A small sample data-driven method
User needs elicitation from online reviews in new product iteration

Cong, Yangfan; Yu, Suihuai; Chu, Jianjie; Su, Zhaojing; Huang, Yuexin; Li, Feilong

DOI
10.1016/j.aei.2023.101953

Publication date
2023

Document Version
Final published version

Published in
Advanced Engineering Informatics

Advanced Engineering Informatics 56 (2023) 101953

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

A small sample data-driven method: User needs elicitation from online reviews in new product iteration

Yangfan Cong [a,*], Suihuai Yu [a,*], Jianjie Chu [a,*], Zhaojing Su [a], Yuexin Huang [a,b], Feilong Li [a]

[a] Key Laboratory of Ministry of Industrial Design and Ergonomics, Ministry of Industry and Information Technology, Northwestern Polytechnical University, Xi'an 710072, China

[b] School of Industrial Design Engineering, Delft University of Technology, Delft 2628CE, the Netherlands

ARTICLE INFO

Keywords:
Small sample data-driven
User needs elicitation
New product iteration
ERNIE
Improved SIFRank

ABSTRACT

Eliciting user needs from mass online reviews is playing a significant role in the product iteration process. Efficient user needs elicitation does achieve considerable benefits for maintaining higher competitiveness and a speedier lifecycle. However, there is inevitably an online review scarcity about new products due to the short time on the market and low buyer recognition compared with commonly used products. This paper proposes a small sample data-driven method for user needs elicitation from online reviews in new product iteration. In the first stage, a scraped initial online review dataset is pre-processed roughly to improve the data quality. And then, reviews are classified into multiple categories according to different topics using ERNIE. In the second stage, each topic-based dataset is reprocessed in detail. Thereafter, the key user needs set is determined and facilitated by extracting key product information phrases from every single dataset using improved SIFRank. Moreover, the case study of a smart cat feeder is carried out to demonstrate the feasibility and potential of the ERNIE-ISIFRank methodology. Finally, comparative experiments are conducted to verify the advantages of the proposed method which is primarily based on the pre-trained language model to enhance the deep understanding of the semantics of online reviews. The experimental results confirm that the proposed method can assist in identifying key user needs with high efficiency.

1. Introduction

Under the conditions of fierce competition in modern markets, new products are consequently hitting the market to meet diverse user needs [1]. The concept of new products in this paper is defined as a totally new type of product that has never existed before. If a new product is successful, competitors will quickly imitate it and directly shorten its lifecycle. Therefore, industrial companies always face ever-increasing challenges iterating on their products and maintaining high competitiveness [2]. Product iteration refers to the replacement of the original product with a new generation that has better quality, higher efficiency, and more excellent functions [3]. This is not only the pattern of product development but also an important countermeasure for enterprises in competition [4–6]. From the perspective of product lifecycle theory, a product will always withdraw from the market someday. And with the rapid advance of technology and living standards, the product lifecycle is getting shorter. Therefore, it objectively requires continuous product iteration to follow the market trends [7].

Due to the robust development of online shopping, an increasing number of people post reviews in cyberspace. These reviews contain enormous available information, which is of great use in understanding users' real-time opinions [8]. Online reviews suggest ideas for new product features. Therefore, they are a potentially rich and reliable source of identifying user needs for future product releases [9]. It has been currently recognized that analyzing online reviews is an effective way to reduce the gap between what is required by users and what can be provided by designers [10].

As information and communication technology is booming, Internet technology has been embedded in natural language processing (NLP) [11]. The involvement of online reviews transcends what a finite sample of statistical prediction and estimation can achieve [12]. By taking advantage of the collection, management, and extraction of review data, it is quite effective to quantify users' perceptual information and needs. This can serve as a way to promote product iteration and also provides new patterns for industrial design [13]. The innovative design pattern "product-review-product" uses online review data as the driving force

* Corresponding authors.
E-mail addresses: yangfan_cong@sina.com (Y. Cong), ysuihuai@vip.sina.com (S. Yu), cjj@nwpu.edu.cn (J. Chu).

https://doi.org/10.1016/j.aei.2023.101953
Received 28 November 2022; Received in revised form 16 February 2023; Accepted 18 March 2023
Available online 31 March 2023
1474-0346/

for product development. Since implicit user needs are hidden in online reviews, the data-driven method can effectively enable the elicitation of explicit user needs and subsequently drive a new round of product iteration. The new generation product can continuously obtain new feedback and efficiently realize the latent value of online reviews in a dynamic and sustainable iteration cycle, as shown in Fig. 1.

Take baby bottles as an illustrative example, a piece of review is “There is a weird smell of plastic inside, will it be toxic?”. It just reflects the user’s subjective feelings when using the product. The real needs that are hidden behind the semantic meaning are regarded as implicit needs. By comprehending the review text, the key user needs can be expressed as “Users need the material to be safe and baby-friendly”, which are regarded as explicit needs. Based on explicit needs, we can target the product features and update them.

Traditional product optimization primarily focuses on common products which have been iterated for many generations. Take the air-conditioner as an example, it has come into being and has been fully developed for many years. Especially for a well-known brand, the product owns a substantial buyer base as a result of its long existence and high public recognition. Therefore, previous studies could draw on a large number of online reviews. However, a new product’s online reviews are usually not as many as a mature product in the same span (e.g. one year) because of its low credibility and acceptance. Hence, we describe the number of online reviews of new products collected within no longer than one year as small-sample on the basis of the comparison with commonly used products. To our best knowledge, the existing methods usually can’t perform well enough in analyzing small-volume review data when extracting user needs.

To bridge those research gaps, the primary motivation of this paper is to develop a systematic small sample data-driven method which can be applied to user needs elicitation from online reviews. As we all know, user needs acquisition is the basis and prerequisite of the whole process of new product iteration. The method is expected to prompt the efficiency and quality of the iteration from the perspective of user needs elicitation. It is also hoped to further make up for the current ignorance of acquiring user needs effectively for new products. The method we proposed is a generic framework so it can be extended to other practices of new industrial products.

The following parts of the paper are organized as follows. Section 2 reviews the literature of related research. Section 3 proposes a user needs elicitation method for new product iteration based on a small sample of online reviews. Technically speaking, the method is primarily supported by the pre-trained language model. Section 4 gives an illustrative example of a smart cat feeder. It is used to validate the feasibility and advantages of the proposed method. Section 5 conducts comparative experiments and discusses the merits and limitations of the method. Section 6 concludes the main contributions of this paper and highlights future research.

2. Literature review

In recent years, researchers have shown a growing interest in extracting user needs from review data. A host of relevant works is implementing NLP techniques to address these questions. From the perspective of key technologies, we present a brief background on the use of online reviews for user needs elicitation and the associated challenges.

2.1. Traditional methods to identify user needs from online reviews

Various approaches have been proposed to extend traditional models, such as Quality Function Deployment (QFD) and KANO model. They rely on input data obtained from user reviews. Analyzing and processing these data can assist a lot in mining user needs in a timely manner.

QFD is commonly used in manufacturing industries to identify user needs [14]. Özdağoğlu Güzin [15] integrated QFD and topic modeling to facilitate analyzing online reviews to reveal true user needs for developing products. Latent Dirichlet Allocation (LDA) is used to group reviews as topics, and then QFD is employed to clarify these reviews.

KANO model is a useful tool for classifying and prioritizing user needs. Anna Martí Bigorra et al. [16] presented a methodology for autonomously classifying extracted aspects from textual data into Kano categories. To improve the accuracy and efficiency of classifying online reviews, Yanlin Shi and Qingjin Peng [17] improved KANO model by applying a new classification method based on sentence meaning using the word vector. Focusing on mining changes in user expectation, Tianjun Hou et al. [18] proposed an approach for capturing ever-changing user needs on product affordances based on the online reviews for two generations of products. From the product improvement perspective, Jiayin Qi et al. [19] improved KANO model by integrating it with the classical conjoint analysis model. And it is then applied to analyze online reviews to develop appropriate product improvement strategies. In order to avoid uncertainty, Yuanju Qu et al. [20] integrated fuzzy Kano and fuzzy analytic hierarchy process to evaluate the requirements of smart manufacturing systems. Taking affective factors into consideration, Jinming Zhang et al. [21] combined sentiment analysis with fuzzy Kano to obtain the users’ different attitudes toward aspects of the product from online reviews.

It is obvious that researchers have tried to refine traditional methods from distinct angles by drawing on intelligent algorithms. Although these existing methods have been examined to be helpful for exploring user needs, they can hardly identify latent needs due to difficulties in understanding precise information from unstructured text data. What’s more, these methods can be expensive, time-consuming, and labor-intensive.

2.2. NLP-based methods to identify user needs from online reviews

Precise processing of online reviews can improve the effectiveness of

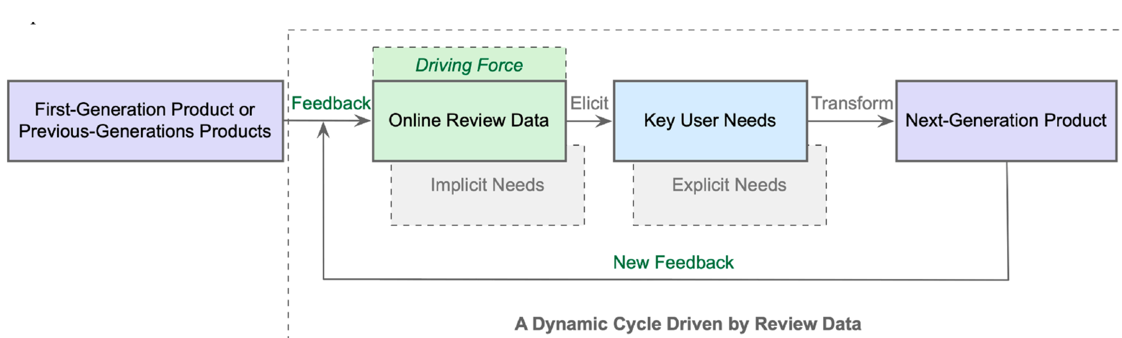

Fig. 1. The dynamic cycle of user needs elicitation driven by online review data.

capturing user needs. In order to mine latent user needs from online reviews efficiently, various methods have been proposed. NLP-based models have been employed as an inexpensive mechanism to support corresponding tasks.

