Reliability Engineering and System Safety 199 (2020) 106925

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Assessing the reliability of electronic products using customer knowledge discovery

Xing Pan[a], Huixiong Wang[a], Weijia You[b,*], Manli Zhang[a], Yuexiang Yang[c]

[a] School of Reliability and Systems Engineering, Beihang University, Beijing 100191, PR China
[b] School of Economics and Management, Beijing Forestry University, Beijing 100083, PR China
[c] School of Management, China University of Mining and Technology (Beijing), Beijing 100083, PR China

ARTICLE INFO

Keywords:
Reliability analysis
Parameter estimation
Customer knowledge
Text mining

ABSTRACT

Reliability is an essential aspect of product quality that concerns both customers and manufacturers. Other than test-based reliability data, online reviews reflect the actual operating status of all products, where the sample is all delivered products, and the test condition is equivalent to the service condition. This manuscript proposes the framework of a reliability analysis based on online reviews, and combines the statistical reliability analysis with current text mining technology. The proposed method adopts lexicon-based text mining technology to extract the failure-related customer knowledge from product users' online reviews. Using the information on the symptom and time of each failure experienced by customers, we classify the failure for each component and analyze the reliability using the estimated parameters of failure distributions. A comparative analysis is proposed to eliminate the uncertainty accompanying the review information. The application of the proposed framework is demonstrated by a case study of two similar mobile phone products. The results indicate that the consideration of failure distribution affects the analytical results significantly, and that the type of components, rather than the product model, has a greater impact on the product reliability.

1. Introduction

Since its first appearance in the 1950s, the term "reliability" has attracted considerable attention in both academia and industry. The reliability is the capacity of an item (component, subsystem, or system) to perform its intended function for a given period of time under a specified operating environment [1]. With the improvements in the development of design and manufacturing technology, the reliability of the product has also increased. Therefore, the method of using qualitative metrics is no longer able to meet the changing requirements of the reliability analysis. Consequently, it is necessary to quantify the reliability and to make the reliability of a product "visible" and "comparable," by which designers and manufacturers can monitor and control the product quality in life circles [2]. However, it is sometimes difficult to provide sufficient failure data using traditional test-based reliability metrics because of the low failure rate, especially when no failures are observed [3,4]. In addition, to obtain accurate results, the optimal reliability test is supposed to be conducted under normal operating conditions, which is impractical in terms of time and cost. Therefore, analysts have attempted to identify additional sources of failure data other than laboratory tests, such as the field data.

In reliability engineering, especially in the continuous improvement process of reliability, field data collected from the customers comprise an important source of data. As a special form of "test" data, field data provides a remedial approach for uncertainty because they are collected from real life profiles. Specifically, as an accessible source of field data, customer feedback captures combinations of real environmental exposures. According to Roesch et al. [5], customer fallout is identified as "natural" because the life profile is not accelerated in the usage, and it is free from data manipulation or intentional enhancement compared to other forms of reliability data. Furthermore, comparative analyses of customers' feedback enable analysts to determine the advantages and disadvantages of a product, compared with other competitive products [6]. Therefore, it is important to develop suitable procedures for collecting the field data during the reliability design and verification [7].

In recent times, with the maturing Internet technology, customers are encouraged to express their opinions on the products and services acquired, which facilitates the customer knowledge discovery as well as product improvement. For manufacturers, online reviews have become an important source of customer knowledge, and many efforts have been made to obtain valuable feedback from these reviews [8,9]. Previous studies [10–12] also emphasized customers' reviews in product

* Corresponding author at: No. 35, Tsinghua East Road, Haidian District, Beijing 100083, PR China.
E-mail address: wjyou@bjfu.edu.cn (W. You).

https://doi.org/10.1016/j.ress.2020.106925
Received 15 June 2019; Received in revised form 13 January 2020; Accepted 12 March 2020
Available online 14 March 2020
0951-8320/

knowledge discovery, and it is verified that the "word of mouth" (WOM) has a significant impact on sales. The growing reliance on online reviews is playing an increasingly important role in the business world, generating large quantities of data that outline the quality of products. In addition, the advancement of quantitative text processing techniques has also contributed to a more precise understanding of customers' experiences with products. With the development of text mining, machine learning helps to extract semantics, sentiments, and emotions from large numbers of reviews [13,14]. In the new era of big data, the use of automatic or semiautomatic methods of knowledge discovery for online review data will benefit both manufacturers and merchants significantly in terms of maintaining their advantages with their products.

In this study, we propose a framework to evaluate the reliability of electronic products based on customer knowledge discovery. Text analyses of customers' online reviews are employed to obtain information on product failures experienced by the customers, based on which statistical reliability analyses are conducted. A comparative analysis is proposed in order to eliminate the uncertainty contained in the review information. A case study on two smart phone models is conducted, from which we find that the failure distribution affects the analytical results of product reliability, and that the reliability varies with the type of component rather than product models. The remainder of the manuscript is organized as follows: Section 2 introduces the related work on data-based reliability analysis, big data in the online review, and the text mining technology for analyzing the failure-related information. Section 3 illustrates the framework of metrics used in this work. Section 4 focuses on the text analysis method, which is used to recognize the failure information in the review data. Section 5 introduces the process of parameter estimation and reliability analysis based on review information. In Section 6, we demonstrate the proposed methodology using a comparative case study of two similar smart phone models. Section 7 discusses the advantages and limitations of the proposed method, and offers directions for future research.

2. Related work

2.1. Data-driven reliability analysis

Since the concept of reliability engineering first appeared as a scientific discipline in the 1950s [2], there have been extensive studies of data-driven reliability analysis, including data collection and data processing approaches. To obtain the life and failure data, two basic sources are used to acquire reliability data: laboratory test data and field data. Zhang et al. [15] reviewed the taxonomy, method, and effect of existing mechanical reliability tests, and discussed the potential disadvantages of laboratory tests on reliability. In determining the scope of reliability tests, Vintr [16] summarized the methodologies of data analysis on the basis of current reliability test designs. Further, Elsayed [17] presented relevant models for reliability prediction through degradation data and life data derived from different reliability test plans. The above research on test-based reliability analyses provides methods to collect data and to evaluate the product reliability from laboratory tests.

In addition to laboratory tests, another important data source is the field data, which is generated under real product life profiles, e.g., warranty data, failure reports, and maintenance data. As the incompleteness and ambiguity in information affect analysis results, a majority of recent works have attempted to recognize and eliminate the uncertainty when extracting product information from the field data. Wang et al. [18] proposed a flexible failed-but-not-reported model to perform reliability analyses, which eliminates the errors in customers' behavior models. Tian [19] developed a data cluster-based reliability model to perform reliability analyses with the lack of long-term dependency. Altun et al. [20] proposed a reliability prediction model that can make long-term (3 years) reliability predictions using short-term field data of as short as three months. The above works discuss the methods to process the field data from different perspectives of uncertainty. In addition, data analysis is another important aspect in the reliability analysis. Dai et al. [21] proposed to perform reliability analyses on the basis of two-dimensional warranty data collection. Cheng [22] proposed an integrated platform for qualitative and quantitative reliability analyses on the spaceship, in which the data-driven failure mode and effects analysis (FMEA) is implemented at the component level. As a low-cost but informative source of product data, field data have shown the ability to depict the product quality in the product design stage. In this study, we utilize customers' online reviews as the data source in reliability evaluation and product comparison.

2.2. Customer knowledge in online-review

The emergence of electronic commerce and the digitalization of traditional retail businesses have generated a large amount of field data for product research. As online shopping has received growing popularity, the use of online reviews has been a widely accepted source for customers to identify the quality of goods ahead of making purchase decisions. According to Abrahams et al. [10], automotive customers use social media to obtain the information about vehicle defects, which influences customers' impressions of products. Pavlou et al. [11] found that the criticality of reviews has a significant impact on customers' trust in merchants; further, customers can choose between merchants by browsing and comparing the online reviews. In particular, online reviews of high involvement products affect product sales significantly [12]. Online reviews include a large amount of information about the product performance, quality, and customers' experiences with products. Therefore, as an important reference when customers are making purchase decisions, online reviews are valuable assets in the customer knowledge discovery, as well as in product and service improvement.

Apart from the attention from the customers, a growing number of merchants and manufacturers have also begun to realize the value of online reviews. To understand the customers' experience and evaluation of products from the perspectives of manufacturers, customer knowledge has attracted much attention in both business and industry. First defined by Wayland et al. [23], customer knowledge management (CKM) is defined as the effective use of information and experience in the acquisition, development, and retention of a profitable customer portfolio. Customer knowledge can be categorized into three separate areas, i.e., knowledge for customers, knowledge about customers, and the knowledge possessed by customers [24]. In this study, we focus on the third part because the feedback from the customer has been recognized as an important source for quality improvement [8,9] as customers are the natural inspectors of product failures and defects. Hence, online product reviews are important for both manufacturers and merchants because they provide numerous resources in product iteration and verification.

2.3. Text mining for product defect

Several efforts have been made to obtain customer knowledge for defect discovery and quality management from social media. Abrahams et al. [25] proposed a framework based on online reviews called SMART, which can quantify social media content, identify, and analyze product defects. Jin et al. [26] proposed an approach to automatically translate online reviews into engineering characteristics (ECS) for quality function deployment (QFD), which helps designers to focus more on the customers' need in product design. Zheng et al. [27] proposed a design defect management framework to identify and analyze design defects. Kirmani [28] focused on how a firm may signal the unobservable quality of its products by using several marketing-mix variables. Liu et al. [29] proposed a new method to derive contextual features from replies and used a multi-view ensemble learning method to identify product defects from social media.

