1: Are those stains thrown out by the belt? I occur this problem every time I start my car.
Reply 2: 4S service center said there might be some problems with the fuel seal.
Reply 3: I had a warranty to repair my car for free.
Reply 4: I just bought the car and encountered this problem. It seems that the compressor is leaking fuel.

Annotated sentences containing DR words in Thread A

(1) *There seem to be <u>fuel stains</u>* **[S]** *here.*
(2) *Are those <u>stains</u>* **[S]** *thrown out by the <u>belt</u>* **[C]** *?*
(3) *I occur this problem every time I <u>start</u>* **[S]** *my car.*
(4) *4S service center said there might be some problems with the <u>fuel seal</u>* **[C]**.
(5) *I had a warranty to <u>repair</u>* **[R]** *my car for free.*
(6) *It seems that the <u>compressor</u>* **[C]** *is <u>leaking fuel</u>* **[S]**.

Table 10
Top 5 DR words of D8.

Words	top1	top2	top3	top4	top5
Component	fuel cap	fuel seal	remote control	fuel pipe	solenoid valve
Symptom	burning	flash	fuel stain	start	fuel leakage
Solution	replace	unlock	repair	check	complain

Table 7
Example of top 5 words of D1 and T1.

Words	top1	top2	top3	top4	top5
General	deep	refit	drive	appear	dry
Component	chassis	connecting rod	transmission	regulator	control lever
Symptom	bulge	abnormal sound	stuck	cracking	rust
Solution	quality testing	warranty	lower	handle	clean
Topic	insurance	tire repair	4S	fee	preferential

Table 11
Normalized FMEA results of 20 defects.

Defect	*S*	*O*	*D*	RPN	Rank
D1	0.2000	0.9879	0.9496	0.1876	18
D2	0.3000	0.9636	0.9517	0.2751	15
D3	0.8000	0.9691	0.9488	0.7356	2
D4	0.6000	0.9636	0.9498	0.5491	8
D5	0.8000	0.9349	0.9486	0.7095	3
D6	0.2000	0.4823	0.9497	0.0916	19
D7	0.3000	0.7660	0.9502	0.2184	17
D8	0.5000	0.0508	0.9506	0.0241	20
D9	0.7000	0.9415	0.9495	0.6258	6
D10	0.8000	0.7274	0.9489	0.5522	7
D11	0.5000	0.8234	0.9499	0.3911	11
D12	0.7000	0.4514	0.9496	0.3001	14
D13	0.6000	0.9636	0.9494	0.5489	9
D14	0.6000	0.8852	0.9488	0.5039	10
D15	0.4000	1.0000	0.9518	0.3807	12
D16	0.9000	0.8985	0.9491	0.7674	1
D17	0.7000	0.9415	0.9495	0.6258	5
D18	0.4000	0.9338	0.9499	0.3548	13
D19	0.3000	0.9249	0.9511	0.2639	16
D20	0.7000	0.9636	0.9538	0.6433	4

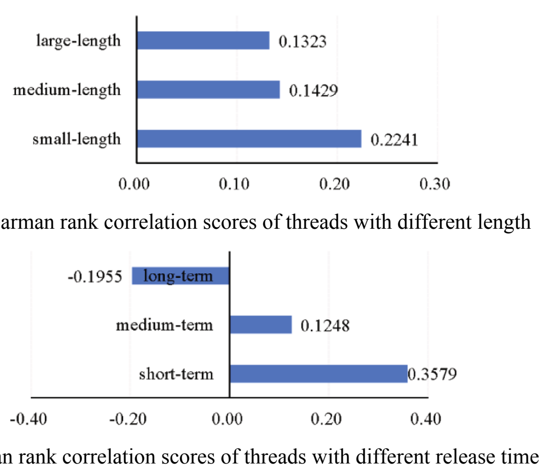


Fig. 10. Spearman rank correlation scores.

the defect discussed by most sentences in Thread A, we decide the defect type mentioned in Thread A is D8. The defective component may be the fuel cap and the defect symptom is fuel leakage. By observing the solution words of D8, we find that customers encountering these defects usually ask for aftersales to replace the defective component.