2.2.1. Machine learning

Machine learning (ML) algorithms are statistics-based in essence to achieve the understanding of natural language. Scholars have applied Naïve Bayes, Support Vector Machine (SVM), TF-IDF, TextRank, etc. to analyze product reviews directly and uncover user needs [22]. Hui Sun et al. [23] proposed a creative method for dynamically mining user needs while using TF-IDF to measure the importance weights. Feng Zhou et al. [24] presented a two-layer model for latent user needs elicitation through use case analogical reasoning. And Fuzzy SVM was developed to build sentiment prediction models based on a list of affective lexicons. Considering the perspective of Kansei engineering, W.M. Wang et al. [25] aimed to combine it with text mining approaches to automatically convert unstructured product-related texts to feature-affective opinions.

The keywords that contain underlying users' opinions on products can be used to reveal user needs [26]. Irma Patricia Delgado-Solano et al. [27] presented a method for extracting keywords from users' written requirements using TextRank and inverse frequency analysis. Junegak Joung et al. [28] proposed automated keyword filtering to identify product attributes from online reviews based on LDA. In order to implement LDA, Feng Zhou et al. [29] combined LDA with a rule-based unsupervised ML technique, VADER, to perform sentiment analysis on informative review data. There are also some methods that suggest how to effectively categorize user reviews. Mengsi Cai et al. [30] proposed a product-and-user approach for requirement analysis. It can automatically elicit, classify, and rank product requirements from online reviews, and then identify requirement differences among users. Jim Buchan et al. [31] introduced a binary classification approach to large volumes of unstructured review data. In general, those methods are carried out from traditional perspectives, like product shape, color, etc. Some scholars have also explored novel perspectives to bring out the greater potential for online reviews. To identify perceptions of sustainable features, Nasreddine El Dehaibi et al. [32] detailed a design method by collecting online reviews, manually annotating them using crowd-sourced work, and processing the annotated review fragments.

Although tremendous efforts have been devoted to leveraging tools and technologies to enhance implicit user needs elicitation, still there is much space for improvement. In terms of feature extraction, these ML-based models still require a large number of manual efforts to engage in labeling the data [33]. What's more, traditional ML algorithms need intervention or guidance from domain experts on information extraction so as to achieve decent results [34]. The process is sometimes accompanied by problems such as the curse of dimensionality, data sparseness, and word isolation. Unfortunately, the accuracy of these ML-based tools is imperfect, which jeopardizes the reliability of the analysis results.

2.2.2. Deep learning

Deep learning (DL) algorithms used in the NLP field are actually context-based modeling approaches. User needs are often expressed in linguistic terms which tend to be abstract, fuzzy, and conceptual. Therefore, a growing number of scholars are studying online reviews by adopting DL algorithms, such as Convolutional Neural Network (CNN), Long Short-Term Memory Network (LSTM), Recurrent Neural Network (RNN), etc. W.M. Wang et al. [35] proposed a heuristic DL method that extracted affective opinions from product reviews and then categorized them into seven pairs of affective attributes. Mattia Atzeni et al. [36] presented a DL approach for sentiment polarity classification based on a two-layer bidirectional LSTM (BiLSTM). Artem Timoshenko and John R. Hauser [37] identified informative sentences with a CNN in order to extract user needs from user-generated content. Yue Wang et al. [38] build a BLSTM CRF-based classifier to extract keywords of user needs to boost the efficiency of mapping them to design parameters. Aiming at the specific design, Xinjun Lai et al. [39] utilized a BiLSTM-CRF model to extract keywords that describe design elements or user opinions from reviews to facilitate Kansei engineering for new energy vehicle exterior design.

When eliciting user needs, the quality of the feature representation of review data has a direct effect on the results. Compared with ML-based methods, DL-based ones have accomplished considerable improvement. However, there are still certain weaknesses: online reviews vary from product to product and need to be retrained according to product types. In addition, word vectors cannot guarantee the representation of complete semantic information, resulting in poor robustness and generalization ability of DL models.

2.2.3. Pre-trained language models

Pre-trained language models enable word vectors to dynamically adapt to the context and address the polysemy problem. Pre-trained models' training process is based on large-scale unlabeled corpora, which essentially belong to transfer learning [40]. Knowledge learned from the large-scale corpora partly compensates for the limitation of DL algorithms which do not have access to outside knowledge. Previous studies have substantiated that pre-trained language models can achieve state-of-the-art results on a wide range of language modeling tasks, including classification, summarization, similarity learning, recommendation systems, etc.

Some researchers have made progress in applying corresponding modeling architectures, such as Embeddings from Language Models (ELMo) [41], Bidirectional Encoder Representation from Transformers (BERT) [42], and StarSpace [43]. Yi Han and Mohsen Moghaddam [44] proposed an efficient and scalable methodology for automated and large-scale elicitation of attribute-level user needs. The methodology is built on BERT, with a new convolutional net and named entity recognition layers for extracting attribute, description, and sentiment words from user reviews. Considering that the amount of feedback typically obtained is too large to be processed manually, Rohan Reddy Mekala et al. [45] proposed a DL-backed pipeline for needs extraction in low-volume training dataset environments. Aiming to provide a novel research perspective, Xuanyu Wu et al. [46] proposed a semantic-driven method for user requirements revelation by combining BERT and improved LDA. Inspired by the masking strategy of BERT, Yu Sun et al. proposed a model called enhanced representation through knowledge integration (ERNIE) by using knowledge masking strategies [47]. It is a knowledge-enhanced semantic representation model established on the self-attention mechanism [48] with the strength of pre-training. The complete semantic representation of concepts can be learned by the modeling of prior knowledge in massive data.

Established methods show a deeper insight into language context and have achieved some results in user needs elicitation. However, there is relatively little and limited research in the area of product design using pre-trained language models. It is currently still in the exploration and development phase.

2.3. A brief summary of reviewed works

By analyzing the previous literature, two points are summarized as follows.

(1) From the perspective of the research trigger, considerable work has been done on mining user needs based on large-scale online reviews. And a plentiful supply of reviews can provide enough information for extant methods to study. For a new product, user feedback plays an essential role in user needs elicitation when upgrading products. But a greater number of reviews is usually unavailable because new products can't open the market quickly. And previous methods usually fail to perform well enough in the small-sample data environment. Facing the contradiction, the research problem will be looking at a way to elicit more valuable

user needs from limited review data to facilitate new product iteration.

(2) From the perspective of key technology, research attention has transferred from applying traditional methods to adopting NLP techniques. The development of NLP techniques is substantially a continuous evolution of how to accurately characterize natural language. Most scholars have targeted to improve or combine different techniques to learn and express the semantics of large-volume user reviews. Unlike previous studies, this paper is driven by small-sample reviews, and luckily, the pre-trained language model shows greater advantages in such conditions. It has been trained in advance on several large corpora. When applied, it can enhance the deep understanding of the semantics of online reviews by taking into account prior knowledge [47]. So it can rely on a smaller number of data to achieve desired results faster.

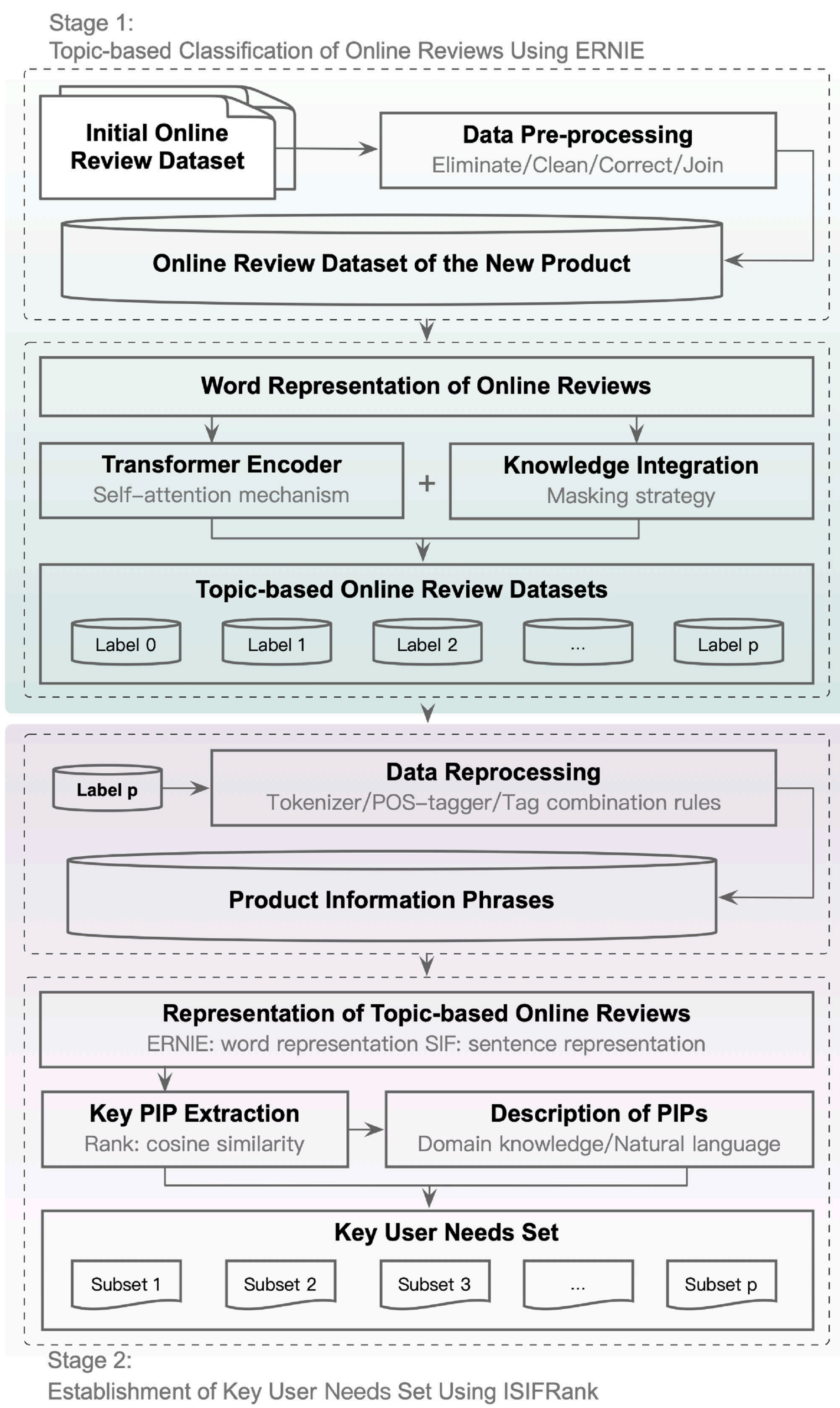

Fig. 2. Proposed analytical framework of the ERNIE-ISIFRank methodology.