Text mining has been a useful tool to analyze online reviews and support decision-making, and information extraction is a crucial technology in mining online reviews. As a common procedure in customer knowledge discovery, feature extraction has developed mature approaches to determine whether a text string contains specific information that is needed. The following are commonly used methods of feature extraction in the literature: (a) Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF is a numerical statistic measure that aims to reflect how important a word is to a document in a collection or corpus. Reyes et al. [30] applied TF-IDF to find frequent sequences of words considering n-grams of different orders. You et al. [13] defined the evaluation function as an extension of the classical TF-IDF to extract customer knowledge from online reviews. (b) Information Gain (IG). Abrahams et al. [31] used IG and three other methods to select the most important term features in their research, and algorithms with IG feature extraction achieved the highest classification performance. (c) Mutual Information (MI). In probability theory and information theory, MI is a measure of the mutual dependence between two variables. Peng et al. [32] applied MI to the feature selection, and demonstrates the feasibility of the method. (d) Chi-square (CS). Abrahams et al. [31] compared chi-square to three other methods to select the most important term features, and showed the superiority of this method. (f) Latent Dirichlet Allocation (LDA). LDA is a generative statistical model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar [33]. Dey et al. [34] used this method to extract topics of social media content. (g) Bayesian network. The Bayesian network is a probabilistic graphical model (a type of statistical model) that represents a set of variables and their conditional dependencies via a directed acyclic graph. Zheng et al. [27] applied Bayesian networks to the construction of a defect assessment model.

3. Framework

This Section provides an introductory framework to the proposed reliability analysis method.

In the proposed method, we propose to retrieve the life time-related reliability information from customers' online reviews, and utilize the retrieved information to conduct a reliability analysis. Fig. 1 illustrates the input/output and the procedures in each step of the method. The proposed framework has four steps:

Step 1: Filter and extract. The initial inputs of the proposed method are the customer review and the sales data, including the text of the review and the corresponding time of both the review and purchase. To extract the failure information from online reviews, we analyze the review using text mining and determine whether each review reports a product failure.
Step 2: Correction. The failure details obtained in the review analysis are inaccurate because they display the review time instead of failure time, and the review data are biased according to the customers' experience and understanding of the product. Therefore, we correct the failure information based on three assumptions on the review behavior of the customer: 1. Accuracy: the product failures are reported impartially and described accurately; 2. Equal deviation: deviations in the failure details are stable among similar products; 3. Timeliness: reviews are strongly associated with the failure time.
Step 3: Estimation. When the corrected failure information is obtained, we can observe a censored sample under the test condition, which is completely consistent with the actual life profile. Then, we estimate the reliability parameters using maximum likelihood estimation (MLE), and the reliability of the product is quantified using comparable parameters, i.e., the mean time between failure (MTBF) and failure rate.
Step 4: Analysis. Finally, with the estimated life parameters of the products, a comparative analysis is performed to evaluate the reliability of the products. The weak link, measures of improvement, and failure constitution of the products are obtained by making a comparison of life parameters. Uncertainty analysis is conducted to find the source of and solutions to the uncertainty introduced by the review analysis.

4. Failure information extraction from online reviews

This Section introduces the process of failure information extraction, in which dictionaries are constructed and used to recognize the failure-related information from the text. The procedure of information extraction addresses the grammatical structure of the text, and determines whether each review reports a product failure.

4.1. Failure dictionary construction

The failure dictionary is composed of three parts. The first part, which is called a context-irrelevant failure dictionary, covers the general keywords describing the failure of all types of products. The second part is the product-specific failure dictionary, which is provided by industry experts. The product-specific keywords are more precise, describing the failure of specific parts of the product. The third part is the corpus-specific failure dictionary, which is obtained using the tool word2vec [35] based on the above two sets of keywords. In word2vec, the semantic distance between each word or group of words in the review and the existing failure-related vocabulary is calculated based on the similarity, and a threshold of the similarity needs to be assigned to make the new keywords appropriate in screening the failures from the online reviews. There are two sources of failure-related vocabulary: one is to adopt FMEA, which is a structural method employed to recognize and classify all dominant or potential failures; the other source is the official documentation on the product warranty issued by the relevant regulator, which specifies the definition of the product failure and conditions of degraded performance applicable to repair and return policy. With the product structure, function, and failure analysis, the failures of a product are categorized by the component, failure keywords are associated with the failure symptoms and components. To eliminate the one-to-many correspondence between failure keywords and components, we propose a principle in the construction of failure dictionary, that the integrality of the components should be kept in the dismantlement of the product, so as to merge the components that cause the same failure, and try to make sure that each failure keyword corresponds to only one component.

(1) Context-irrelevant failure dictionary

For ordinary customers who are not experts, some universal words are used to describe a product failure, e.g., "it doesn't work!" and "it's out of order." Therefore, the general words and phrases that report the product failure from the user's perspective, while not addressing any specific symptoms or components of the failure, are included in the context-irrelevant failure dictionary. We summarized the phrases expressing the general complaints about the product failure by scanning the review texts of the products to be investigated. To obtain a complete context-irrelevant failure dictionary, seminal keywords from FMEA records [36,37] are taken as the starting point, and by scanning the reviews, the validity of these words as context-irrelevant keywords is ensured. We propose that 10% of all available reviews are inspected manually to ensure the completeness and dependability of the dictionary.

(2) Product-specific failure dictionary

Besides the general keywords about failure, some complaints about a specific part of a product can also be found from the online reviews,

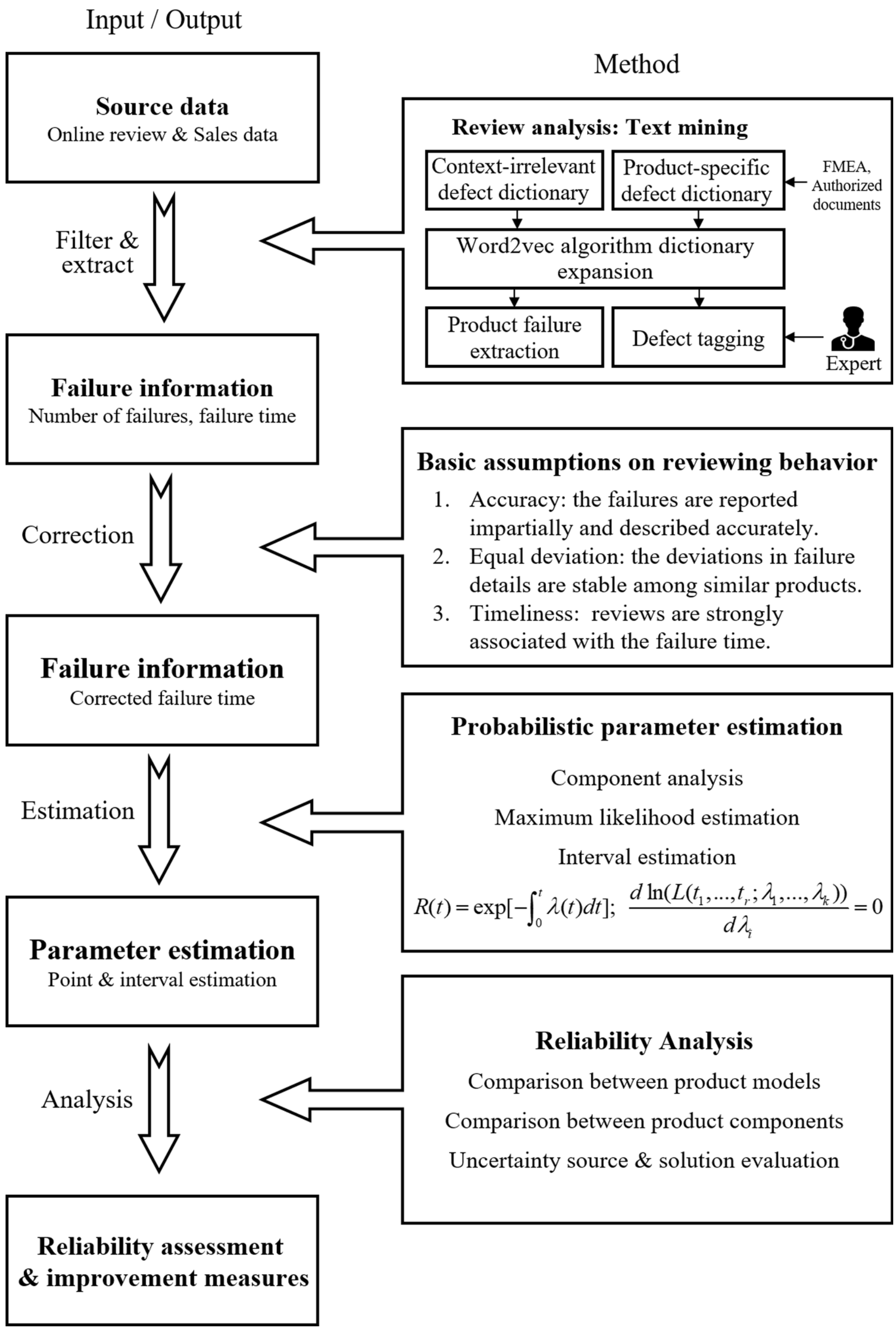

Fig. 1. Framework of the customer knowledge-based reliability analysis.

like "the screen is too bright," "the brake is too sensitive," and "the battery died." In order to extract this type of failure information, experts should provide some keywords which are capable of describing the symptoms of a failure specific to one product. The failure entities will help to turn the text of reviews into a structured dataset for further analysis. As one product usually has multiple models in the market simultaneously, we add the model information to the entity to improve the accuracy. The component here refers to the product feature that customers are discussing, and the "symptom" is the keyword or phrase describing the problems of a product. In building a product-specific failure dictionary, we propose the use of multiple text sources such as the product FMEA, review texts, and product troubleshooting information.

$$\text{Defect Entities:}\{\text{Product}(\textit{model}),\ \text{Component},\ \text{Symptom}\} \tag{1}$$

(3) Corpus-specific failure dictionary

As there are many synonyms, acronyms, and Internet slang in online reviews, we adopt Word2vec, which is an algorithm proposed by Google [35], to extract as much information about failure as possible. Word2vec is an open-source tool that implements the representation of large-scale real-valued vectors with rich semantic information by training large-scale corpora. It simplifies the processing of text content into vector operations, calculates the similarity in vector space, and uses the similarity of space vectors to represent the similarity of text semantics.