4.5. Defect severity analysis via FMEA

We use the result of PDIAM when D is 20 and ε is 0.9 to identify the severity of defects. We select the component and symptom words for each defect to generate component-symptom pairs and display the normalized results of S, O, and D in Table 11. We rank 20 defects based on their RPN values and analyze their severity. Among these 20 defects, D16 is the most severe defect not only for the large probabilities of occurrences but also for the high detection probability and the serious consequences. D3 and D5 are also severe defects which may cause great remedial costs and potential damage to word-of-mouth. Besides, manufacturers need to put more attention to D10 for its great severity. In contrast, D8 is the most negligible one. The less severe consequences and the probability of occurrence make it not so threatening than other defects.

4.6. The influence of text features on defect severity results

To further explore the influences of text features including thread length and release time on results, we conduct experiments on threads with different features.

To evaluate the impact of thread length, we divide threads into three categories: small-length, medium-length, and large-length threads. Word numbers of a thread are used to measure thread length. A small-length thread contains 1–10 words while a medium-length thread has 20–30 words. The large-length thread owns more than 40 words. We extract 300 threads with various lengths from the whole dataset randomly. Similarly, we classify threads with different release times into short-term, medium-term, and long-term threads. Short-term threads are released within 30 days before June 28, 2017 (the latest release date of threads) while medium-term threads are released within 60–90 days. The release time of long-term threads is longer than 120 days. For each type of release time, we collect 300 threads randomly for experiments. To estimate the result difference between threads with various text features and the whole dataset, we use Spearman's Rank Correlation coefficient (J & S, 2018) for measurement and present experimental results in Fig. 10.

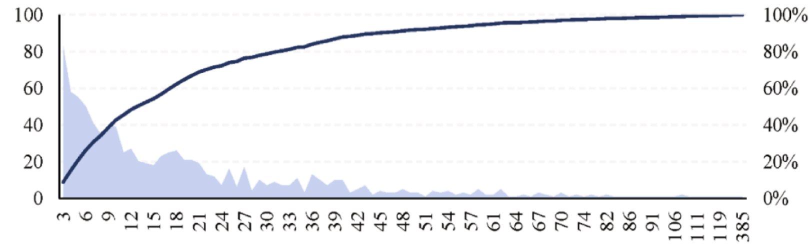
From Fig. 10 **(a)** we observe that results derived from small-length threads are closest to all threads. And a distinct difference between large-length threads and the whole dataset is discovered. In the aspect of release time, short-term threads have results closest to all threads. On the contrary, the results of long-term threads are totally different from results derived from the whole threads.

For further exploring why results derived from threads with various text features have significant differences, we calculate the length and release time distributions of all DR threads (as shown in Fig. 11). Both distributions illustrate that most DR threads are small-length and short-term, which means most defect information is delivered in small-length and short-term threads. Hence, the results of small-length and short-term threads are closest to the whole dataset.

