# User experience evaluation method based on online product reviews

Fangmin Cheng[a,*], Suihuai Yu[a], Shengfeng Qin[b], Jianjie Chu[a] and Jian Chen[c]
[a]*Shaanxi Engineering Laboratory for Industrial Design, Northwestern Polytechnical University, Xi'an, China;*
[b]*School of Design, Northumbria University, Newcastle upon Tyne, UK;*
[c]*College of Design and Art, Shaanxi University of Science & Technology, Xian, Shaanxi, China*

**Abstract.** Evaluating the quality of the user experience (UX) of existing products is important for new product development. Conventional UX evaluation methods, such as questionnaire, have the disadvantages of the great subjective influence of investigators and limited number of participants. Meanwhile, online product reviews on e-commerce platforms express user evaluations of product UX. Because the reviews objectively reflect the user opinions and contain a large amount of data, they have potential as an information source for UX evaluation. In this context, this study explores how to evaluate product UX through using online product reviews. A pilot study is conducted to define the key elements of a review. Then, a systematic method of product UX evaluation based on reviews is proposed. The method includes three parts: extraction of key elements, integration of key elements, and quantitative evaluation based on rough number. The effectiveness of the proposed method is demonstrated by a case study using reviews of a wireless vacuum cleaner. Based on the proposed method, designers can objectively evaluate the UX quality of existing products and obtain detailed suggestions for product improvement.

Keywords: User experience (UX) evaluation, Online product reviews, Opinion mining, UX aspect, Product design

## 1. Introduction

As human society gradually entered the era of experience economy, product users are no longer merely satisfied with functional use of products but pursue a good experience when interacting with products [1]. This shift prompts manufacturing companies to pay attention to improving the user experience (UX) of products to fulfil the needs of users. UX is a concept that includes all aspects of how individuals interact with a product [2]. Evidence shows that providing a positive UX can increase user satisfaction and brand loyalty, thus promoting the commercial success of the company [3]. In this context, the UX quality of products has become one of the critical factors for companies to achieve competitive edge in the market [4].

Evaluation of product UX can assist the developing team in solving UX design problems, thereby improving the UX quality of the product [5]. For products under development, designers can identify and address product defects based on UX design evaluation results to improve the UX quality of the products. For products already on the market, designers can identify the defects in current products based on UX evaluation results and improve the UX quality of new products. Previous studies have proposed many UX evaluation methods to measure UX, mainly including questionnaire, interview, and expert review [6]. Although these methods are feasible, their defects are obvious. First, most of these methods require a large amount of intervention by the surveyors in the evaluation process. Surveyors are responsible for setting the experimental environment, recruiting participants, formulating interactive tasks, formulating questionnaire items, and designing interview questions. In this way, user feedback is inevitably affected by surveyors' subjective views, thus affecting the objectivity of UX evaluation results. Second, most of these methods can only recruit a small number of users or experts to evaluate UX, and this low number of partici-

[*] Corresponding author. E-mail: chengfm1991@163.com.

pants may lead to inaccurate evaluation results. In response to these issues, a novel UX evaluation approach is needed.

Online product reviews on e-commerce platforms can be a source of user feedback about the UX of the products already on the market. On e-commerce platforms such as Amazon.com and Taobao.com, customers are encouraged to share their opinions on products by writing reviews. The reviews discuss product functions, product components, and user's opinions, which reflect the UX of the product [7]. In addition, the reviews are written freely by consumers without any purposeful guidance [8]. As a result, the reviews can be considered as a relatively realistic and objective reflection of users' opinions. Meanwhile, the development of e-commerce has attracted a large number of consumers, which makes a massive amount of online product reviews available and can be used as a source of information representing the opinions of a big volume of users. Therefore, the use of large-scale online product reviews for product UX evaluation has the potential to address the problems embedded in the current UX evaluation methods. While many studies focus on the analysis of online product reviews, few studies analyze reviews from the perspective of UX evaluation.

To fill this research gap, this study explores a UX evaluation method based on online product reviews for products already on the market. The rest of this paper is organized as follows. In Section 2, related studies are briefly reviewed. Section 3 defines the key elements of reviews through a pilot study, which lays the foundation for the construction of the evaluation method. Section 4 proposes a systematic UX evaluation method based on online product reviews. In Section 5, the proposed method is demonstrated by a case study using vacuum cleaner reviews. Discussion is stated in Section 6. Conclusion is stated in Section 7.