Considering the two points above, the original contributions of this research can be concluded. It is the first time to combine ERNIE and improved SIFRank (ISIFRank) to elicit more valuable user needs in the environment of small-sample review data for new product iteration. The ERNIE-ISIFRank methodology is established on the basis of the pre-trained language model, which possesses the potential to understand the semantic meaning of online reviews accurately and effectively.

3. Methodology

3.1. Overview of the ERNIE-ISIFRank methodology for user needs elicitation from small-sample online reviews

It is very time-consuming and labor-intensive for professionals to manually browse thousands of online reviews and summarize user needs. The proposed methodology's logic is to first classify the disorganized reviews according to topics and then extract key phrases in each category. Thereafter, professionals can quickly capture the focus of each topic in a short time. On this basis, the scattered key phrases are supplemented and completed into natural language, i.e. user needs.

In detail, the proposed methodology of user needs elicitation from small-sample online reviews contains two main stages: (1) topic-based classification of online reviews; (2) establishment of key user needs set. In the first stage, online reviews are collected as an initial dataset which is pre-processed roughly to improve the data quality on the whole. And then online reviews are classified into multiple categories according to topics using fine-tuned ERNIE. In the second stage, every category of online reviews is reprocessed, including tokenization, part-of-speech (POS) tagging, and combining tokens as product information phrases (PIPs). By extracting key PIPs from every category, the key user needs set can be determined and facilitated utilizing ISIFRank. The research framework is illustrated in Fig. 2.

3.2. Topic-based classification of online reviews using ERNIE

In this section, ERNIE is utilized to carry out the topic-based classification of online reviews. The procedure consists of four parts. Firstly, initial online reviews are pre-processed. Secondly, unstructured review text is presented in the form of vector representation as the input of ERNIE. Thirdly, the word representation of online reviews is obtained based on the Transformer encoder and knowledge integration. Finally, the softmax classifier is added to the output end of fine-tuned ERNIE to support the classification task. Topic-based classification of online reviews using ERNIE is demonstrated in Fig. 3.

3.2.1. Data pre-processing and input representation

Authoritative e-commerce websites are the main sources of online reviews and the Scrapy crawler was adopted for data collection. The initial data contain disturbing information and spam comments. So they need to be pre-processed [49]. The specific steps are as follows.

Step 1: Eliminate default, repetitive, and false reviews, leaving behind useful ones.
Step 2: Remove invalid and broken characters and whitespace from reviews to complete data cleaning.
Step 3: Identify spelling, lexical, syntactic, and semantic errors to complete correction.
Step 4: Join different descriptions of the same entity together under a single name.

After pre-processing, an online review dataset with good quality is obtained. However, computers are not able to process unstructured

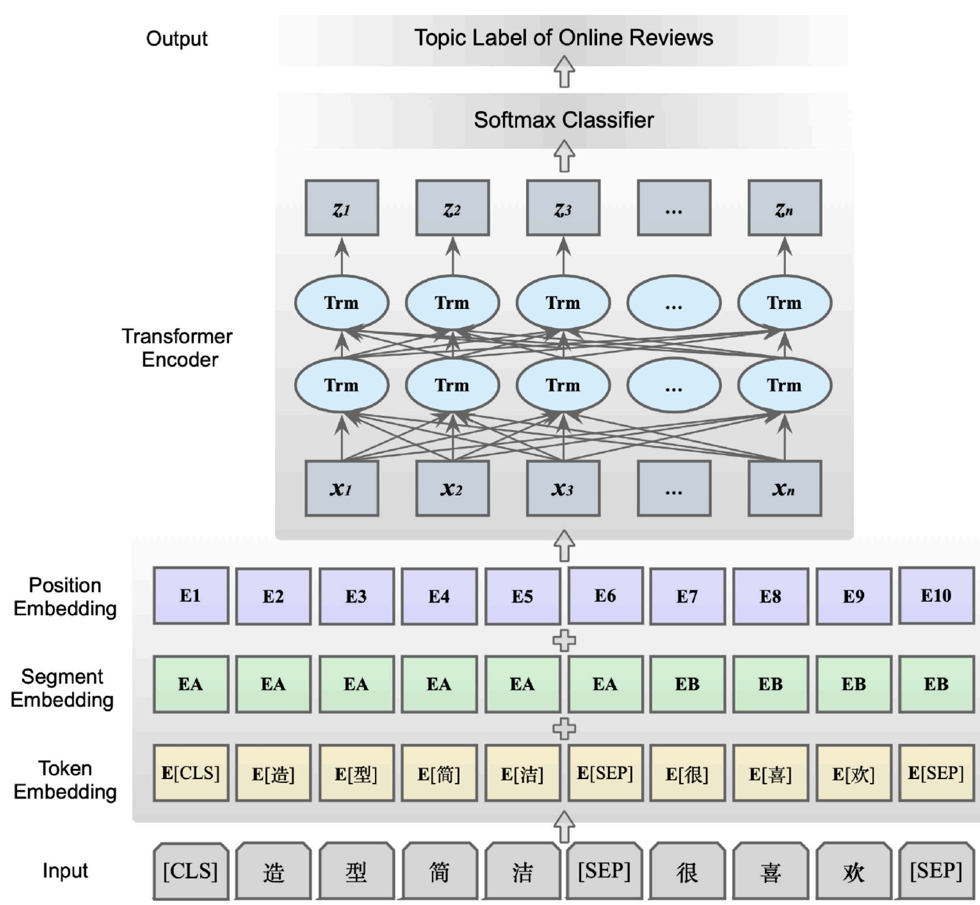

Fig. 3. Topic-based classification of online reviews using ERNIE.

online reviews directly. Therefore, it is necessary to convert the unstructured data into vector representation.

For a given input online review $T = (t_1, t_2, \cdots, t_n)$, t_n is regarded as a character in Chinese. For instance, T is "造型简洁, 很喜欢。" (The shape is simple, and I like it very much). The input representation $X = (x_1, x_2, \cdots, x_n)$ is constructed by summing the corresponding token embedding, segment embedding, and position embedding. x_n is the input representation of t_n. To represent each word, the model uses a 768-dimensional embedding vector. That is to say, the dimension of the input representation is 768.

To be clear, [CLS] and [SEP] are contained in every token embedding. [CLS] is added as a classification token at the first place of every review. And [SEP] is added as a special separator token to cut the review sentence [42]. In addition, segment embedding is used to distinguish whether the words belong to different sentences. And position embedding encodes position information. A visualization of this construction can be seen in Fig. 3.

3.2.2. Transformer encoder for word representation

The input representation can be then put into and understood by ERNIE to get word representation. ERNIE deploys the Transformer encoder as the framework for word representation, mainly consisting of self-attention mechanism and feed forward neural networks [46]. First and foremost, the input representation $X = (x_1, x_2, \cdots, x_n)$ comes into the self-attention mechanism layer, obtaining $Y = (y_1, y_2, \cdots, y_n)$. Next, $Y = (y_1, y_2, \cdots, y_n)$ comes into the feed forward neural networks, acquiring the output representation of ERNIE $Z = (z_1, z_2, \cdots, z_n)$. In this way, we take $Z = (z_1, z_2, \cdots, z_n)$ as the word representation of the input online review $T = (t_1, t_2, \cdots, t_n)$. Fig. 4 depicts the architecture of the Transformer encoder.

Furthermore, the self-attention mechanism is good at calculating the relationship between any two words. It can capture feature information at a longer distance within the review text [50]. Meanwhile, it can also distinguish the ambiguity problem caused by words with multiple interpretations. Thus, it helps ensure a more complete word representation. The calculation formula inside the unit can be expressed as follows:

$$\begin{cases} Q = X \bullet W^{Q} \\ K = X \bullet W^{K} \\ V = X \bullet W^{V} \end{cases} \tag{1}$$

$$Y = Softmax\left(\frac{Q \bullet K^{T}}{\sqrt{d_k}}\right) \bullet V \tag{2}$$

$$Z = W \bullet Y + b \tag{3}$$

where Q is the query matrix, K is the key matrix, and V is the value matrix. Specifically, through weight matrices W^Q, W^K, and W^V, input representation X is transformed into Q, K, and V, respectively. d_k is the dimension of input representation. $\sqrt{d_k}$ is designed to avoid the situation that leads to an undesirable outcome of softmax, either 0 or 1. W is the weight matrix and b is the bias value.

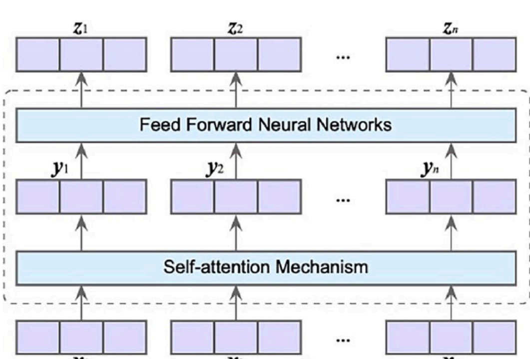

Fig. 4. The architecture of the Transformer encoder.