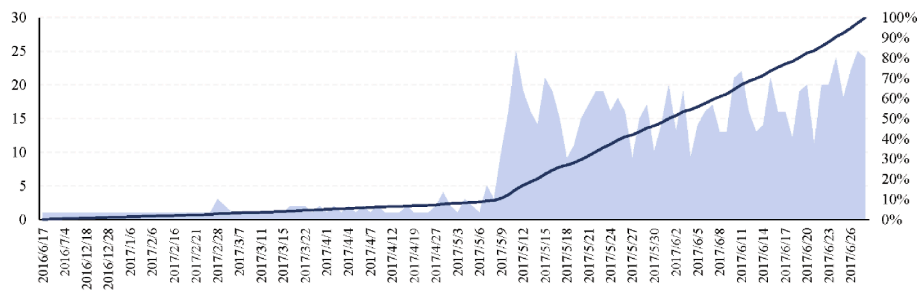
5. Conclusion

In this study, we propose a PGM named PDIAM and integrate it with FMEA to identify product defects from social media data. To solve the deficiencies of the extant PGMs, PDIAM first discovers DR texts and then derives defect information from these texts. Collecting defect information only from DR texts enhances the accuracy and comprehensiveness of results. In the generative process of PDIAM, the classification of general, component, symptom, and solution words ensures more detailed defect information including defect types, defective components, defect symptoms, and defect solutions can be provided by PDIAM. Moreover, to measure defect severity, we utilize FMEA based on the results of PDIAM and alleviates the inherent subjectivity brought by expert evaluation. A case study in the automobile industry is used to validate the effectiveness of our approach. In contrast to baseline methods, our method not only has an overwhelming and stable performance in DR thread identification but also derives more detailed defect information which helps manufacturers to discover product defects promptly and take remedial actions properly.

There are some limitations in our work that need to be addressed in the future. Firstly, our study doesn't provide a method to decide the number of defects or topics. How to select the proper defect and topic number to achieve the best performance of defect identification remains a question. Secondly, for the defects that have never appeared before, how to discover these defects also need further research. In future work, how to estimate the severity of failure modes in a more scientific way can be further explored.



(a) Length distribution of DR threads



(b) Release time distribution of DR threads

Fig. 11. Length and release time distributions of DR threads.

CRediT authorship contribution statement

Lu Zheng: Conceptualization, Methodology, Software, Visualization, Writing - original draft. **Zhen He:** Writing - review & editing, Validation, Funding acquisition. **Shuguang He: . :** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos.71661147003, 71532008, 71872123 and 71572122).

References

Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., & Jiao, J. (2015). An Integrated Text Analytic Framework for Product Defect Discovery. *Production and Operations Management, 24*(6), 975–990. https://doi.org/10.1111/poms.12303

Abrahams, A. S., Jiao, J., Fan, W., Wang, G. A., & Zhang, Z. (2013). What's buzzing in the blizzard of buzz? Automotive component isolation in social media postings. *Decision Support Systems, 55*(4), 871–882. https://doi.org/10.1016/j.dss.2012.12.023

Abrahams, A. S., Jiao, J., Wang, G. A., & Fan, W. (2012). Vehicle defect discovery from social media. *Decision Support Systems, 54*(1), 87–97. https://doi.org/10.1016/j.dss.2012.04.005

Alkahtani, M., Choudhary, A., De, A., & Harding, J. A. (2018). A decision support system based on ontology and data mining to improve design using warranty data. *Computers & Industrial Engineering*. https://doi.org/10.1016/j.cie.2018.04.033

Amplayo, R. K., Lee, S., & Song, M. (2018). Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis. *Information Sciences, 454–455*, 200–215. https://doi.org/10.1016/j.ins.2018.04.079

Bhattacharjee, P., Dey, V., & Mandal, U. K. (2020). Risk assessment by failure mode and effects analysis (FMEA) using an interval number based logistic regression model. *Safety Science, 132*, Article 104967. https://doi.org/10.1016/j.ssci.2020.104967

Bi, J.-W., Liu, Y., Fan, Z.-P., & Cambria, E. (2019). Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model. *International Journal of Production Research, 57*(22), 7068–7088. https://doi.org/10.1080/00207543.2019.1574989

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Chen, T.-H., Shang, W., Nagappan, M., Hassan, A. E., & Thomas, S. W. (2017). Topic-based software defect explanation. *Journal of Systems and Software, 129*, 79–106. https://doi.org/10.1016/j.jss.2016.05.015

Chenghua Lin, & Yulan He. (2009). Joint sentiment/topic model for sentiment analysis. Proceeding of the 18th ACM conference on Information and knowledge management, Hong Kong, China.