We applied word2vec by taking reviews as the input and producing a model file as the output. Based on the above two types of failure keywords, the Word2vec constructs a vocabulary from the training file, and provides high-dimensional vector representations of words based on pre-specified parameters. By calculating the similarity between the keywords in the failure dictionary and the words in the review, we are able to measure the semantic distance between the failure keywords and the review text. By roughly checking the output of new keywords found by the word2vec tool, we set the threshold as 0.98 and obtained new keywords to construct the corpus-specific failure dictionary, and we depicted product failures in the customers' innovative way.

4.2. Failure information extraction

Note that reviews containing the keywords in the dictionaries may not necessarily contain a product failure, and the influence of linguistic phenomenon, e.g. the structure, style, and sentiment of the language, should be avoided. Using the dictionary developed in Section 4.1, lexicon-based failure information is obtained from the online reviews with the following steps.

Step 1: Word segmentation.

For Chinese online reviews, word segmentation is the basis for the word-level analysis. In this study, the method of conditional random fields (CRFs) [38] is adopted to divide the sentence of the online reviews into words. A CRF is a probability statistical model for marking and decomposing serialized data based on the maximum entropy model. It is usually used to calculate the joint probability of the entire labeled sequence given that the observation sequences are labeled. The distribution condition attribute of the labeled sequence allows the CRFs to fit the actual data well, and in these data, the conditional probability of labeled sequences relies on features that are interdependent and interacting, and indicates the importance of features by assigning different weights.

Step 2: Part-of-speech tagging.

The Part-of-Speech (POS) tagging aims to extract the product features, which determines the meaning of each word when the meaning of the word varies with its part-of-speech. Usually, special attention is given to nouns, verbs, and adjectives. There are two major methods for Chinese text POS tagging, namely the statistical-based method and the rule-based method. Statistical natural language processing is widely used in disambiguation and syntactic analysis. For a given input string, all possible POS strings are determined first, and the tagging with the highest score is selected as the best output. The most widely used methods are the hidden Markov model (HMM) method and the conditional random field (CRF) method. However, when the training corpus reaches a certain scale, it is difficult to increase the precision by expanding the data scale. We need to design a more precise feature template for POS tagging. The rule-based method of text POS tagging is a traditional approach, and the tagging result is directly affected by the accuracy of the pre-specified rule set. This method takes advantage of existing linguistic achievements and utilizes widely accepted rules. First, a dictionary is used to segment and label the corpus. Then, all possible POSs of the words are listed. According to the context information, the ambiguity is eliminated in combination with the rule base, and the most appropriate POS is retained. If the rules are too special for a certain context, the coverage of the rules will be greatly reduced. As a result, the effectiveness of the rules will also diminish significantly. In this study, we applied the method of POS tagging based on Standards of Chinese text POS tagging [39].

Step 3: Complaint marking and failure entity identification.

To distinguish the complaints from the positive comments, sentiment analysis is carried out based on an emotional vocabulary ontology database proposed by the Information Retrieval Laboratory of Dalian University of Technology [40]. Different from the failure dictionary we constructed in Section 4.1, this database focuses on sentiment analysis only, and it helps to measure the sentiment of each review by counting the words in the review that are also listed in the positive/negative dictionary in the database. By obtaining the negative reviews, we focus on the reviews that discuss the failures more efficiently. Once a keyword is matched, three words before the word are checked to see whether any of them are a negation, and the sign of the score of the sentiment word is changed if there is a negation. The sentiment of the sentence is the cumulative score of each word in the sentence. The reviews with a negative sentiment are saved for further analysis, which improves the effectiveness of failure extraction.

$$TermSentiment_i = \begin{cases} 1, \\ \text{if } term\ \ i\ \ \text{in}\{positive\}\ \text{ and } term_k \neq negation;\ k = \{i-1, i-2, i-3\}\ \text{ if } i \geq 3 \\ -1, \\ \text{if} term\ \ i\ \ \text{in}\{negative\}\ \text{ and } term_k \neq negation;\ k = \{i-1, i-2, i-3\}\ \text{ if } i \geq 3 \end{cases} \tag{2}$$

$$Sentiment\ \ \text{of}\ Sentence = \sum_{i=1}^{j} Term\ \ Sentiment_i \tag{3}$$

Based on the lexicon constructed in Section 4.1, defect entities are extracted from the negative reviews. By using POS tagging, each failure entity is turned into a triple, which includes the product model, failure symptoms, and possibly the specified components. Experts are invited to manually confirm and combine the failures. In order to further understand the customer's dissatisfaction and find potential omissions of failure entity, the negative reviews without extant defect entities are further explored by the experts to double check if more failures, especially those not in the failure dictionary yet, should be extracted. After this step, a data set is constructed for the subsequent analysis, in which all of the review items that report the product failure, symptoms, and the corresponding review with purchase time are included.

5. Reliability parameter estimation on uncertain failure information

5.1. Hypotheses on review data

Compared with the test-based reliability data, there remain some uncertainties in the failure information obtained from the above extracting process. Such uncertainty includes the tendency, promptness and impartiality of the customers to provide a review, which is caused by various factors, e.g., whether there is a product failure, the customer's experience with the product features. Specifically, the uncertainty brought by the review information is classified as aleatory and epistemic uncertainties [3,41]. Note that the tendency of customers to submit a review on a product is significantly influenced by customers' experience and perception of the product; thus, the customers who never provide a comment are regarded as censored data. To address such an uncertainty, we adopt a comparative analysis to eliminate the impact of the biased review information, instead of providing only the exact values of life estimations as the final result. Detailed discussions about the uncertainty and its influence, source, and possible solutions are presented in Section 6.3. To conduct a comparative analysis, we tentatively omit the uncertainty in review information, and there are three basic assumptions on the review data:

Assumption 1. Accuracy. We assume that the product failures are reported impartially and described accurately in the review, i.e., the review process and data are free from external intervention. The customers' subjectivity, tendency, and other mental factors are also proved to influence customers' tendency to provide reviews [42], resulting in biased reliability estimation. Thus, the estimation of the parameters in the proposed method is currently used to conduct a comparative analysis between the products. However, previous studies proposed quantified approaches to measure how sentiment [42,43], product involvement [44,45], and prices [46] influence the product reviews, which indicates the potential to improve the accuracy of review analysis.

Assumption 2. Equal deviation. We assume that the deviations of the review data at depicting the number of failures are equal in different product models, i.e., the review data from Products 1 and 2 will conceal (or exaggerate) approximately the same proportion of product failures. In the context of customer reviews, such deviations are inevitable because the failure data are collected from the customer descriptions, which are infected by the customer's perception. However, when the analysis is conducted with a high volume of review data, such deviations tends to be stable [12]. Therefore, we assume that the customers for different product models report product failures with a constant probability c. Assumption 2 is mathematically represented as

$$\left.\frac{r_{review}}{r}\right|_{\text{Product}i} = \left.\frac{r_{review}}{r}\right|_{\text{Product}j} \equiv c \tag{4}$$

where r represents the actual number of failure entities, and r_{review} represents the number of failures reported by the review data.

Assumption 3. Timeliness. The review time is usually associated with the failure time [47]. We assume that the time difference between review time and failure time is normally distributed. This can be mathematically represented as

$$t = t_{rev} - t_{lag} \tag{5}$$

$$t_{lag} \sim N(\mu, \sigma^2) \tag{6}$$

where t represents the failure time of the product, t_{rev} represents the review time, and t_{lag} represents the time between the review time and failure time. Compared to the variance of failure time, the difference between t_{lag} for each review is relatively small ($<24\ h$), and is thus omitted in the reliability analysis, i.e., $\mu = 0$.

The above assumptions provide approaches to utilize the review information as failure data in the reliability analysis. On the one hand, the uncertainties introduced by the review analysis can be partially omitted in terms of the comparative analysis. On the other hand, previous studies [14,42–45] have provided insight into the elimination of these uncertainties, which is not an emphasis in this framework study on reliability analysis.

5.2. Parameter estimation and goodness-of-fit test

Since the emergence of statistical quality management, the parameter estimation has been a commonly used tool to determine and compare the reliability of a product. In the reliability analysis of mobile phones, we dismantle the product into components and analyze the failure data respectively. Because the components vary with respect to their material and structure, they follow distinct failure distributions. Therefore, we adopt different distributions in the analysis, namely the exponential distribution and Weibull distribution.

5.2.1. Exponential distribution

Way et al. [48] established that electronic devices with a lack of wear properties and without infant mortality follow an exponential distribution. We obtain the MLE of θ:

$$\hat{\theta} = \frac{\sum_{i=1}^{r} t_i + (n-r)t_c}{r} \tag{7}$$

where $\hat{\theta}$ represents the MLE of the MTBF of the product, n represents the number of all sold products with or without a review, t_i represents the failure time of the i^{th} failure product, and t_c represents the censoring time.

For an exponential distributed censored sample, we derive the interval estimation based on the following χ^2 distributed pivotal quantity:

$$H = \frac{2[\sum_{i=1}^{r} t_i + (n-r)t_c]}{\theta} \sim \chi^2 \tag{8}$$

According to the χ^2 distribution, we obtain the confidence interval of parameter θ:

$$\hat{\theta}_U = \frac{2[\sum_{i=1}^{r} t_i + (n-r)t_c]}{\chi^2_{\alpha/2}(2r)}$$
$$\hat{\theta}_L = \frac{2[\sum_{i=1}^{r} t_i + (n-r)t_c]}{\chi^2_{1-\alpha/2}(2r)} \tag{9}$$

Given a confidential level $\alpha = 0.05$, we can obtain the confidential interval of the estimated parameter.