What has been illustrated above is the working principle of one-head. ERNIE employs multi-head architecture. It comprises 12 layers of the Transformer encoder. Each Transformer encoder contains 12 attention heads in total [45]. To be clear, the Transformer encoder uses a twelve-head mechanism to obtain corresponding word representation Z_i $(i = 1, 2, \cdots, 12)$ by invoking different mapping matrices in parallel to complete Eqs. (1)–(3). Finally, the word representation Z is produced as follows:

$$Z = concat(Z_1, Z_2, \cdots, Z_{12})W^{0} \tag{4}$$

where $concat(Z_1, Z_2, \cdots, Z_{12})$ means to concatenate twelve matrices, and W^0 is a weight matrix.

3.2.3. Knowledge integration of online reviews

ERNIE's own pre-training corpora are not specific and targeted enough for product reviews. To improve its generalization ability, an additional corpus of online reviews from common products is used for knowledge integration. Hence, the knowledge of real-world products' language descriptions can be contained in the word representation of online reviews.

A multi-stage knowledge masking strategy is embedded in knowledge integration. The strategy enhances word representation through knowledge-based reasoning. To be specific, knowledge masking refers to training the model to predict the masked part by masking some characters randomly. And then it can learn the contextual information of the masked part. ERNIE inherits the character-level masking from BERT and enriches the strategy with entity-level masking and phrase-level masking. The diversity of words can be represented by these three masking strategies. And the word representation can retain the correlation between components in online reviews [51].

Masking and predicting are the key to this strategy. The first learning stage is basic-level masking. It treats an online review as a sequence of basic language units. We randomly mask 15 percent of the basic language units. And other basic units in the sequence are used as inputs to train a transformer to predict the masked units. For Chinese, the basic language unit is a Chinese character. The second stage is entity-level masking. Entities can be abstract or have a physical existence of the specific new product. The third stage is phrase-level masking. A phrase is a small group of Chinese characters together acting as a conceptual unit. In the training process, those three kinds of language units are randomly masked to train a Transformer to predict them. After the three-stage learning, high-level semantic knowledge is easy to be fully modeled in the word representation [47]. Multi-stage knowledge masking strategy is described in Fig. 5.

3.2.4. Fine-tuning and topic-based classification

The parameters of the model can be acquired through the pre-training process. However, the downstream task in this study will involve a specific new product. In order to take full advantage of the linguistic knowledge learned during the pre-training process and transfer it to the downstream task, it is necessary to fine-tune the parameters on the collected small sample review data. A re-initializing strategy for fine-tuning partial parameters is adopted to balance the gap between upstream and downstream tasks [52].

Specifically, among the 12 layers of the Transformer encoder, lower layers learn more general features, such as linguistic knowledge of part of speech, morphology, etc. Higher layers that are closer to the output will tend to learn knowledge about the downstream tasks. As a result, we reserve the weights of the bottom layers. And for the weights of the top layers (1 to 6 layers), we re-initialize them randomly and let these parameters be relearned on the downstream task [52]. Such a strategy can help to make the model become much more stable when fine-tuning in few-sample settings.

By adding the softmax classifier to the output end of ERNIE, the

Sentence	喂	食	器	颜	色	高	级	，	外	观	简	约
Basic-level Masking	[MASK]	食	器	[MASK]	色	高	级	，	外	观	[MASK]	约
Entity-level Masking	[MASK]	[MASK]	[MASK]	颜	色	高	级	，	外	观	简	约
Phrase-level Masking	喂	食	器	[MASK]	[MASK]	[MASK]	[MASK]	，	[MASK]	[MASK]	[MASK]	[MASK]

Fig. 5. Multi-stage knowledge masking strategy of an online review.

topic-based classification task can be carried out. Disordered online reviews are rearranged and stored according to different topics. Every classified online review dataset has the same topic, such as shape, color and interaction.

3.3. Establishment of key user needs set using ISIFRank

The reviews in each topic-based category directly reflect the level of user satisfaction, rather than user needs. Therefore, ISIFRank [53] is proposed to facilitate the establishment of key user needs set. The whole procedure consists of four parts. Firstly, topic-based online reviews are input and reprocessed. Secondly, representation of topic-based online reviews is required, including word representation and sentence representation. Thirdly, key PIPs are extracted in each category using SIF and Rank algorithms. Finally, key PIPs are connected by natural language to realize the establishment of key user needs set. As for innovation, ISIFRank introduces ERNIE to optimize word representation. Moreover, tag combination rules are specially designated, making the model more compatible with product design. The extraction of key PIPs using ISIFRank is shown in Fig. 6.

3.3.1. Data reprocessing

The prerequisite for the establishment of the key user needs set is the extraction of key PIPs from topic-based online reviews. The online reviews of each topic are stored in the same text document, where the review data still remain in natural language. To facilitate the training of online reviews in the ISIFRank model, it needs to be processed again. The specific steps are as follows.

Step 1: The online review is tokenized by THU Lexical Analyzer for Chinese (THULAC).

Step 2: Every token is POS tagged via Natural Language Toolkit (NLTK).

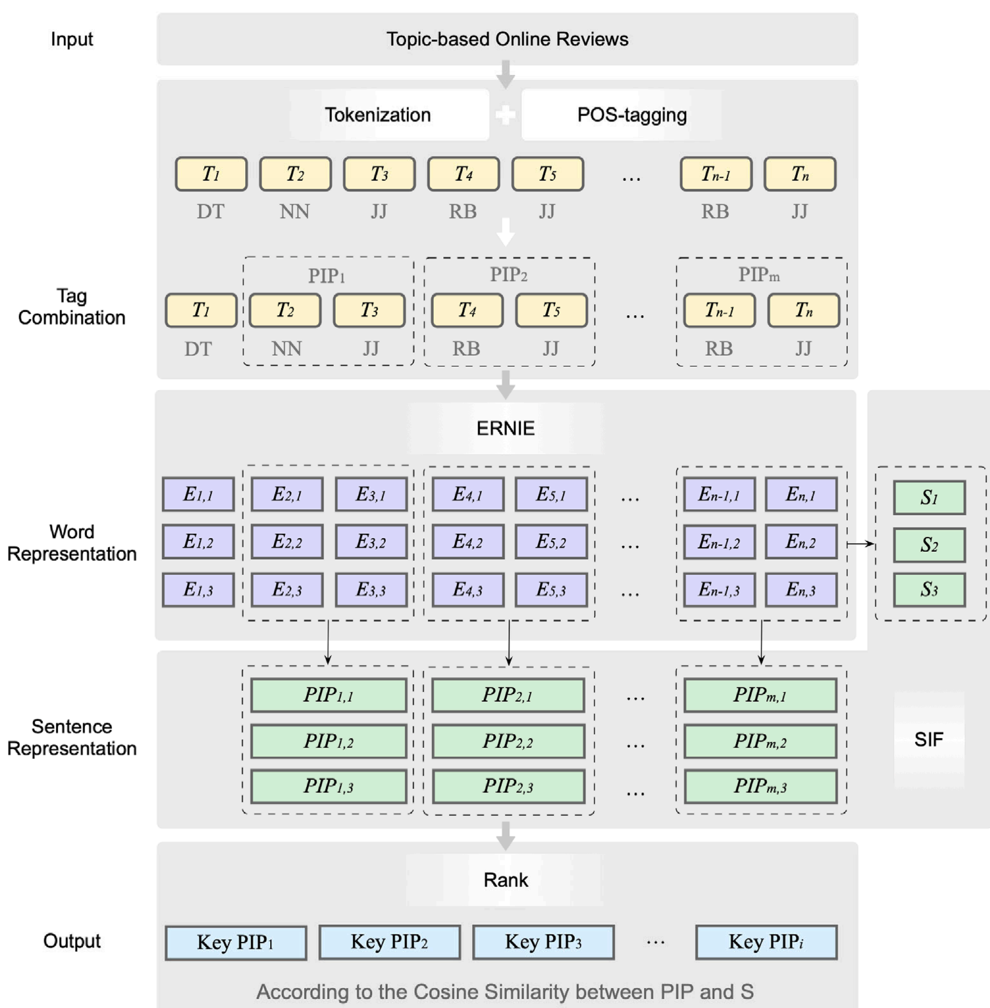

Fig. 6. Extraction of key PIPs using ISIFRank.

Step 3: Combine tokens as PIPs from the online review according to the POS tags using regular expression. The extracted PIPs are the candidate key PIPs.

The regular expression is created on the basis of novel tag combination rules. Considering the daily language usage in the sphere of product, we defined eight tag combination rules. They indicate user needs mainly around adjectives and verbs, according to Chinese linguistic habits and lexical matching principles. Four rules are about adjectives and another four rules are about verbs. The set of PIPs can be established according to the tag combination rules, where $PIP = (PIP_1, PIP_2, \cdots, PIP_m)$. The rules are proposed and defined as:

$\mathrm{JJ}\begin{cases}\mathrm{JJ}\\ \mathrm{NN}+\mathrm{JJ}\\ \mathrm{RB}+\mathrm{JJ}\\ \mathrm{DT}+\mathrm{JJ}\end{cases}\mathrm{VB}\begin{cases}\mathrm{VB}\\ \mathrm{RB}+\mathrm{VB}\\ \mathrm{VB}+\mathrm{RB}\\ \mathrm{VB}+\mathrm{NN}\end{cases}$ where JJ is the adjective, NN is the noun, RB is the adverb, VB is the verb, and DT is the determiner. All of these are based on NLTK lexical notation symbols.

When designing these rules, user sentiments are also considered at the same time. Formed according to the designated rules, some PIPs are phrases that can indicate user sentiments (positive or negative) to some extent. What's more, the extracted PIPs are not merely limited to noun phrases as usual. Different kinds of PIPs can reflect user needs more exactly.

3.3.2. Representation of topic-based online reviews

The vector representation of topic-based online reviews consists of two parts: word representation and sentence representation.