Du, L., Buntine, W., Jin, H., & Chen, C. (2012). Sequential latent Dirichlet allocation. *Knowledge and Information Systems, 31*(3), 475–503. https://doi.org/10.1007/s10115-011-0425-1

Farhadloo, M., Patterson, R. A., & Rolland, E. (2016). Modeling customer satisfaction from unstructured data using a Bayesian approach. *Decision Support Systems, 90*, 1–11. https://doi.org/10.1016/j.dss.2016.06.010

Fattahi, R., & Khalilzadeh, M. (2018). Risk evaluation using a novel hybrid method based on FMEA, extended MULTIMOORA, and AHP methods under fuzzy environment. *Safety Science, 102*, 290–300. https://doi.org/10.1016/j.ssci.2017.10.018

F. Wang R. Liu Y. Zuo H. Zhang H.e. Zhang W.u. Junjie (2016). Robust Word-Network Topic Model for Short Text. IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI) 2016 CA, USA San Jose.

Gao, S., Tang, O., Wang, H., & Yin, P. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management, 71*, 19–32. https://doi.org/10.1016/j.ijhm.2017.09.004

Goldberg, D. M., & Abrahams, A. S. (2018). A Tabu search heuristic for smoke term curation in safety defect discovery. *Decision Support Systems, 105*, 52–65. https://doi.org/10.1016/j.dss.2017.10.012

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

J, A. K., & S, A. Aspect-based opinion ranking framework for product reviews using a Spearman's rank correlation coefficient method Information Sciences 460–461 2018 23 41 10.1016/j.ins.2018.05.003.

James, T. L., Villacis Calderon, E. D., & Cook, D. F. (2017). Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Systems with Applications, 71*, 479–492. https://doi.org/10.1016/j.eswa.2016.11.004

Jin, J., Ji, P., Liu, Y., & Johnson Lim, S. C. (2015). Translating online customer opinions into engineering characteristics in QFD: A probabilistic language analysis approach. *Engineering Applications of Artificial Intelligence, 41*, 115–127. https://doi.org/10.1016/j.engappai.2015.02.006

Jo, Y., & Oh, A. (2011, February 9). Aspect and sentiment unification model for online review analysis. WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining, HongKong, China.

Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. In MIT Press.

Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications, 116*, 472–486. https://doi.org/10.1016/j.eswa.2018.09.037

Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies, 87*, 105–122. https://doi.org/10.1016/j.trc.2017.12.018

Law, D., Gruss, R., & Abrahams, A. S. (2017). Automated defect discovery for dishwasher appliances from online consumer reviews. *Expert Systems with Applications, 67*, 84–94. https://doi.org/10.1016/j.eswa.2016.08.069

Liu, Y., Jiang, C., & Zhao, H. (2018). Using contextual features and multi-view ensemble learning in product defect identification from online discussion forums. *Decision Support Systems, 105*, 1–12. https://doi.org/10.1016/j.dss.2017.10.009

Liu, Y., Jiang, C., & Zhao, H. (2019). Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media. *Decision Support Systems, 123*, Article 113079. https://doi.org/10.1016/j.dss.2019.113079

Lo, H.-W., & Liou, J. J. H. (2018). A novel multiple-criteria decision-making-based FMEA model for risk assessment. *Applied Soft Computing, 73*, 684–696. https://doi.org/10.1016/j.asoc.2018.09.020

Lo, S. (2008). Web service quality control based on text mining using support vector machine. *Expert Systems with Applications, 34*(1), 603–610. https://doi.org/10.1016/j.eswa.2006.09.026

Newman, D., Noh, Y., Talley, E., Karimi, S., & Baldwin, T. (2010). June 21). *Evaluating topic models for digital libraries*. In *Proceedings of the 2010 Joint International Conference on Digital Libraries, JCDL 2010*. https://doi.org/10.1145/1816123.1816156

Nicholas G. Rupp. (2004). The attributes of a costly recall: Evidence from the automotive industry. Review of Industrial Organization, 25(1), 21–44. https://doi.org/DOI 10.1023/B:REIO.0000040514.22968.e1.

Nord, J. H., Lee, T.-R. (Jiun-S., Çetin, F., Atay, Ö., & Paliszkiewicz, J. (2016). Examining the impact of social technologies on empowerment and economic development. International Journal of Information Management, 36(6, Part A), 1101–1110. https://doi.org/10.1016/j.ijinfomgt.2016.08.001.