To conduct a goodness-of-fit test on an exponential distributed sample, we suggest the F-test proposed by Zhao et al. [49]:

Let $r_1 = r_2 = 0.5r$ if $r = 2r_1$, and $r_2 = r_1 + 1$ if $r = 2r_1 + 1$. For an appropriate quantity φ,

$$\varphi = \frac{r_2 \sum_{i=1}^{r_1} (n-i+1)(t_i - t_{i-1})}{r_1 \sum_{i=r_1}^{r} (n-i+1)(t_i - t_{i-1})} \tag{10}$$

If φ satisfies

$$1 < \varphi < F_{1-\alpha/2}(2r_1, 2r_2), \text{or} 1 < \frac{1}{\varphi} < F_{1-\alpha/2}(2r_2, 2r_1) \tag{11}$$

Then, the hypothesis of the exponential distribution is accepted; otherwise, the hypothesis is rejected.

5.2.2. Weibull distribution

The failures of mechanical components, batteries, and fatigue materials follow a failure distribution in which the failure rate λ changes over time; thus, we adopt the Weibull distribution to estimate their reliability parameters. The MLE of parameters is given by the following transcendental equation set:

$$\begin{cases} \frac{d\ln(L(x_1,x_2,\ldots,x_r;m,\eta))}{dm} = \sum_{i=1}^{r}\left[\frac{1}{m} + \ln t_i - \ln\eta - \left(\frac{t_i}{\eta}\right)^m \ln\frac{t_i}{\eta}\right] - (n-r) \\ \qquad \left(\frac{t_c}{\eta}\right)^m \ln\left(\frac{t_c}{\eta}\right) = 0 \\ \frac{d\ln(L(x_1,x_2,\ldots,x_r;\mu,\eta))}{d\eta} = -\frac{mr}{\eta} + \frac{m}{\eta}\sum_{i=1}^{r}\left(\frac{t_i}{\eta}\right)^m + (n-r)\left(\frac{t_c}{\eta}\right)^m \frac{m}{\eta} = 0 \end{cases} \tag{12}$$

where m is the shape parameter, η is the scale parameter, and L represents the likelihood function of m and η.

To obtain the confidential interval of m and η, we adopt the interval estimation based on asymptotic normality:

$$[\hat{m} + u_{\alpha/2}\hat{\sigma}_m, \hat{m} + u_{1-\alpha/2}\hat{\sigma}_m], [\hat{\eta} + u_{\alpha/2}\hat{\sigma}_\eta, \hat{\eta} + u_{1-\alpha/2}\hat{\sigma}_\eta] \tag{13}$$

where

$$\hat{\sigma}_m^2 = Var(\hat{m}) = \frac{I_{\eta\eta}}{(I_{mm}I_{\eta\eta} - I_{m\eta}^2)}$$
$$\hat{\sigma}_\eta^2 = Var(\hat{\eta}) = \frac{I_{mm}}{(I_{mm}I_{\eta\eta} - I_{m\eta}^2)} \tag{14}$$

$$I_{mm} = -\frac{\partial^2 \ln L}{\partial^2 m^2}\bigg|_{\hat{m},\hat{\eta}},$$
$$I_{m\eta} = -\frac{\partial^2 \ln L}{\partial m \partial \eta}\bigg|_{\hat{m},\hat{\eta}},$$
$$I_{\eta\eta} = -\frac{\partial^2 \ln L}{\partial^2 \eta^2}\bigg|_{\hat{m},\hat{\eta}} \tag{15}$$

where $u_{\alpha/2}$ denotes the $\alpha/2$ quantile of the standard normal distribution; $Var(\hat{m})$ and $Cov(\hat{m})$ denote the asymptotic variance and covariance of m, respectively.

To conduct the goodness-of-fit test on a Weibull distributed sample, we suggest to employ the χ^2 test and determine the appropriate quantity B^2/c, which follows the χ^2 distribution [50]:

$$B^2/c = \frac{2(r-1)\lg[\sum_{i=1}^{r-1} V_i/(r-1)] - 2\sum_{i=1}^{r-1}\lg V_i}{1 + \frac{r}{6(r-1)}} \tag{16}$$

where

$$V_i = (r-i)(\ln t_{r-i+1} - \ln t_{r-i}) \tag{17}$$

If B^2/c satisfies

$$\chi^2_{\alpha/2}(r-2) < B^2/c < \chi^2_{1-\alpha/2}(r-2) \tag{18}$$

Then, the hypothesis of the Weibull distribution is accepted; otherwise, the hypothesis is rejected.

6. Case study

In this Section, we conduct a case study to validate the proposed method. The case study is conducted on a data set of the online review data collected from an e-commerce website. The data set contains 2602 records of customers' review from two smart phone models (1650 reviews for Product 1, whose sales volume was 10,481; and 952 reviews for Product 2, whose sales volume was 5376), which were generated during a specific date interval. The data of products that customers never provide a review are treated as censored data. We conducted the failure information extraction and reliability analysis on the two models, and made a comparative analysis based on the results. For reasons of confidentiality, the name of the retailer and the models of the products are concealed in this manuscript.

6.1. Failure information extraction

Following the steps described in Section 4.1, the failure dictionary is composed of three parts. Table 1 shows the context-irrelevant dictionary of failure of electronic products. For mobile phone products, the keywords in the failure-related dictionary (Tables 1 and 2) are extracted and expanded with reference to the warranty policies issued by the manufacturers and the official documents on the product warranty issued by relevant regulator [51].

In the interviews with three experts in the mobile phone industry, we obtain a list of keywords to express the failure in the smart phone industry. Table 2 shows part of the keywords in the failure dictionary of smartphone.

By employing the word2vec tool in the reviews with a similarity threshold 0.98, we obtain the keywords in Table 3 similar to the keywords in Tables 1 and 2.

By combining the keywords in Tables 1–3, we obtain an integrated dictionary of product failures, with a total of 124 keywords. Then, the sentiment analysis is conducted after the word segmentation and POS tagging. Complaints are screened out and saved for further analysis. Table 4 shows some of the complaints we received from the reviews. The failures of mobile phone are attributed to nine main components, and the components on the mainboard are regarded as part of the mainboard, so as to keep the one-to-one correspondence between failure keywords and failure components.

Then, a lexicon-based method is employed to extract the failure entities from the complaints, and experts are invited to manually confirm and combine the failures. The experts removed 30 (10.6%) and 18 (8.3%) records from product 1 and 2, respectively, because those reviews actually did not report any failure. These misclassifications can be attributed to three reasons: 1. The review contains a failure keyword, and negative sentiment is detected, but the keyword is used to describe non-product issues. Such misclassifications usually happen when the text contains anaphora [52]; 2. The review reports a product failure but the description of the failure is a quotation from other people or is describing other products for comparison. Semantic analysis, especially the recognition algorithm for the quotations and comparisons, is the possible solution to such misclassification [53]; 3. Some keywords detected in the reviews are actually typos. Table 5 lists the examples of failure entities after confirmation.

Besides, we propose confusion matrix to examine the accuracy of the recognition. A sample containing 100 original reviews was manually analyzed by three experts in mobile phone design and manufacturing. Table 6 exhibits the confusion matrix, which indicates that the proposed extraction algorithm is sensitive to the failed products (false negative rate = 0%), while a small proportion of non-failed products are detected as failed ones (false positive rate = 4%). Table 7 exhibits the three misclassified records in Table 6.

6.2. Reliability estimation and analysis

6.2.1. Failure summary and distribution selection

With the above process, we determined the review time and the corresponding failure component of each malfunctioning product. The consumption of each product can be deemed as a test, and the results of the tests are contained in the customer's review. From the perspective of the reliability test, the failure data in the review forms a right-censored sample, in which the censoring point is the time at which the last review is submitted. In the dataset we used, each customer is assigned an equal deadline within which to make the review, i.e., 180 days, before which the customers have multiple chances to submit reviews. Considering that the maximum length of time between the review and purchase in the data set is 178 days (4283 h), which approximates the deadline, we determine the censoring point as 180 days (4320 h). In this case, we only collect the data of products sold before 180 days ago, so that all products have reached the censoring point, regardless of whether the customer provided a review. Meanwhile, for customers, any malfunctioning component will result in a product failure; therefore, the mobile phone product can be regarded as a series system. In this series system, different components follow different probabilistic distributions; therefore, we classify the failure of the mobile phone based on each component, and we consider their failures respectively.

Fig. 2 provides an overview of the collected failure data, which simply summarizes the number of failures caused by different components. As shown in Fig. 2, the mainboard accounts for the most failures in both products, while the compositions of other component failures vary between the two product models. Besides, the compositions of failure modes relevant to the mainboard are also different. In Fig. 3, a histogram analysis is used to determine the failure time distribution of each component.

Table 1
Context-irrelevant dictionary of failure.

Keyword	Description
Something wrong with	Abnormal function
Is broken	Damaged or not working correctly
Malfunction	Abnormal function
Is not working	Abnormal function
…(8 items)	…

Table 2
Smart phone-specific dictionary of failure.

Keyword	Remark	Failure component	Failure mode
Frequent restart	Automatically restart frequently	Mainboard	Chip failure
Black line	Defective pixels cluster	LCD	Defective pixel
Cannot take photos	Camera failing to take photo	Camera	CMOS sensor failure
Cannot adjust volume	Volume key failure	Keyboard	Keyboard failure
Cannot switch application	Operation system malfunction	Software	Software function disorder
…(99 items)	…	…	…

Table 3
Product-specific dictionary expansion.