(1) Word representation.

Traditional SIFRank obtains word representation by ELMo. However, the inference speed of ELMo is slow. In order to make up for this shortage, ERNIE is transplanted to replace ELMo to obtain the word representation of online reviews. To be specific, the semantic integrity and generalization of word representations are ensured by characterizing the diversity of online reviews from three aspects: token, segment, and position (see Section 3.2.2–3.2.4 for details).

(2) Sentence representation.

A piece of user review is regarded as a sentence. PIP is a phrase that can also be regarded as a part of the sentence in essence. Therefore, the vector representations of PIPs and topic-based online reviews are both called sentence representations. For an online review s, its word representation is v_w, which is acquired through the ERNIE algorithm. Its sentence representation is v_s, and the sentence representation of related PIP is v_{PIP}. They are both acquired through the SIF algorithm. Traditionally, sentence representation is obtained by summing and averaging the word representation. The SIF algorithm varies in this part. It obtains the sentence representation by weighted averaging each word vector, which can better capture the central idea of the online review. The equation is expressed as follows:

$$Weight(w) = \frac{a}{a + p(w)} \tag{5}$$

where a is a constant, which is probably suitable in $[10^{-3}, 10^{-4}]$, and $p(w)$ is the term frequency. The sentence representation of the online review is expressed as follows:

$$v_s = \frac{1}{|s|}\sum_{w\in s}\frac{a}{a + p(w)}v_w = \frac{1}{|s|}\sum_{w\in s}Weight(w)v_w \tag{6}$$

It works the same way with the sentence representation of PIP v_{PIP}. At this point, all of these representations have the same number of layers and dimensions [53].

3.3.3. Extraction of key PIPs

Key PIPs are extracted through the Rank algorithm. For a given topic-based online review and its PIPs, the rank of PIPs is defined as the similarity or correlation score between v_s and v_{PIP}:

$$Rank(v_{PIPi}, v_s) = Similarity(v_{PIPi}, v_s) \tag{7}$$

The similarity can be generally calculated by cosine distance:

$$Similarity(v_{PIPi}, v_s) = cos(v_{PIPi}, v_s) = \frac{\overrightarrow{v}_{PIPi} \bullet \overrightarrow{v}_s}{\|\overrightarrow{v}_{PIPi}\|\|\overrightarrow{v}_s\|} \tag{8}$$

when the Euclidean distance is used to calculate the similarity, the weight of the representations should be normalized [53].

The value of the rank of key PIP is between 0 and 1. The closer it is to 1, the more relevant the key PIP is to the topic of the online review. Conversely, the closer the value is to 0, the more irrelevant the key PIP is to the topic.

Define a threshold ε, the PIP whose cosine similarity is greater than ε is selected as a key PIP finally.

3.3.4. Establishment of key user needs set

Extracted key PIPs focus on the most critical points, which are perceived by users in the course of using the new product. But the information is fragmented, scattered, and not a complete expression of user needs. Technically, it is hard to generate accurate user needs automatically with key PIPs because it is a process that requires the involvement of design experience and knowledge for certain products. In order to generate a high-quality key user needs set, the key PIPs are manually transformed into key user needs based on natural language descriptions. During the transformation, expert knowledge and domain knowledge of new products are incorporated.

The key user needs subset B can be created from the respective topic-based online review datasets. The composition of B includes topic t, key PIPs $KeyPIP$, normalized weight (N-Weight) w, and natural language description NLD, i.e. $B = \{t, KeyPIP, w, NLD\}$. The key user needs set A is made up of all subsets B, i.e. $A = \{B_1, B_2, \cdots, B_j\}$. It can potentially allow designers, engineers, and decision-makers to have access to useful reference information to benefit the iterative upgrading of new products.

4. Case study

An illustrative example is presented to elaborate on how the ERNIE-ISIFRank methodology works. The motivation of the empirical study is to confirm that the proposed methodology is feasible and promising. In China, with the increasing number of cat owners in recent years, the pet-related industry is rapidly evolving. This phenomenon leads to the diversification in offerings specifically designed for cats. Smart cat feeders are one of the most newly created intelligent products. The latest generation of smart feeders called Real from a bestselling brand Homerun, launched in November 2021, is chosen as the target product in this case study.

4.1. Data preparation

Based on the two principles of high authority and excellent review quality, well-known e-commerce websites in China were selected as the source of online reviews, including JD, Suning, and Tmall. The smart cat feeder has a serious shortage of online reviews within three months, which will affect the performance of the ERNIE-ISIFRank model. Therefore, we decide to set the collection time to half a year.

Web crawler technology is adopted to collect user reviews of the Homerun Real Smart Cat Feeder, which were posted on its online retail stores from December 15th, 2021 to June 31st, 2022. There are 6754 pieces of reviews in the initial online review dataset.

4.2. Key user needs elicitation of smart cat feeder

4.2.1. Results of the topic-based classification of online reviews

According to Section 3.2.1, initial online reviews are pre-processed, getting the online review dataset of the smart cat feeder. There is a total of 5000 pieces of online reviews in the dataset. Considering the relevant characteristics of the smart cat feeder, the topic labels are determined from three dimensions: appearance, function, and emotion after professional discussion. We assign eight labels from 0 to 7, which are shape (S), color (C), material (M), interactive operation (IO), intelligent feeding (IF), installation and maintenance (IM), value for money (VM), and practicality (P). And then, 5000 online reviews are randomly divided into a training set and a test set in a ratio of 8:2. The examples of online review labels are shown in Table 1.

According to Section 3.2.2–3.2.4, online reviews of the smart cat feeder are classified into eight categories which are saved in eight text files. That is to say, eight topic-based online review datasets are obtained.

4.2.2. Results of the establishment of key user needs set

According to Section 3.3.1, topic-based online reviews are reprocessed. Citing the "Shape" topic as an illustration, tokenization is completed through THULAC, and POS tagging is carried out through NLTK. PIPs are obtained by integrating regular expression and tag combination rules. The processing results are shown in Table 2.

According to Section 3.3.2 and 3.3.3, word representation and sentence representation of the online reviews are calculated via ERNIE and SIF algorithms, respectively. On the basis of the Rank algorithm, key PIPs are extracted in descending order of the weight score. We manually check all of the key PIPs thoroughly and get rid of the wrong ones. By setting the threshold as 0.45, the key PIP whose cosine similarity exceeds 0.45 is selected as the main element to be considered when establishing the key user needs set. Extracted key PIPs and their weights of all labels are shown in Table 3.

Extracted key PIPs are product-related information with high user interest and play an essential role in establishing key user needs set. The analysis of Table 3 finds that different PIPs reflect different user attitudes. As a matter of fact, all of the PIPs of label 0, label 1, and label 5 belong to positive descriptions, indicating that users are satisfied with these aspects of the smart cat feeder to some extent. It means that the existing product has largely met user needs. User needs generalized from this situation are called "met needs". However, some of the PIPs of label 2, label 3, label 4, label 6, and label 7 belong to positive description and the others belong to negative description, which signifies that user experience is inconsistent. It indicates that the existing product does not fully meet most user needs in these aspects, and there is still much room for improvement. User needs generalized from this situation are called "unmet needs". "Met and unmet needs" together constitute the key user needs.

Table 1
The examples of online review labels.

Dimension	Label	Topic	Examples of online review
Appearance	0	S	造型简洁，很喜欢。 (The shape is simple, I like it.)
	1	C	喂食器颜色高级，十分百搭。 (The color of the smart feeder is fancy, it is quite versatile.)
	2	M	塑料材质，对猫咪不友好，担心安全问题。 (Plastic is not friendly to cats. I kind of worry about safety issues.)
Function	3	IO	可以通过APP操作，对上班族来说很方便。 (It can be operated by APP, which is convenient for office workers.)
	4	IF	不仅自动出粮，还能称重，真是太棒了。 (Not only does it automatically feed cat, but it also weighs cat food. So great.)
	5	IM	有教程，安装简单好上手，爱了爱了。 (Tutorials are available and it is easy to put together. Love it.)
Emotion	6	VM	总体来说，这款喂食器，性价比一般。 (Generally, this feeder is not cost-effective enough.)
	7	P	简直太实用了，解放双手了，哈哈哈。 (It's so practical and can free my hands, ha-ha.)

Table 2
An example of reprocessing results.

Online review	这款喂食器造型简洁，很喜欢!
Tokenization	{"这款", "喂食器", "造型", "简洁"}, {"很", "喜欢"}
POS tagging	{"DT", "NN", "NN", "JJ"}, {"RB", "VV"}
Tag combination rules	"NN" + "JJ" → "造型" + "简洁" "RB" + "VB" → "很" + "喜欢"
Combination results	{"这款", "喂食器", "造型简洁"}, {"很喜欢"}
PIPs	"造型简洁", "很喜欢"

Table 3
Key PIPs and their weights (in Chinese).

Label	Key PIP (Weight)
0	造型简洁(0.84) 颜值高(0.76) 高级(0.74) 质感好(0.54) 容量适中(0.46)
1	好看(0.82) 漂亮(0.67) 配色高级(0.51) 高档(0.49)
2	材质安全(0.90) 不黑下巴(0.85) 担心安全(0.51)
3	方便(0.92) 界面简单(0.87) 按键清晰(0.83) 远程操作(0.83) 连接无线网(0.79)按键灵敏(0.65) 连网使用(0.64) 手动出粮(0.46) 经常断连(0.77)
4	智能(0.93) 自动出粮(0.92) 出粮顺畅(0.89) 出粮口小(0.74) 不密封(0.73) 卡粮(0.72)不保鲜(0.68) 猫碗低(0.68) 称重(0.67) 减肥(0.58) 出粮失败(0.56) 卡冻干(0.47)
5	安装方便(0.90) 很方便(0.89) 简单(0.72) 安装简单(0.65) 容易清洗(0.50)
6	性价比一般(0.89) 满意(0.83) 性价比低(0.81) 略贵(0.73) 价格略高(0.63) 性价比高(0.46)
7	实用(0.88) 实用性强(0.82) 好用(0.77) 方便(0.68) 简单(0.55) 一般(0.52) 不实用(0.45)

Note: Label 2, 3, 4, 6, and 7 contain both positive and negative PIPs, Label 0, 1, and 5 only contain positive PIPs.