## 2. Theoretical background

### *2.1. Construction of UX evaluation method*

Many methods have been introduced to evaluate UX. In early UX research, the concept of UX was often mixed with the concept of usability [9]. Therefore, usability measurement methods were also regarded as UX evaluation methods, such as usability reviews [10] and cognitive walkthrough [11]. These evaluation methods cannot provide a broad perspective of the UX and can only be performed by experts rather than users. With the development of the concept of UX, UX evaluation methods with a broader perspective were proposed. The most frequently used method is the standardized questionnaire, in which end-users describe their perception regarding UX aspects. "Standardized" means that these questionnaires are not a more or less random or subjective collection of questions, but result from a careful construction process [12]. The most recognized standardized questionnaires include AttrakDiff [13], UEQ [14], and meCUE [15]. These standardized questionnaires provide a broader view of UX. Moreover, the questionnaire enables users to participate in the evaluation process and obtain quantitative evaluation results. Another kind of questionnaire, the satisfaction scale, is also widely used in product UX analysis [16]. However, questionnaires can only collect subjective data of users. Some researchers introduced psychophysiological and psychometric methods, such as electroencephalography (EEG) [17], electrodermal activity (EDA) and electromyography (EMG) [18] into UX evaluation, thereby providing relatively objective UX evaluation results. However, since psychophysiology and psychometrics methods can only focus on individual sensory experiences, they can only evaluate the UX in a limited dimension, rather than the overall UX. In addition to the above methods, methods such as UX curve [19], user interview and indirect observation [20] are also used to evaluate UX in some studies.

### *2.2. Definition of UX aspects*

For feasibility of the UX evaluation, UX should be divided into multiple dimensions or aspects. At the conceptual level, the UX aspects can be defined in two ways. The first defines UX aspects as components of the user's subjective response to the product. For example, Hassenzahl et al. [21] argued that the UX of interactive products has two aspects: practical quality and hedonic quality. The second defines UX aspects as factors that affect the overall UX. The purpose of this research is to evaluate the UX quality of a product, which is an influencing factor of UX rather than a component of a user's subjective response. Therefore, we define UX aspects according to the second definition.

In pioneering UX studies, academic researchers proposed various UX models that define multiple UX aspects [22, 3]. In the following research on UX evaluation, these models and aspects were adjusted and improved according to diverse research purposes and application scenarios. Laugwitz et al. [14] identified

26 items that were closely related to UX, and grouped them into six aspects, including attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. A UX questionnaire based on these items was eventually constructed. Park et al. [23] defined the UX aspects as usability, affect, and user value. Tuch et al. [24] argued that need fulfilment, technologies involved in the experience, and effect were important UX aspects and analyzed the user-generated narratives to identify the influence of these aspects on positive and negative experiences. Through a literature review, Winckler et al. [25] identified a set of UX aspects that were central for interactive systems, including visual and aesthetic experience, emotion, stimulation, identification, meaning and value, and social relatedness. In addition, context of use [26], brand [27], cultural background [28] and other factors are also defined as UX aspects by some research, but these aspects are not related to the product.

In existing UX questionnaires, aspects are defined based on the opinions of a small number of experts or users, and are applied to UX evaluation of all products. However, different UX aspects are of different importance to different products, and thus it is unreasonable to use a unified set of UX aspects to evaluate different products. Hence, we attempt to extract UX aspects directly from the reviews based on the existing definition of aspects, which is more in line with the characteristics of the product and the views of users.

### *2.3. Mining online product reviews*

With the increasing importance of online product reviews, the amount and diversity of research in this area has increased dramatically. The purposes of these works are to analyze user characteristics [29], provide product purchase suggestions [30], and improve product design [31], etc.

Among the above research, the research with the purpose of improving product design is relatively abundant and has a great correlation with our research. Most such studies focus on summarizing user opinions or identifying customer needs based on attribute identification and sentiment analysis to provide valuable information for designers. Qi et al. [31] performed attribute identification and sentiment analysis, and utilized the KANO model to analyze the data to develop appropriate product improvement strategies. Li et al. [32] proposed a sentiment analysis approach based on Kansei Engineering and machine learning to extract and measure users’ affective responses to products from online reviews. Jin et al. [33] extracted aspects of product features and detailed reasons of consumer dissatisfaction to inform designers regarding what leads to unsatisfied opinions. The researchers conducted identification of product features and the sentiment analysis with the help of pros and cons reviews. Then the approach of conditional random fields was employed to detect aspects of product features and detailed reasons. Jin et al. [34] proposed a framework to identify comparative customer requirements from product online reviews for competitor analysis. Li et al. [35] proposed a model for identifying critical customer requirements based on online reviews and KANO model. Yang et al. [36] proposed a methodology of establishing a UX knowledge base from online customer reviews to support UX-centered design activities. Their work provided an approach to automatically discover valuable UX information from online reviews.

Although much research on online product reviews has aimed at design improvement, only a few studies have been based on the UX perspective or proposed operable UX evaluation methods.

## 3. Definition of key elements of reviews

The purpose of this research is to propose a quantitative evaluation method for UX quality of existing products based on a massive amount of online product reviews to help designers improve the design of new products. The data source is Chinese reviews on Chinese e-commerce platforms. For the sake of understanding, the sentences and words of Chinese reviews in this paper are translated into English without affecting the accuracy of the described research process.