Keyword in Tables 1 and 2	Corpus-specific keyword	Similarity
Noises	Cracking voice	0.99556
Signal on and off	Flash screen	0.99568
Loose rare cover	Baggy	0.99678
Short standby time	Not sufficient battery	0.98300
…(17 items)	…	…

Table 4
Exemplary complaints from reviews.

Date	Review
20xx/8/14 13:08:00	'The packaging sucks!'
20xx/8/12 9:00:00	'The battery is not durable!'
20xx/8/10 19:30:00	'Always rebooting by itself.'
…	…

In Fig. 3, we illustrate the failure statistics by classifying the failures based on different components. As shown in Fig. 3(a), the overall failure percentages of the two products are 2.4% (Product 1) and 3.7% (Product 2), respectively. Because some components expose few failures, in Fig. 3(f), other failures include the failures of battery, shell, speaker, earpiece, and microphone. From the above failure statistics, we can observe the distribution of the failure data and find that failures of different components follow different failure probability distributions. Combining the observed distribution from the above data with the standard document [51], Table 8 exhibits the assumed failure distribution of each product entity (whole system, specific components, and other components) in the following analysis. Since the Weibull distribution is not additive (i.e., a series system consisting of different Weibull distributed components does not follow Weibull distribution), the system failure is assumed to follow exponential distribution, so as to simplify the parameters in the estimation.

6.2.2. Estimations of reliability parameters and MTBFs

When the failure distribution is preliminarily determined, the parameters can be estimated using MLE and interval estimation. With the estimating approach introduced in Section 5.2, the estimated parameters and the corresponding MTBF can be obtained, as shown in Table 9. In the goodness-of-fit test, the LCD of Product 1 is rejected for the Weibull distribution hypothesis because the number of failures is 5, which is too small for the proposed test, as proposed in [50]. It can also be proved that the simplification of the system failure distribution does not significantly impacts the estimation, because the estimated reliability of LCD is significantly better than other components, and the distribution of software failure is close to exponential distribution (m is close to 1).

Table 6
Confusion matrix of the proposed failure extraction method.

		Actual class Failure	No failure
Predicted class	Failure	18	3
	No failure	0	79

6.2.3. Comparison between products on component failure contribution

As the uncertainty introduced by the review information affects the accuracy of the estimated value of the parameters and MTBF, a comparison between similar products is an important approach that we propose in the review-based analysis. When the failure distribution and parameters of all the components are determined, we can depict the contribution of each failure component to the gross failure with the precise estimated failure distribution. As introduced above, the mobile phone product can be seen as a series system. Therefore, the product reliability depends on the reliability of each component. This can be mathematically represented as

$$R_s(t) = R_{c1}(t)\cdot R_{c2}(t)\cdot\ldots\cdot R_{ck}(t) \tag{19}$$

where $R_s(t)$ and $R_{ci}(t)$ represent the reliability of the system and component i, respectively.

We use the following equation to depict the contribution of the reliability of component i to the product reliability:

$$\mathrm{C}_i = \int_0^\infty \frac{R_s(t)}{R_{ci}(t)}dt = \int_0^\infty \frac{R_{c1}(t)\cdot R_{c2}(t)\cdot\ldots\cdot R_{ck}(t)}{R_{ci}(t)}dt \tag{20}$$

Summarizing the contribution C_i of all components to the gross failure, Fig. 4 illustrates a more accurate composition of the failure from the perspective of failure distribution, in which not only the number of failures are considered, but the time of each failure is also utilized to achieve more preciseness in assessing the contribution of each component. Being different from the results in Fig. 2, Fig. 4 reveals that "others" and "software failure" contribute the greatest to the product failure in Products 1 and 2, respectively, while LCD failure has a much less influence on the reliability of both products.

Furthermore, in Fig. 5, components of both products are compared by the estimated failure rates. To examine the influence of the uncertainty of the estimated parameters, the confidence intervals, as introduced in Section 6.2.2, are also used in the comparison. Because the failure rates of each component are of the order of different

Table 5
Failure information from reviews.

Date of purchase	Time of review	Time to failure (h)	Review text	Component of failure	Type of failure
20xx/8/6	20xx/8/14 16:45	196.75	'The new cell phone frequently restarts randomly, at least twice a day.'	Mainboard	Instable operating
20xx/8/5	20xx/8/13 01:33	181.55	'Sent to AS service center for testing. Conclusion: dead pixel!'	LCD	Dead pixel
20xx/5/10	20xx/5/14 13:32	97.53	'The front camera is not working.'	Camera	CMOS sensor failure
…	…	…	…	…	…

Table 7
Misclassified samples, keywords and reasons.

Misclassified review text	Keywords	Reasons of misclassification
"The website showed that I had a 20-yuan coupon, but it doesn't work on the App on my phone."	Does not work	It is the coupon that does not work, not the product.
"The SIM slot is too loose so the phone does not identify my old SIM card used for many years. But it works after I use a new SIM card, so it is OK."	Does not identify	The malfunction of the product is caused by non-product reasons.
"Folks tell me that the keyboard is difficult to use, but I don't think so. But I have to complain about the delivery of the platform. The response is too slow!"	Slow response	The detected keyword describes the service, not the product.

magnitudes, the estimations are plotted on a logarithmic scale, where a higher value represents a higher failure rate. From Fig. 5, we can see that the intervals of the average failure rate $\bar{\lambda}$ overlap in both products, especially the intervals of LCD and software. For other components, the overlap is insignificant. In addition, the lengths of the intervals are different in each component, even between components that have approximate point estimations, because the data sample and the information entropy vary with the components and product models.

6.3. Result evaluation and uncertainty analysis

The above statistics show interesting results. First, Figs. 2 and 4 show distinct results with respect to the constitution of the failures. Fig. 2 reveals that the mainboard accounts for 32.9% and 23.5% of failures for Products 1 and 2, respectively, as opposed to 15.9% and 11.3% in Fig. 4, which is deduced from the same sample. The discrepancy could be attributed to the consideration of the failure distribution. Specifically, the results in Fig. 2 are obtained by summarizing the number of occurrences of the failures, while Fig. 4 shows the failure distribution estimated using both the number and time of failures. When the failure time of each entity is incorporated into the calculation, the estimated parameters approximates the exact value. It can therefore be assumed that the missing failure information in the reliability analysis will reduce the preciseness of the analysis and provide biased results. Meanwhile, more precise and impartial customer feedback can help to provide more precise and comprehensive failure data, improving the accuracy of the reliability analysis.

Secondly, the calculated failure rate (or MTBF) varies widely with different components, e.g., the estimated MTBFs of LCD and software differ by two orders of magnitude. With respect to the estimation process, the reason for the discrepancy is that the number of reported LCD failures is substantially smaller than the number of instances of mainboard failure. Specifically, only five (Product 1) and 22 (Product 2) customers reported LCD failures, while 82 (Product 1) and 47 (Product 2) review records illustrate product malfunctions associated with the mainboard. Besides, one unanticipated finding is that the estimated MTBF of the LCD is abnormally large, which is evidently not in line with the actual condition of the product. Therefore, it is important to collect and analyze sufficient product failure data during the product improvement and reliability analysis stages.

Another significant phenomenon is that although different components may have distinct reliabilities, the reliability of the same component for different product models is approximate. This may indicate that the reliability of components tends to be subject to their inherent characteristics, rather than the material or manufacturing process of the component, which is in accordance with the theoretical depiction in the previous research [54]. In the case study, the reliability of all components and the system of Product 1 are all inferior to that of Product 2, and the estimated system MTBF is close to the estimated MTBF of software, which indicates that the general reliability of a product depends on the component which possesses the lowest reliability, although the reliability of the product is a combination of the reliabilities of all individual components. Therefore, when developing high-reliability systems, designers should focus on the product's weak links because it is the most economical and efficient approach to improve the system reliability.

The above results and evaluations also indicate that our hypotheses on the data introduce some uncertainties to the results. To address these uncertainties, Fig. 6 illustrates the categorization, sources, and solutions to these uncertainties. When utilizing the review data in reliability analysis, uncertainties arise mainly from two sources: cognitive bias and review behavior of the customers, and the performance of the text mining method. If an analyst requires the precise values of parameters, there are certain approaches to eliminate the uncertainties, e.g., encouraging customers to post reviews or analyzing the operating time on a component level.

The uncertainty analysis in Fig. 6 also indicates that the comparison between different products is an effective approach to eliminating the uncertainty introduced by the collection and processing of review information. In conclusion, the software is the weak link of both product models in the case study, which causes the most reliability loss. Besides, the reliability of different components varies with the type of

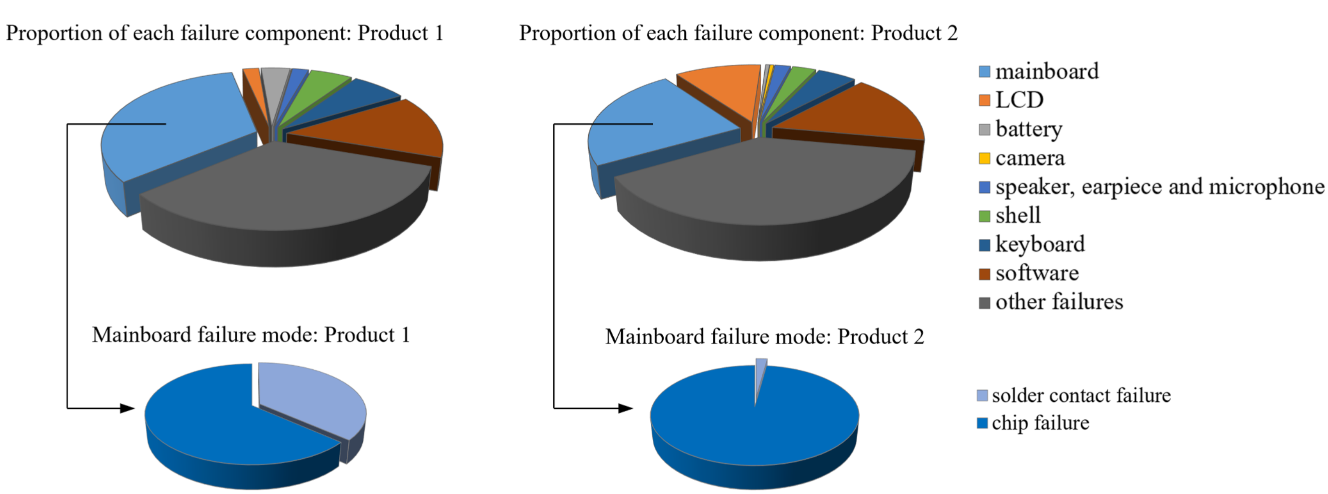

Fig. 2. Proportion of each failure component and the corresponding failure mode.