According to Section 3.3.4, considering the product knowledge related to smart feeders, key PIPs are summarized and merged into natural language. They are transformed into complete and comprehensible sentences which are called key user needs. Normalize the weight of the key PIPs within the topic, and then establish a subset of key user needs. Each topic-based subset is created in the same way. All of these subsets form the key user needs set of the smart cat feeder, as shown in Fig. 7.

4.2.3. Analysis of the results of the case study

By applying the proposed methodology to the practical example of a smart cat feeder, the analysis of messy online reviews is fulfilled via the classification and extraction process. Therefore, the exploration of explicit user needs is completed expeditiously based on low-volume online reviews. ERNIE can effectively assist in the task of classifying disorganized online reviews according to different topic labels, saving both time and labor costs. ISIFRank performed well enough to extract high-quality key PIPs, ensuring a fairly accurate and quick establishment of key user needs set.

As shown in Fig. 7, the application results are presented as a set of key user needs from three dimensions, including appearance, function, and emotion. Different forms of information are contained in every dimension, including topic, key PIPs, N-Weight, and natural language description. All the information can be used as the basic data resource for the analysis of the smart cat feeder. And the developers can refer to the information for related iteration work. In this case, designers can

Appearance Needs			
Topic	Key PIPs	N-Weight	Key User Needs in Nature Language
Shape	造型简洁	0.2515	UN1:用户需要智能喂食器具有简洁的造型，希望整体造型美观、高级并体现产品质感。(Users want the shape of the smart cat feeder to be infuded with simplicity and beauty, which could creat a premium tecture.)
	颜值高	0.2275	
	高级	0.2216	
	质感好	0.1617	
	容量适中	0.1377	UN2:用户较为关注喂食器的容量，部分用户偏向其容量适中，可较为实用。(Users are concerned with the capacity for cat food. Some users prefer a moderate capacity which could be more practical.)
Color	好看	0.3293	UN3:用户需要喂食器具有美观且高级的配色。(Users need beautiful and premium color schemes.)
	漂亮	0.2691	
	配色高级	0.1092	
	高档	0.1049	
Material	材质安全	0.3982	UN4:用户需要喂食器的材质安全可靠，这与猫咪健康息息相关。(Users need safe and reliable materials for the feeder, which is closely related to the cat's health.)
	担心安全	0.2257	
	不黑下巴	0.3761	UN5:用户希望猫碗的材质不要导致猫咪黑下巴(一种由猫碗材质引起的猫咪疾病)。(Users need that the material of the bowl does not cause black chin in cats which is a cat disease caused by plastic bowl.)

(a) Appearance dimension

Fig. 7. Key user needs set of the smart cat feeder.

develop improvement strategies for the next-generation product to meet critical user needs. And decision-makers can make calculated and intelligent decisions on improving versions of the smart cat feeder. Acquiring user needs is an essential and necessary part of new product iteration. Different roles in the company can think from their own perspectives around the key user needs set to advance the iterative upgrade process of the smart cat feeder, scientifically, efficiently, and quickly.

5. Discussion

5.1. Comparative experiments

In order to demonstrate the advantages of ERNIE-ISIFRank methodology based on the pre-trained language model, the comparison is carried out between the proposed method and a list of commonly used methods.

With respect to the topic-based classification of online reviews using ERNIE, we consider the evaluation as a series of eight-class classification problem. We choose SVM, FastText, BiLSTM, TextCNN, DPCNN and Bert as baseline methods. The performance is measured by precision, recall, and F1-score, which are calculated by Eqs. (9), (10), and (11), respectively,

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F_1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

For SVM, gamma is set to 1.67 and the penalty term is set to 1.32. For FastText, we use the default setting. For BiLSTM, the dropout rate is 0.5, the learning rate is $1e-3$, and the batch size is 18. For TextCNN, the dropout rate is 0.5, the learning rate is $1e-3$, and the batch size is 32. For DPCNN, the dropout rate is 0.5, the learning rate is $3e-5$, and the batch size is 32. For Bert, the dropout rate is 0.5, the learning rate is $3e-5$, and the batch size is 18. For ERNIE, the dropout rate is 0.5, the learning rate is $5e-5$, and the batch size is 16.

With respect to the extraction of key PIPs using ISIFRank, we choose TF-IDF, YAKE, TextRank, TopicRank, PositionRank and SIFrank as baseline methods to conduct comparative experiments. To objectively assess the extraction effectiveness, precision, recall, and F1-score of different methods are recorded, which are calculated by Eqs. (12), (13), and (14) respectively,

$$Precision = \frac{C_{\text{right}}}{C_{\text{extract}}} \tag{12}$$

$$Recall = \frac{C_{\text{right}}}{C_{\text{standard}}} \tag{13}$$

$$F_1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{14}$$

where C_{right} is the number of all rightly extracted key PIPs, C_{extract} is the number of all extracted key PIPs, and C_{standard} is the number of all manually labeled standard answers.

For TF-IDF, the n-gram window length is set to 3. For YAKE, the window size is 1, the deduplication threshold is 0.85, and the n-gram length is 3. TextRank has window sizes of 2. The minimum clustering similarity threshold of TopicRank is set to 0.75. PositionRank has a window size of 10. In the SIFRank, we use the original pre-trained ELMo model. As for the ISIFRank, we use the pre-trained ERNIE model in our paper. All the models use the same tools for tokenizing and POS tagging

Function Needs			
Topic	Key PIPs	N-Weight	Key User Needs in Nature Language
Interactive Operation	方便	0.1429	UN1:操作方便是用户对交互操作的重点要求。用户需要简单的APP操作界面以确保操作便利性。(Users need a simple and convenient app interface, allowing for easy opration.)
	界面简单	0.1351	
	远程操作	0.1289	UN2:连接无线网的功能满足了用户进行远程操作的需求，但用户需要更加稳定的连接，以确保喂食器的正常运行。(The ability to connect to a wireless network meets the user's need for remote operation, but users need a more stable connection to ensure proper operation of the feeder.)
	连接无线网	0.1227	
	经常断连	0.0699	
	连网使用	0.9994	
	按键清晰	0.1289	UN3:该喂食器具有自动喂食和手动喂食两种模式，在手动喂食时，用户需要灵敏的按键提升用户体验。(The smart feeder has two modes: automatic feeding and manual feeding. When feeding manually, users need sensitive buttons to enhance the user experience.)
	按键灵敏	0.1010	
	手动出粮	0.0994	
Intelligent Feeding	智能	0.1085	UN4:自动出粮是用户对喂食器的首要诉求，且要确保猫粮可以顺畅掉落，避免卡在出粮口。冻干体积比猫粮大一些，在有冻干与猫粮混合的情况下，尤其要确保冻干不会卡在出粮口。(The ability to automatically dispense cat food is the primary user need for the feeder. It is necessary to ensure that the cat food can fall smoothly and avoid getting stuck in the spout. Freeze-dried food is a bit larger than traditional cat food, so it is especially important to make sure that the freeze-dried food does not get stuck in the spout when it is mixed with cat food.)
	自动出粮	0.1076	
	出粮顺畅	0.1039	
	出粮口小	0.0863	
	卡粮	0.0840	
	出粮失败	0.0653	
	卡冻干	0.0548	
	不密封	0.0852	UN5:用户需要喂食器密封性好，可以确保猫粮新鲜。(Users need the feeder to be well sealed and can ensure that the cat food kept fresh.)
	不保鲜	0.0793	
	猫碗低	0.0793	UN6:用户需要猫碗的高度再高一点。(Users need the height of the bowl to be a little higher.)
	称重	0.0782	UN7:部分用户有猫咪减肥的需求，需要该款喂食器具有猫粮称重的功能。(Some users have the need of losing cats' weight and need a function of weighing cat food.)
	减肥	0.0677	
Installation and Maintenance	安装方便	0.2459	UN8:用户需要喂食器的安装过程方便、简单。(Users need a convenient and easy installation process for the feeder.)
	很方便	0.2432	
	简单	0.1967	
	安装简单	0.1776	
	容易清洗	0.1366	UN9:用户需要喂食器容易清洗干净。(Users need the feeder to be easy to clean.)

(b) Function dimension

Emotion Needs			
Topic	Key PIPs	N-Weight	Key User Needs in Nature Language
Value for Money	性价比一般	0.2046	UN1:用户需要喂食器具有高性价比，价格合理，不要太高。(Users need the feeder to be cost-effective and reasonably priced. A too high price is not accepted.)
	满意	0.1908	
	性价比低	0.1862	
	略贵	0.1678	
	价格略高	0.1448	
	性价比高	0.1057	
Practicality	实用	0.1884	UN2:用户需要喂食器实用性强，方便好用。(Users need the feeder to be practical, useful and convenient on the whole.)
	实用性强	0.1756	
	好用	0.1649	
	方便	0.1456	
	简单	0.1178	
	一般	0.1113	
	不实用	0.0964	

(c) Emotion dimension

Fig. 7. (*continued*).

under the same environment.

For these comparative experiments, we use 10-fold cross validation in the evaluation. Moreover, every experiment is performed 10 times to obtain the average performance metrics mentioned above. The final results are shown in Tables 4 and 5.

In the first place, we conduct the analysis from precision and recall.

Table 4
Metrics to characterize the performance of different models for topic-based classification.

Model	Precision	Recall	F1-score
SVM	0.6812	0.4934	0.5741
FastText	0.8311	0.6350	0.7227
BiLSTM	0.8443	0.6722	0.7543
TextCNN	0.8528	0.7154	0.7796
DPCNN	0.8915	0.7831	0.8314
Bert	0.9045	0.8332	0.8611
ERNIE	0.9248	0.8845	0.8934

Table 5
Metrics to characterize the performance of different models for key PIPs extraction.