Peeters, J. F. W., Basten, R. J. I., & Tinga, T. (2018). Improving failure analysis efficiency by combining FTA and FMEA in a recursive manner. *Reliability Engineering & System Safety, 172*, 36–44. https://doi.org/10.1016/j.ress.2017.11.024

Pernkopf, F., Peharz, R., & Tschiatschek, S. (2014). Introduction to Probabilistic Graphical Models. In Academic Press Library in Signal Processing (Vol. 1, pp. 989–1064). https://doi.org/10.1016/B978-0-12-396502-8.00018-8.

Pozna, C., & Precup, R. E. (2014). Applications of signatures to expert systems modelling. *Acta Polytechnica Hungarica, 11*(2), 21–39. https://doi.org/10.12700/aph.11.02.2014.02.2.

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association, 111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., … Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science, 58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Shams, M., & Baraani-Dastjerdi, A. (2017). Enriched LDA (ELDA): Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction. *Expert Systems with Applications, 80*, 136–146. https://doi.org/10.1016/j.eswa.2017.02.038

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handb. Latent Semant. Anal, 427*, 424–440.

Topîrceanu, A., & Precup, R.-E. (2020). A framework for improving electoral forecasting based on time-aware polling. *Social Network Analysis and Mining, 10*(1), 39. https://doi.org/10.1007/s13278-020-00646-7

Wang, W. M., Li, Z., Tian, Z. G., Wang, J. W., & Cheng, M. N. (2018). Extracting and summarizing affective features and responses from online product descriptions and reviews: A Kansei text mining approach. *Engineering Applications of Artificial Intelligence, 73*, 149–162. https://doi.org/10.1016/j.engappai.2018.05.005

Wang, Y., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems, 105*, 87–95. https://doi.org/10.1016/j.dss.2017.11.001

Winkler, M., Abrahams, A. S., Gruss, R., & Ehsani, J. P. (2016). Toy Safety Surveillance from Online Reviews. *Decision Support Systems, 90*, 23–32. https://doi.org/10.1016/j.dss.2016.06.016

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Zuo, Yuan, Zhao, Jichang, & Ke, Xu. (2016). Word network topic model a simple but general solution for short and imbalanced texts. *Knowledge & Information Systems, 48* (2), 379–398.

Zhang, L., Chu, X., & Xue, D. (2018). Identification of the to-be-improved product features based on online reviews for product redesign. *International Journal of Production Research, 57*(8), 2464–2479. https://doi.org/10.1080/00207543.2018.1521019

Zhang, W., Xu, H., & Wan, W. (2012). Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications, 39*(11), 10283–10291. https://doi.org/10.1016/j.eswa.2012.02.166

Zhang, X., Qiao, Z. L., Ahuja, A., Fan, W. G., Fox, E. A., & Reddy, C. K. (2019). Discovering Product Defects and Solutions from Online User Generated Contents. 3441–3447. https://doi.org/10.1145/3308558.3313732.

Zhang, Xuan, Niu, S., Zhang, D., Wang, G. A., & Fan, W. (2015, May 19). Predicting Vehicle Recalls with User-Generated Contents: A Text Mining Approach. Intelligence and Security Informatics: Pacific Asia Workshop, PAISI 2015, Ho Chi Minh City, Vietnam.

Identifying product defects from user complaints a probabilistic defect model, 9074, (2016), 41–50. https://doi.org/10.1007/978-3-319-18455-5_3

Zheng, L., He, Z., & He, S. (2020). A novel probabilistic graphic model to detect product defects from social media data. *Decision Support Systems, 137*, Article 113369. https://doi.org/10.1016/j.dss.2020.113369

Zoghbi, S., Vulić, I., & Moens, M.-F. (2016). Latent Dirichlet allocation for linking user-generated content and e-commerce data. *Information Sciences, 367–368*, 573–599. https://doi.org/10.1016/j.ins.2016.05.047