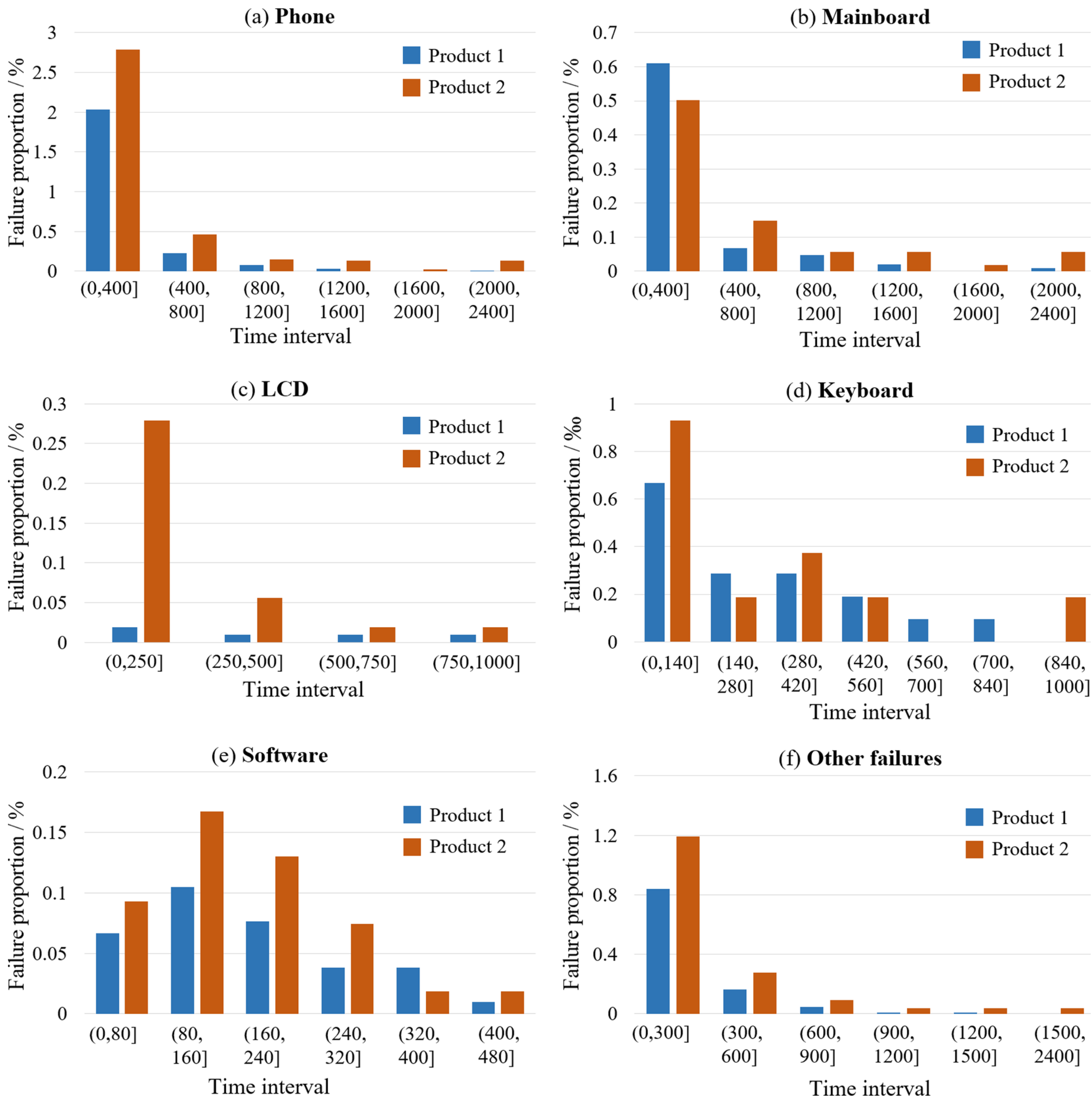

Fig. 3. Failure statistics of different component.
†According to an industry report issued in 2018 by Blancco [56], which is a data security company, the failure probability of mobile phone products varies from 1% to 34%.

Table 8
Assumed failure distributions of each product entities.

Product entity	Failure distribution
Phone (system)	Exponential
Mainboard	Exponential
LCD	Weibull
Keyboard	Exponential
Software	Weibull
Others	Exponential

component rather than the product model, thus improving the working principle of each component may produce more reliability increase than improving the manufacturing process. Further, enhancing the acquisition of customer feedback is necessary, not only because a better understanding of customers' perception is beneficial to the product improvement, but also because more failure information can improve the preciseness of reliability assessment. In the design and improvement of mobile phones, more attention should be given to the software because it contributes the greatest to the product failure.

7. Discussion

In this study, we proposed a framework to analyze the product reliability by utilizing the information contained in customers' online reviews. The case study on a comparison between two similar mobile phone models verifies the significance and feasibility of the proposed method, from which we know the review information is not only important to the product improvement, but it also serves to supplement

Table 9
Estimated parameters and MTBF of each product entity and model.

Product entities	Product model (1 or 2)	Number of failure records	Distribution and parameters			Point estimation of MTBF ($\times 10^5$h)	Interval estimation of MTBF ($\times 10^5$h) (at $\alpha = 0.05$)		Goodness-of-fit test (at $\alpha = 0.05$)	
			Exponential	Weibull			Lower limit	Upper limit	Observation (control)	Result
			θ ($\times 10^5$)	η ($\times 10^5$)	m					
Phone	1	252	1.76	–	–	1.76	1.56	2.00	0.059 (0.781)	Accept
	2	200	1.12	–	–	1.12	0.981	1.30	0.0809 (0.757)	Accept
Mainboard	1	83	5.42	–	–	5.42	4.42	6.80	0.052 (0.648)	Accept
	2	47	4.90	–	–	4.90	3.76	6.68	0.099 (0.557)	Accept
LCD	1	5	–	502	0.705	631	479	812	0.16 (0.216, 9.35)	Reject
	2	22	–	245	0.544	425	306	764	9.58 (8.23, 31.5)	Accept
Keyboard	1	17	26.6	–	–	26.6	17.4	45.7	0.180 (0.368)	Accept
	2	10	23.2	–	–	23.2	13.6	48.3	0.141 (0.269)	Accept
Software	1	35	–	2.60	0.898	2.74	1.10	4.63	20.5 (19.0, 50.7)	Accept
	2	32	–	1.49	0.910	1.55	0.740	3.05	18.4 (13.8, 41.9)	Accept
Others	1	112	4.00	–	–	4.00	3.35	4.86	0.160 (0.651)	Accept
	2	89	2.51	–	–	2.51	2.07	3.12	0.198 (0.641)	Accept

Note: the exponential distribution hypothesis is accepted if the observed appropriate quantity is not greater than the control value; the Weibull distribution hypothesis is accepted if the observed appropriate quantity is in the control interval.

the test-based reliability analysis.

The case study on the proposed method reveals some measures to improve the acquisition of customer knowledge. Compared to the test-based reliability analysis, the failure data derived from the online review results in broader confidence intervals for the estimated lifetime, which indicates that for precise analytical processes, the present failure data fails to provide a sufficient test sample. Besides, possible sources of the uncertainty of the proposed estimation are the customer's perception of the product, promptness to provide a review, inadequate data, etc. In order to address these uncertainties, recent studies on the customers' behavior in the B2C context provided possibilities to correct the presence of biased review data from a cognitive and behavioral perspective. However, when examining the failure data in the case study, we concluded that the total test sample in the case study is significant (number of test subjects ≈ 1000), while the observed failure times of the censored sample are short (for some components $t_r < 300\ h$). Therefore, if more failures can be reported, e.g., encouraging customers to voice their concerns and report the abnormalities, more information and knowledge related to the product failure will be obtained, by which more precise and accurate results of life estimation can be achieved.

Another approach to strengthen and extend the application of the proposed method is to combine the statistical techniques with qualitative methods. In product engineering, a focus group is an efficient yet low-cost survey method to collect user feedback, and can help to understand the motivations of users and their perceptions of product [55]. With the failure information, the proposed quantitative method provides a quick and automatic mapping of the failure distribution, and indicates which component possesses the most reliability issues. Meanwhile, focus groups can discuss the issues that are of concern, providing detailed suggestions about the reliability of the product, and possibly the corrective actions. While on the one hand, the focus group method can give experienced and targeted conclusions on the products, on the other hand, a statistical approach can help to compensate for the subjectivity and homogenization of group discussions. Therefore, a combination of qualitative and quantitative methods can improve the dependability of the results in reliability analysis.

In order to provide a more precise analytical result of the reliability, an emotion analysis of the customers' review behavior can be conducted. As introduced in the above uncertainty analysis, customers may have different promptness, experience, and subjectivity regarding the use and the failure of the products, which is reflected in the emotions in the review. Thus, an emotion analysis can provide a balanced review sample. Besides, a probabilistic mapping between failure mode and failure component can strengthen the correspondence between failure keywords and components, in which keywords may correspond to different components according to the probability. This probabilistic mapping can be constructed on the basis of the FMEA statistics and product diagnostics data, by which the reliability analysis is improved.