Model	Precision	Recall	F1-score
TF-IDF	0.7839	0.7123	0.7417
YAKE	0.8123	0.7364	0.7734
TextRank	0.8224	0.7534	0.7844
TopicRank	0.8321	0.7545	0.7880
PositionRank	0.8429	0.7636	0.7980
SIFRank	0.8826	0.8012	0.8422
ISIFRank	0.9049	0.8143	0.8534

As shown in Table 4, it is easy to find that ERNIE scores the highest values on precision (0.9248) and recall (0.8845). Closely relying on the quality of the word representation, the other models perform poorly compared with Bert and ERNIE. As Bert and ERNIE are based on pre-trained language model, they own stronger generalization ability. And the Transformer-based encoder can adaptively extract features of review text to improve the classification accuracy. Furthermore, ERNIE possesses richer knowledge masking strategies than Bert, exhibiting a deeper understanding of language context. As such, ERNIE outperforms Bert a little further, drastically reducing the costs and improving the performance of the classifier.

As shown in Table 5, obviously, ISIFRank stands out on precision (0.9049) and recall (0.8143). TF-IDF and YAKE perform worst. Given that the keyword does not appear all the time in a single online review, these three methods depend on term frequency to extract keywords. Therefore, they are unable to understand the semantics in Chinese, which leads to extracting unsatisfactory key PIPs. TextRank, TopicRank, and PositionRank are all graph-based, but they mainly focus on the rank of importance of different words. Taking such shortcomings into consideration, SIFRank utilizes ELMo to obtain various syntactic and semantic information in Chinese. ELMo is a pre-trained language model that can ensure the accurate extraction of key PIPs. Also, SIFRank compensates for the inability of the other methods to combine two words together for a complete expression of key PIPs. This is another reason that explains why SIFRank performs dramatically better. Despite the merits of SIFRank, ISIFRank still outperforms SIFRank to some degree because of two improvements. Firstly, we introduce ERNIE to replace ELMo to address its poor inference problem while obtaining word representation. ELMo was essentially based on BiLSTM. Luckily, ERNIE has been confirmed better than BiLSTM in Table 4. Secondly, we devised novel tag combination rules, in which the linguistic description characteristics of product reviews are involved. So, it further ensures the integrity and accuracy of the extracted key PIPs.

In the second place, we summarize the analysis from F1-score which considers precision and recall at the same time. It delivers a comprehensive reflection of the model performance. As shown in Tables 4 and 5, the F1-score of ERNIE and ISIFRank is 0.8934 and 0.8534, respectively. This is a robust and comprising result in comparison with baseline methods. Overall, the ERNIE-ISIFRank-based method proved effective and efficient in a small-sample environment. It indicates that using pre-trained embeddings before training our core models is a very productive way of achieving state-of-the-art performance metrics. Our research also serves as an important step towards the broader adoption of approaches based on pre-trained language models in new product iteration. The high precision, recall, and F1-score achieved prove the very benefits of low costs and high efficiency during evaluation concerning a small scale of online reviews.

To further demonstrate the validity and necessity of fine-tuning for the small-sample data, we performed an extra experiment in which the ERNIE model was tested without fine-tuning. Precision, recall, and F1-score are achieved with measures of 0.8954, 0.8050, and 0.8478, respectively. Conferencing the results of fine-tuned ERNIE in Table 4, there is an increase of 0.0294 in precision, 0.0795 in recall, and 0.0456 in F1-score. The positive results show that the fine-tuning strategy we use does improve the performance of ERNIE in a small-sample setting. It is obvious that fine-tuning plays a vital role in the topic-based classification of few-sample online reviews.

5.2. Limitations

Despite the advantages of the ERNIE-ISIFRank methodology, some limitations should be acknowledged. (1) THULAC is used for tokenizing online reviews in this paper. As there is no obvious boundary between words in Chinese text, the result of tokenization of the same user review usually varies from one method to another. To the best of our knowledge, different results have an effect on the semantic expression, which subsequently impacts the key PIPs extraction. And that is also a great difficulty in the current research of Chinese keyword extraction. (2) The automatic sentiment analysis of key PIPs was not taken into consideration while establishing the key user needs set. Some extracted key PIPs may not contain useful information that must be ruled out manually. Meanwhile, it must be pointed out that the designated tag combination rules are not well-round enough to reflect user needs. (3) The initial review data are only scraped from online retail stores. Other data collected from sources such as user reports and social media should be incorporated to address user needs more comprehensively.

6. Conclusion

A small sample data-driven method is proposed in this paper to address user needs elicitation in new product iteration from online reviews. It attempts to resolve the difficulty of carrying out much faster processing of limited review data based on pre-trained language models. We found that with the support of the ERNIE-ISIFRank methodology, eliciting user needs from limited online reviews could be executed easily and efficiently. The proposed novel and intelligent methodology proved a boon to new product iteration.

The main theoretical contribution can be summarized into three aspects. (1) Well-defined an innovative iteration design pattern for new products. Driven by online review data, the dynamic cycle of user needs elicitation facilitates upgrading products for the next generation. (2) Presented a generic and compatible framework for extracting user needs drawing on a small sample of online reviews in the context of new product iteration. Topic-based classification and key PIPs extraction are the two stages of this framework. Accurate and efficient analysis of finite review data to uncover valuable information is the core of eliciting key user needs. And it can help to improve product functions and raise its competitiveness in the market. (3) Proposed a systematic methodology primarily based on the pre-trained language model. This is a tentative attempt that integrates both ERNIE and ISIFRank in the process of analyzing a small sample of online reviews. Moreover, we contribute to using ERNIE to improve SIFRank and designating tag combination rules which are suitable for product design. Hence, the proposed method not only enriches the graph of pre-trained language models but also focuses on the sphere of user needs elicitation for new products. It is helpful in reducing labor costs, boosting iteration efficiency, and shortening the design cycle.

As for practical contribution, an empirical case was conducted to illustrate the application of the proposed data-driven method. The study provides guidance and reference for user needs elicitation of new products. The dynamic adjustment of diverse user needs provides designers, engineers, and decision-makers with access to the rational improvement strategies of new products. Therefore, the research method can help to drive the entire product iteration process from the perspective of user needs elicitation. The framework is expected to be regarded as the premise and foundation to support the generation of user needs in new product iteration.

Admittedly, this paper is currently an exploratory study, so it is recommended to take an in-depth look at future work. (1) Take advantage of the knowledge graph to solve the problem that the tokenization of Chinese text is prone to ambiguity. (2) Make use of sentiment analysis to automatically discriminate the sentiment polarity of key PIPs, providing a more specific reference for user needs correction. (3) In terms of the whole process of new product iteration, how to translate key user needs into product characteristics is where the research will go next.

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.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgement

This work was supported by the research and development of the prototype database for classic cultural elements of Chinese design style [grant number D5130200012].