Besides the above improvements on the proposed method, this study will open practical and straightforward extensions. The proposed method can be applied to customer-orientated data in other forms. The warranty records, or third-party maintenance statistics are also valuable data sources for the customer knowledge-based reliability analysis. Similar to the information provided by online reviews, this closed-door information usually includes the failure type, failure symptom, and the failed components, but it is in a more concrete record rather than "hidden" in the text. With purposive collection and summary of the warranty or third-party maintenance data, the failure information constitutes a large test sample and the criterion of failure is 100% applied properly, thus providing significant information for the reliability analysis.

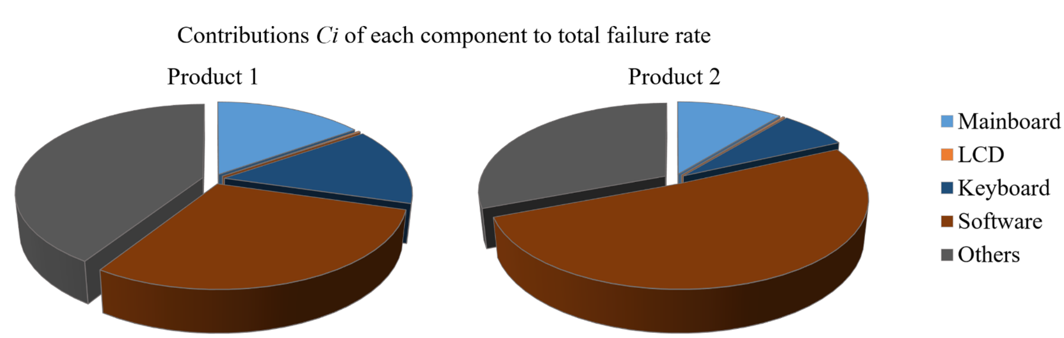

Fig. 4. Contributions C_i of each component to product failure.

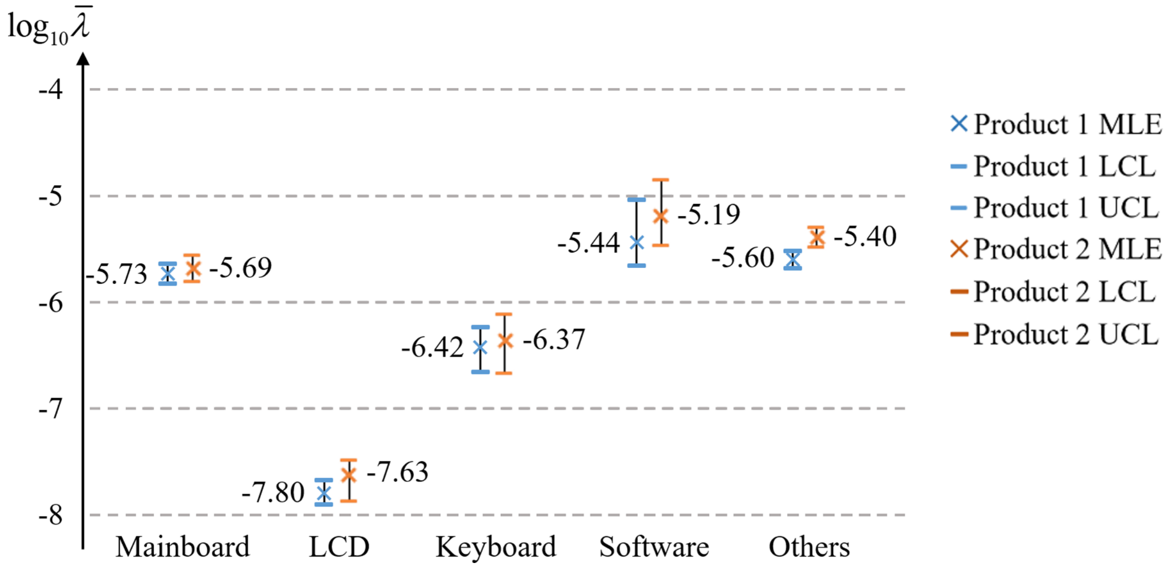

Fig 5. Estimated failure rate and confidential intervals of different components of Products 1 and 2.

CRediT authorship contribution statement

Xing Pan: Conceptualization, Supervision, Methodology. **Huixiong Wang:** Methodology, Formal analysis, Writing - review & editing. **Weijia You:** Methodology, Software, Formal analysis. **Manli Zhang:** Conceptualization, Validation. **Yuexiang Yang:** Conceptualization.

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.

Acknowledgement

The authors are grateful to the anonymous reviewers for their valuable comments which led to an improvement of this paper. This work is supported by the National Key Research and Development Program of China under Grant No. 2017YFF0206504, the National Natural Science Foundation of China under Grant No. 71571004, the Fundamental Research Funds for the Central Universities under Grant No. 2018RW14, the Special Scientific Research Project for Civil Aircraft of the Ministry of Industry and Information Technology, and Forestry Science and Technology Research Project under Grant No. 2019132707.

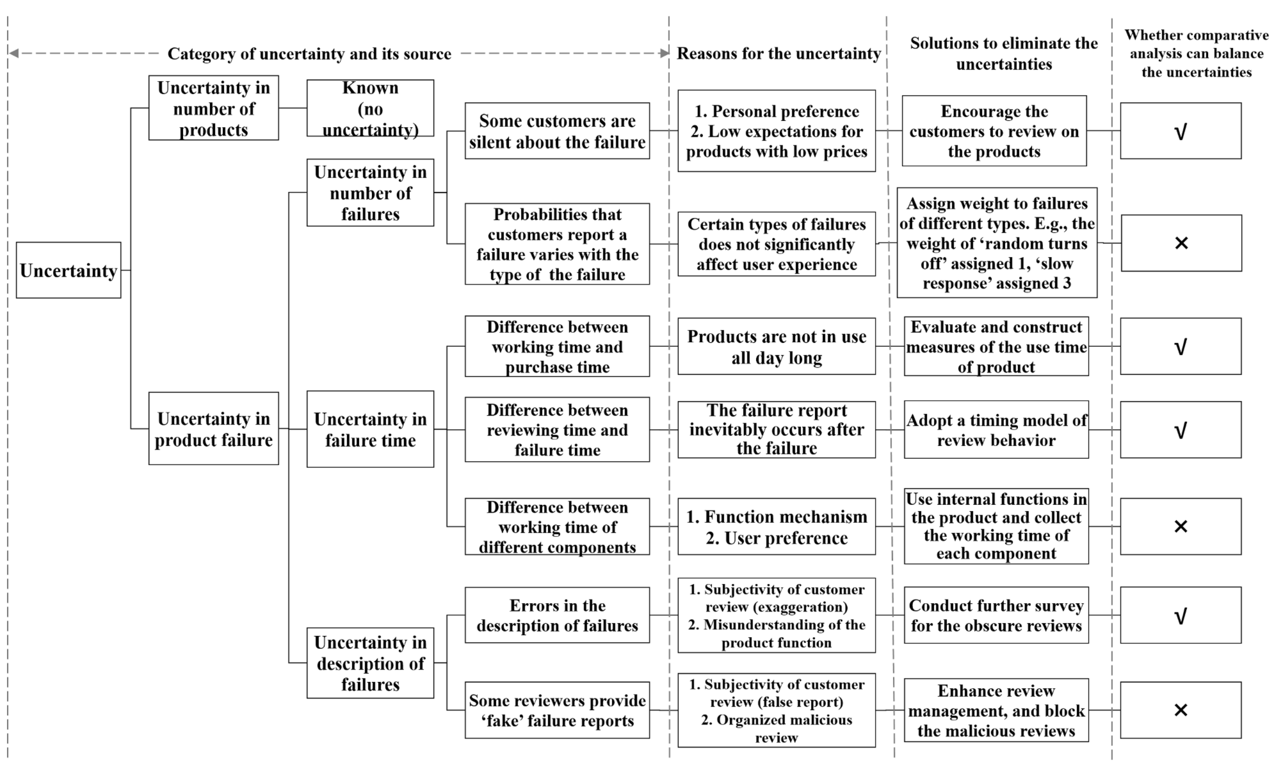

Fig. 6. Uncertainty analysis: category, source, and solutions.

References

[1] Goel A, Graves RT. Electronic system reliability: collating prediction models. IEEE Trans Device Mater Reliab 2006;6:258–65. https://doi.org/10.1109/TDMR.2006.876570.

[2] Saleh JH, Marais K. Highlights from the early (and pre-) history of reliability engineering. Reliab Eng Syst Saf 2006;91:249–56. https://doi.org/10.1016/j.ress.2005.01.003.

[3] Kang R, Zhang Q, Zeng Z, Zio E, Li X. Measuring reliability under epistemic uncertainty: review on non-probabilistic reliability metrics. Chinese J Aeronaut 2016;29:571–9. https://doi.org/10.1016/j.cja.2016.04.004.

[4] Yang L, Bernstein JB. Failure rate estimation of known failure mechanisms of electronic packages. Microelectron Reliab 2009;49:1563–72. https://doi.org/10.1016/j.microrel.2009.06.005.

[5] Roesch WJ, Brockett S. Field returns, a source of natural failure mechanisms. Microelectron Reliab 2007;47:1156–65. https://doi.org/10.1016/j.microrel.2007.02.004.

[6] Mourtzis D, Vlachou E, Zogopoulos V, Gupta RK, Belkadi F, Debbache A, et al. Customer feedback gathering and management tools for product-service system design. Procedia CIRP 2018;67:577–82. https://doi.org/10.1016/j.procir.2017.12.264.

[7] Antonitsin A, State V, Academy D, Science A. Statistical methods in reliability testing. Technometrics 2012;25:305–16.

[8] Joshi AW, Sharma S. Customer knowledge development: antecedents and impact on new product performance. J Mark 2004;68:47–59. https://doi.org/10.1509/jmkg.68.4.47.42722.

[9] Liao SH, Hsieh CL, Huang SP. Mining product maps for new product development. Expert Syst Appl 2008;34:50–62. https://doi.org/10.1016/j.eswa.2006.08.027.

[10] Abrahams AS, Jiao J, Wang GA, Fan W. Vehicle defect discovery from social media. Decis Support Syst 2012;54:87–97. https://doi.org/10.1016/j.dss.2012.04.005.

[11] Pavlou PA, Dimoka A. The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation. Inf Syst Res 2006;17:392–414. https://doi.org/10.1287/isre.1060.0106.

[12] Eslami SP, Ghasemaghaei M. Effects of online review positiveness and review score inconsistency on sales: a comparison by product involvement. J Retail Consum Serv 2018;45:74–80. https://doi.org/10.1016/j.jretconser.2018.08.003.

[13] You W, Xia M, Liu L, Liu D. Customer knowledge discovery from online reviews. Electron Mark 2012;22:131–42. https://doi.org/10.1007/s12525-012-0098-y.

[14] Li X, Wu C, Mai F. The effect of online reviews on product sales: a joint sentiment-topic analysis. Inf Manag 2019;56:172–84. https://doi.org/10.1016/j.im.2018.04.007.

[15] Zhang ZZ, Qin JQ, Di CC, Wang D. An introductory review of mechanical reliability testing. Appl Mech Mater 2014;541–542:1414–8. https://doi.org/10.4028/www.scientific.net/AMM.541-542.1414.

[16] Vintr Z. Rational design of conditions for reliability test. Proc 2009 8th Int Conf Reliab Maintainab Safety, ICRMS 2009 2009. p. 363–6. https://doi.org/10.1109/ICRMS.2009.5270173.

[17] Elsayed EA. Overview of reliability testing. IEEE Trans Reliab 2012;61:282–91. https://doi.org/10.1109/TR.2012.2194190.

[18] Wang X, Ye ZS, Hong YL, Tang LC. Analysis of field return data with failed-but-not-reported events. Technometrics 2018;60:90–100. https://doi.org/10.1080/00401706.2017.1292957.

[19] Tian J. Better reliability assessment and prediction through data clustering. IEEE Trans Softw Eng 2002;28:997–1007. https://doi.org/10.1109/TSE.2002.1041055.

[20] Altun M, Comert SV. A change-point based reliability prediction model using field return data. Reliab Eng Syst Saf 2016;156:175–84. https://doi.org/10.1016/j.ress.2016.07.024.

[21] Dai A, He Z, Liu Z, Yang D, He S. Field reliability modeling based on two-dimensional warranty data with censoring times. Qual Eng 2017;29:468–83. https://doi.org/10.1080/08982112.2017.1319955.

[22] Cheng Z. Reliability and safety data collection and application in shenzhou spaceship. Proc 2009 8th Int Conf Reliab Maintainab Safety, ICRMS 2009 2009. p. 516–8. https://doi.org/10.1109/ICRMS.2009.5270139.

[23] Wayland RE, Cole PM. Customer connections: new strategies for growth. Harvard Business Press; 1997.

[24] García-Murillo M, Annabi H. Customer knowledge management. J Oper Res Soc 2002;53:875–84. https://doi.org/10.1057/palgrave.jors.2601365.

[25] Abrahams AS, Fan W, Wang GA, Zhang Z, Jiao J. An integrated text analytic framework for product defect discovery. Prod Oper Manag 2015;24:975–90. https://doi.org/10.1111/poms.12303.

[26] Jin J, Ji P, Liu Y, Johnson Lim SC. Translating online customer opinions into engineering characteristics in QFD: a probabilistic language analysis approach. Eng Appl Artif Intell 2015;41:115–27. https://doi.org/10.1016/j.engappai.2015.02.006.

[27] Zheng H, Liu W, Xiao C. An activity-based defect management framework for product development. Comput Ind Eng 2018;118:202–9. https://doi.org/10.1016/j.cie.2018.02.027.

[28] Kirmani A, Rao AR. No pain, no gain: a critical review of the literature on signaling unobservable product quality. J Mark 2000;64:66–79. https://doi.org/10.1509/jmkg.64.2.66.18000.

[29] Liu Y, Jiang C, Zhao H. Using contextual features and multi-view ensemble learning in product defect identification from online discussion forums. Decis Support Syst 2018;105:1–12. https://doi.org/10.1016/j.dss.2017.10.009.

[30] Reyes A, Rosso P. Making objective decisions from subjective data: detecting irony in customer reviews. Decis Support Syst 2012;53:754–60. https://doi.org/10.1016/j.dss.2012.05.027.

[31] Abrahams AS, Jiao J, Fan W, Wang GA, Zhang Z. What's buzzing in the blizzard of buzz? automotive component isolation in social media postings. Decis Support Syst 2013;55:871–82. https://doi.org/10.1016/j.dss.2012.12.023.

[32] Hanchuan P, Fuhui L, Chris D. Feature selection based on mutual information : criteria of max-dependency, max-relevance and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27:1226–38. https://doi.org/10.1109/TPAMI.2005.159.

[33] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993.

[34] Dey L, Haque SM, Khurdiya A, Shroff G. Acquiring competitive intelligence from social media. Proc 2011 Jt Work Multiling OCR Anal Noisy Unstructured Text Data - MOCR_AND '11 2011. p. 1. https://doi.org/10.1145/2034617.2034621.

[35] Mikolov T. Google code archive - Word2vec2013. https://code.google.com/archive/p/word2vec/.

[36] Marques L.M.C. FMEA - Mobile Phone 2010:6–14. http://lrss.fri.uni-lj.si/sl/teaching/zzrs/lectures/2010_Marques.pdf.

[37] Bašić A, Popović P. Risk assessment of mobile phones using failure mode and effect analysis. Adv Eng Syst 2016:213–20. https://doi.org/10.15308/sinteza-2016-213-220.

[38] Lafferty JD, McCallum A, Pereira FCN. Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proc. Eighteenth Int. Conf. Mach. Learn. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; 2001. p. 282–9. https://doi.org/10.1109/ICIP.2012.6466940.

[39] Standards of Chinese text POS tagging. Inst Comput Linguist Peking Univ2017. http://icl.pku.edu.cn/icl_res/segtag98/catetkset.html(accessed November 30, 2017).

[40] Xu L, Lin H, Pan Y, Hui R, Chen J. Constructing the affective lexicon ontology. J China Soc Sci Tech Inf 2008;27:180–5.

[41] Der Kiureghian A, Ditlevsen O. Aleatory or epistemic? Does it matter? Struct Saf 2009;31:105–12. https://doi.org/10.1016/j.strusafe.2008.06.020.

[42] Picazo-Vela S, Chou SY, Melcher AJ, Pearson JM. Why provide an online review? An extended theory of planned behavior and the role of big-five personality traits. Comput Human Behav 2010;26:685–96. https://doi.org/10.1016/j.chb.2010.01.005.

[43] Li Y, Feng X, Zhang S. Detecting fake reviews utilizing semantic and emotion model. Proc - 2016 3rd Int Conf Inf Sci Control Eng ICISCE 2016 2016. p. 317–20. https://doi.org/10.1109/ICISCE.2016.77.

[44] Park D-H, Lee J, Han I. The effect of on-line consumer reviews on consumer purchasing intention: the moderating role of involvement. Int J Electron Commer 2007;11:125–48. https://doi.org/10.2753/JEC1086-4415110405.

[45] Thakur R. Customer engagement and online reviews. J Retail Consum Serv 2018;41:48–59. https://doi.org/10.1016/j.jretconser.2017.11.002.

[46] Noone BM, McGuire KA. Effects of price and user-generated content on consumers' prepurchase evaluations of variably priced services. J Hosp Tour Res 2014;38:562–81. https://doi.org/10.1177/1096348012461551.

[47] Li X, Hitt LM. Self-selection and information role of online product reviews. Inf Syst Res 2008;19:456–74. https://doi.org/10.1287/isre.1070.0154.

[48] Way K, Wei-Ting Kary C, Taeho K. Reliability, Yield, and Stress Burn-In: A Unified Approach for Microelectronics Systems Manufacturing and Software Development. Kluwar Science; 1998.

[49] Zhao Y. Data analysis of reliability. Beijing: National Defense Industry Press; 2011.

[50] Minges LM. Electronic materials handbook volume 1: packaging. Materials Park: ASM International; 1989.

[51] General administration of quality supervision. Regulations on the replace and repair duty of mobile phones product. Insp Quar People's Repub China; 2006http://zlgls.aqsiq.gov.cn/cpaqjcpzldb/sbzd/200610/t20061030_20848.htm (accessed April 10, 2019).

[52] Lee C, Jung S, Park CE. Anaphora resolution with pointer networks. Pattern Recognit Lett 2017;95:1–7. https://doi.org/10.1016/j.patrec.2017.05.015.

[53] Xu K, Liao SS, Li J, Song Y. Mining comparative opinions from customer reviews for competitive intelligence. Decis Support Syst 2011;50:743–54. https://doi.org/10.1016/j.dss.2010.08.021.

[54] Kharmanda G, El Hami A. Biomechanics: optimization, uncertainties and reliability. Biomech Optim Uncertainties Reliab 2017;5:121–52. https://doi.org/10.1002/9781119379126.

[55] Engelbrektsson P. Effects of product experience and product representations in focus group interviews. J Eng Des 2002;13:215–21. https://doi.org/10.1080/09544820110108917.

[56] Blancco. State of mobile device repair & security2018. https://download.blancco.com/download/en-rs-q4-2017-state-of-mobile-device-repair-and-security.pdf.