References

[1] C.H. Lee, C.H. Chen, Y.C. Lee, Customer requirement-driven design method and computer-aided design system for supporting service innovation conceptualization handling, Adv. Eng. Inf. 45 (2020), 101117, https://doi.org/10.1016/j.aei.2020.101117.
[2] J. Cong, P. Zheng, Y. Bian, C.H. Chen, J. Li, X. Li, A machine learning-based iterative design approach to automate user satisfaction degree prediction in smart product-service system, Comput. Ind. Eng. 165 (2022), 107939, https://doi.org/10.1016/j.cie.2022.107939.
[3] Z. Wang, C.H. Chen, X. Li, P. Zheng, L.P. Khoo, A context-aware concept evaluation approach based on user experiences for smart product-service systems design iteration, Adv. Eng. Inf. 50 (2021), 101394, https://doi.org/10.1016/j.aei.2021.101394.
[4] X. Li, C.H. Chen, P. Zheng, Z. Jiang, L. Wang, A context-aware diversity-oriented knowledge recommendation approach for smart engineering solution design-ScienceDirect, Knowl.-Based Syst. 215 (2021), 106739, https://doi.org/10.1016/j.knosys.2021.106739.
[5] Z. Wang, C.H. Chen, P. Zheng, X. Li, W. Song, A hypergraph-based approach for context-aware smart product-service system configuration, Comput. Ind. Eng. 163 (2022), 107816, https://doi.org/10.1016/j.cie.2021.107816.
[6] Z. Wang, C.H. Chen, P. Zheng, X. Li, L.P. Khoo, A graph-based context-aware requirement elicitation approach in smart product-service systems, Int. J. Prod. Res. 59 (2021) 635–651, https://doi.org/10.1080/00207543.2019.1702227.
[7] R.J. Calantone, R. Yeniyurt, R.D. Townsend, R.B. Schmidt, The effects of competition in short product life-cycle markets: the case of motion pictures*, J. Prod. Innovat. Manag. 27 (2010) 349–361, https://doi.org/10.1111/j.1540-5885.2010.00721.x.
[8] Y. Dong, Y. Fu, L. Wang, Y. Chen, J. Li, A sentiment analysis method of capsule network based on BiLSTM, IEEE Access 8 (2020) 37014–37020, https://doi.org/10.1109/ACCESS.2020.2973711.
[9] D. Poshyvanyk, A. Marcus, Using information retrieval to support design of incremental change of software, in: Proceedings of the Twenty-second IEEE/ACM International Conference on Automated Software Engineering, 2007, pp. 563–566. doi:10.1145/1321631.1321743.
[10] S. Li, D. Tang, J. Yang, Q. Wang, I. Ullah, H. Zhu, A novel approach for capturing and evaluating dynamic consumer requirements in open design, Adv. Eng. Inf. 39 (2018) 95–111, https://doi.org/10.1016/j.aei.2018.12.002.
[11] Z. Wang, C.H. Chen, P. Zheng, X. Li, L.P. Khoo, A novel data-driven graph-based requirement elicitation framework in the smart product-service system context, Adv. Eng. Inf. 42 (2019), 100983, https://doi.org/10.1016/j.aei.2019.100983.
[12] Z. Wang, P. Zheng, X. Li, C.H. Chen, Implications of data-driven product design: From information age towards intelligence age, Adv. Eng. Inf. 54 (2022), 101793, https://doi.org/10.1016/j.aei.2022.101793.
[13] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, Digital twin-driven product design, manufacturing and service with big data, Int. J. Adv. Manuf. Technol. 94 (2018) 3563–3576, https://doi.org/10.1007/s00170-017-0233-1.
[14] Y. Shi, Q. Peng, A spectral clustering method to improve importance rating accuracy of customer requirements in QFD, Int. J. Adv. Manuf. Technol. 107 (2020) 2579–2596, https://doi.org/10.1007/s00170-020-05204-1.
[15] G. Özdağoğlu, A. Kapucugil-İkiz, A.F. Çelik, Topic modelling-based decision framework for analysing digital voice of the customer, Total. Qual. Manag. Bus. 29 (2018) 1545–1562, https://doi.org/10.1080/14783363.2016.1273106.
[16] A.M. Bigorra, O. Isaksson, M. Karlberg, Aspect-based Kano categorization, Int. J. Inform. Manage. 46 (2019) 163–172, https://doi.org/10.1016/j.ijinfomgt.2018.11.004.
[17] Y. Shi, Q. Peng, Enhanced customer requirement classification for product design using big data and improved Kano model, Adv. Eng. Inf. 49 (2021), 101340, https://doi.org/10.1016/j.aei.2021.101340.
[18] T. Hou, B. Yannou, Y. Leroy, E. Poirson, Mining changes in user expectation over time from online reviews, J. Mech. Design. 141 (2019), https://doi.org/10.1115/1.4042793.
[19] J. Qi, Z. Zhang, S. Jeon, Y. Zhou, Mining customer requirements from online reviews: a product improvement perspective, Inf. Manage. 53 (2016) 951–963, https://doi.org/10.1016/j.im.2016.06.002.
[20] Y. Qu, X. Ming, S. Qiu, Z. Liu, X. Zhang, Z. Hou, Integrating fuzzy Kano model and fuzzy analytic hierarchy process to evaluate requirements of smart manufacturing systems, Concurr. Eng. 27 (2019) 201–212, https://doi.org/10.1177/1063293X19845137.
[21] J. Zhang, D. Chen, M. Lu, Combining sentiment analysis with a fuzzy kano model for product aspect preference recommendation, IEEE Access 6 (2018) 59163–59172, https://doi.org/10.1109/ACCESS.2018.2875026.
[22] Y. Huo, J. Liu, J. Xiong, W. Xiao, J. Zhao, Machine learning and CBR integrated mechanical product design approach, Adv. Eng. Inf. 52 (2022), 101611, https://doi.org/10.1016/j.aei.2022.101611.
[23] H. Sun, W. Guo, H. Shao, B. Rong, Dynamical mining of ever-changing user requirements: a product design and improvement perspective, Adv. Eng. Inf. 46 (2020), 101174, https://doi.org/10.1016/j.aei.2020.101174.
[24] F. Zhou, R. Jianxin Jiao, J.S. Linsey, Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews, J. Mech. Des. 137 (2015), 071401, https://doi.org/10.1115/1.4030159.
[25] W.M. Wang, Z. Li, Z.G. Tian, J.W. Wang, M.N. Cheng, Extracting and summarizing affective features and responses from online product descriptions and reviews: a Kansei text mining approach, Eng. Appl. Artif. Intell. 73 (2018) 149–162, https://doi.org/10.1016/j.engappai.2018.05.005.
[26] Q. Zhou, Z. Xu, N.Y. Yen, User sentiment analysis based on social network information and its application in consumer reconstruction intention, Comput. Hum. Behav. 100 (2018) 177–183, https://doi.org/10.1016/j.chb.2018.07.006.
[27] I.P. Delgado-Solano, A.S. Núñez-Varela, H.G. Pérez-González, Keyword Extraction from Users' Requirements Using TextRank and Frequency Analysis, and their Classification into ISO/IEC 25000 Quality Categories, in: 2018 6th International Conference in Software Engineering Research and Innovation, 2018, pp. 88–92. doi:10.1109/CONISOFT.2018.8645870.
[28] J. Joung, H.M. Kim, Automated keyword filtering in LDA for identifying product attributes from online reviews, J. Mech. Des. 143 (2020), 084501, https://doi.org/10.1115/1.4048960.
[29] F. Zhou, J. Ayoub, Q. Xu, X.J. Yang, A machine learning approach to customer needs analysis for product ecosystems, J. Mech. Des. 142 (2019), 011101, https://doi.org/10.1115/1.4044435.
[30] M. Cai, Y. Tan, B. Ge, Y. Dou, Y. Du, PURA: a product-and-user oriented approach for requirement analysis from online reviews, IEEE Syst. J. 16 (1) (2021) 566–577, https://doi.org/10.1109/JSYST.2021.3067334.
[31] J. Buchan, M. Bano, D. Zowghi, P. Volabouth, Semi-automated extraction of new requirements from online reviews for software product evolution, in: 2018 25th Australasian Software Engineering Conference, 2018, pp. 31–40. doi:10.1109/ASWEC.2018.00013.
[32] N.E. Dehaibi, N.D. Goodman, E.F. Macdonald, Extracting customer perceptions of product sustainability from online reviews, J. Mech. Des. 141 (12) (2019), 121103, https://doi.org/10.1115/1.4044522.
[33] R. Santos, E.C. Groen, K. Villela, An overview of user feedback classification approaches, in: Proceedings of the REFSQ Co-Located Events, 2019, pp. CEUR 2376.
[34] M. Lu, L. Peng, Automatic classification of non-functional requirements from augmented app user reviews, in: Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering, 2017, pp. 344–353. doi:10.1145/3084226.3084241.
[35] W.M. Wang, J.W. Wang, Z. Li, Z.G. Tian, E. Tsui, Multiple affective attribute classification of online customer product reviews: a heuristic deep learning method for supporting Kansei engineering, Eng. Appl. Artif. Intell. 85 (2019) 33–45, https://doi.org/10.1016/j.engappai.2019.05.015.

[36] G. Piao, J.G. Breslin, Domain-aware sentiment classification with GRUs and CNNs, Semant. Web Challenges. Cham 927 (2018) 129–139.
[37] A. Timoshenko, J.R. Hauser, Identifying customer needs from user-generated content, Market. Sci. 38 (2019) 1–20, https://doi.org/10.1287/mksc.2018.1123.
[38] Y. Wang, D.Y. Mo, M.M. Tseng, Mapping customer needs to design parameters in the front end of product design by applying deep learning, CIRP Ann.-Manuf. Techn. 67 (1) (2018) 145–148, https://doi.org/10.1016/j.cirp.2018.04.018.
[39] X. Lai, S. Zhang, N. Mao, J. Liu, Q. Chen, Kansei engineering for new energy vehicle exterior design: an internet big data mining approach, Comput. Ind. Eng. 165 (2022), 107913, https://doi.org/10.1016/j.cie.2021.107913.
[40] C.R. Schaffhausen, T.M. Kowalewski, Large-scale needfinding: methods of increasing user-generated needs from large populations, J. Mech. Des. 137 (7) (2015), 071403, https://doi.org/10.1115/1.4030161.
[41] M. Peters, M. Neumann, M. Iyyer, M. Gardner, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, pp. 2227–2237. doi:10.18653/v1/N18-1202.
[42] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423.
[43] L. Wu, A. Fisch, S. Chopra, K. Adams, J. Weston, StarSpace: Embed All The Things!, arXiv preprint arXiv: 1709.03856, 2017. 10.48550/arXiv.1709.03856.
[44] Y. Han, M. Moghaddam, Eliciting attribute-level user needs from online reviews with deep language models and information extraction, J. Mech. Des. 142 (6) (2020), 061403, https://doi.org/10.1115/1.4048819.
[45] R.R. Mekala, A. Irfan, E.C. Groen, A. Porter, M. Lindvall, Classifying user requirements from online feedback in small dataset environments using deep learning, in: 2021 IEEE 29th International Requirements Engineering Conference (RE), IEEE International Requirements Engineering Conference (RE), 2021, pp. 139–149. doi:10.1109/RE51729.2021.00020.
[46] W. Wu, Z. Hong, Y. Feng, M. Li, S. Lou, J. Tan, A semantic analysis-driven customer requirements mining method for product conceptual design, Sci. Rep. 12 (2022) 10139, https://doi.org/10.1038/s41598-022-14396-3.
[47] Y. Sun, S. Wang, Y. Li, S. Feng, H. Wu, ERNIE: Enhanced Representation through Knowledge Integration, arXiv preprint arXiv: 1904.09223, 2019. doi:10.48550/arXiv.1904.09223.
[48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, arXiv preprint arXiv: 1706.03762, 2017. doi:10.48550/arXiv.1706.03762.
[49] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, Natural Language Processing (almost) from Scratch, arXiv preprint arXiv: 1103.0398, 2011. doi:10.48550/arXiv.1103.0398.
[50] M.L.W. Huang, Y. Wang, Sentiment analysis of Chinese e-commerce product reviews using ERNIE word embedding and attention mechanism, Appl. Sci. 12 (2022) 7182, https://doi.org/10.3390/app12147182.
[51] E. Strubell, P. Verga, D. Belanger, A. Mccallum, Fast and Accurate Entity Recognition with Iterated Dilated Convolutions, arXiv preprint arXiv: 1702.02098, 2017. doi:10.48550/arXiv.1702.02098.
[52] T. Zhang, F. Wu, A. Katiyar, K.Q. Weinberger, Y. Artzi, Revisiting Few-sample BERT Fine-tuning, arXiv preprint arXiv: 2006.05987v2, 2021. doi:10.48550/arXiv.2006.05987.
[53] Y. Sun, H. Qiu, Y. Zheng, Z. Wang, C. Zhang, SIFRank: a new baseline for unsupervised keyphrase extraction based on pre-trained language model, IEEE Access 8 (2020) 10896–10906, https://doi.org/10.1109/ACCESS.2020.2965087.