To explore what are key elements of the reviews can be used for our evaluation, a pilot study was performed. We randomly extracted 100 reviews from three notable Chinese e-commerce platforms (JD.com, Taobao.com, and Suning.com) for three different products (vacuum cleaner, mobile phone, and shoe cabinet), with a total of 300 reviews. We finally obtained 1023 review sentences after sentence segmentation. We labelled the words and phrases that represent the key elements from each review.

Four key terms related to evaluation emerged, which are called product feature term, UX aspect term, attitude term, and degree term. These terms are generally in line with the elements embedded in existing UX questionnaires in terms of product features, UX aspects, and user scores. The product features are the elements of the product itself, including components, functions, and attributes (appearance, size, and material). The user scores represent the user's affective response to the products, which can be expressed by attitude terms and degree terms.

The annotation results of the two review sentences are shown in Figure 1 as examples. We can see that the affective response is reflected by both the attitude term and the degree term. The attitude term represents the polarity of the affective response, and can be divided into positive, neutral or negative. The degree term represents the intensity of the affective response. As shown in Figure 1, the affective responses of review sentences are "very fast" and "very efficient". Although "fast" and "efficient" have different meanings, according to the meanings of the sentences, they represent the same affective response. In addition, we find while that many review sentences do not explicitly contain UX aspect terms, the attitude terms imply UX aspects. As shown in Figure 1, the two review sentences express the same meaning. Review sentence 1 indicates that the UX aspect is "speed", which can be equally considered as "efficiency". Review sentence 2 does not directly indicate the UX aspect, but the attitude term "efficient" implies that the UX aspect being evaluated is "efficiency".

According to statistics, 86% of review sentences contain at least two key elements. For 74% of the review sentences, the key elements in the review can be transformed into the elements of the evaluation system through element mapping. This proves that it is possible to extract four elements from reviews to evaluate product UX. Therefore, we construct the UX evaluation method based on the key elements of reviews. Details of the proposed evaluation method are presented in the next section.

## 4. Product UX evaluation method based on online product reviews

Figure 2 illustrates the framework of the proposed UX evaluation method based on online product reviews. It includes three parts, Part 1 is extraction of key elements in the review text, Part 2 is integration of key elements, and Part 3 is quantitative evaluation.

### *4.1. Extraction of key elements*

In Part 1, we propose a key elements extraction method, consisting of six steps:

(1) Review collection. Collect review data from e-commerce platforms.

(2) Review pre-processing. NLP tools are used to pre-process the review data. Pre-processing includes filtering out fake information, sentence segmentation, Chinese word segmentation, part of speech (POS) tagging, and semantic dependency parsing (SDP).

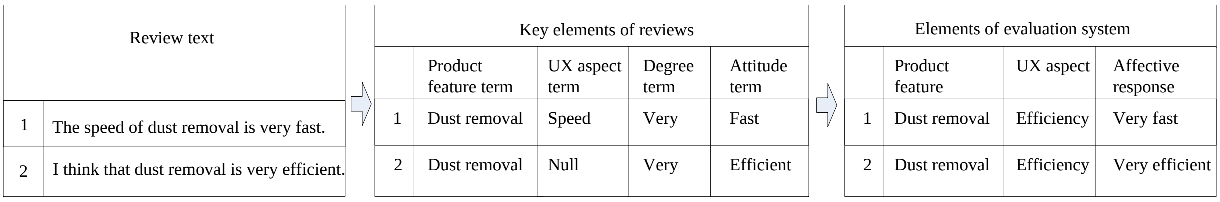
(3) Attitude term extraction. We extract the words that describe the user's attitude in the reviews. We construct a sentiment lexicon to extract attitude terms and determine the attitude polarities of the reviews.

(4) Degree term extraction. The degree terms in the reviews are extracted by using the available degree word lexicon [32].

(5) Product feature term extraction. Critical product feature terms are extracted based on the POS and word frequency of the candidate words.

(6) UX aspect term extraction. UX aspect terms are extracted by analyzing the semantic dependencies between candidate words and product feature terms and attitude terms.

Steps (1), (2), and (4) are relatively common and simple, so we illustrate steps (3), (5), and (6) in further detail.



| Review text | |
|---|---|
| 1 | The speed of dust removal is very fast. |
| 2 | I think that dust removal is very efficient. |

| Key elements of reviews | | | | |
|---|---|---|---|---|
| | Product feature term | UX aspect term | Degree term | Attitude term |
| 1 | Dust removal | Speed | Very | Fast |
| 2 | Dust removal | Null | Very | Efficient |

| Elements of evaluation system | | | |
|---|---|---|---|
| | Product feature | UX aspect | Affective response |
| 1 | Dust removal | Efficiency | Very fast |
| 2 | Dust removal | Efficiency | Very efficient |

Fig.1 Key elements of reviews for evaluation.

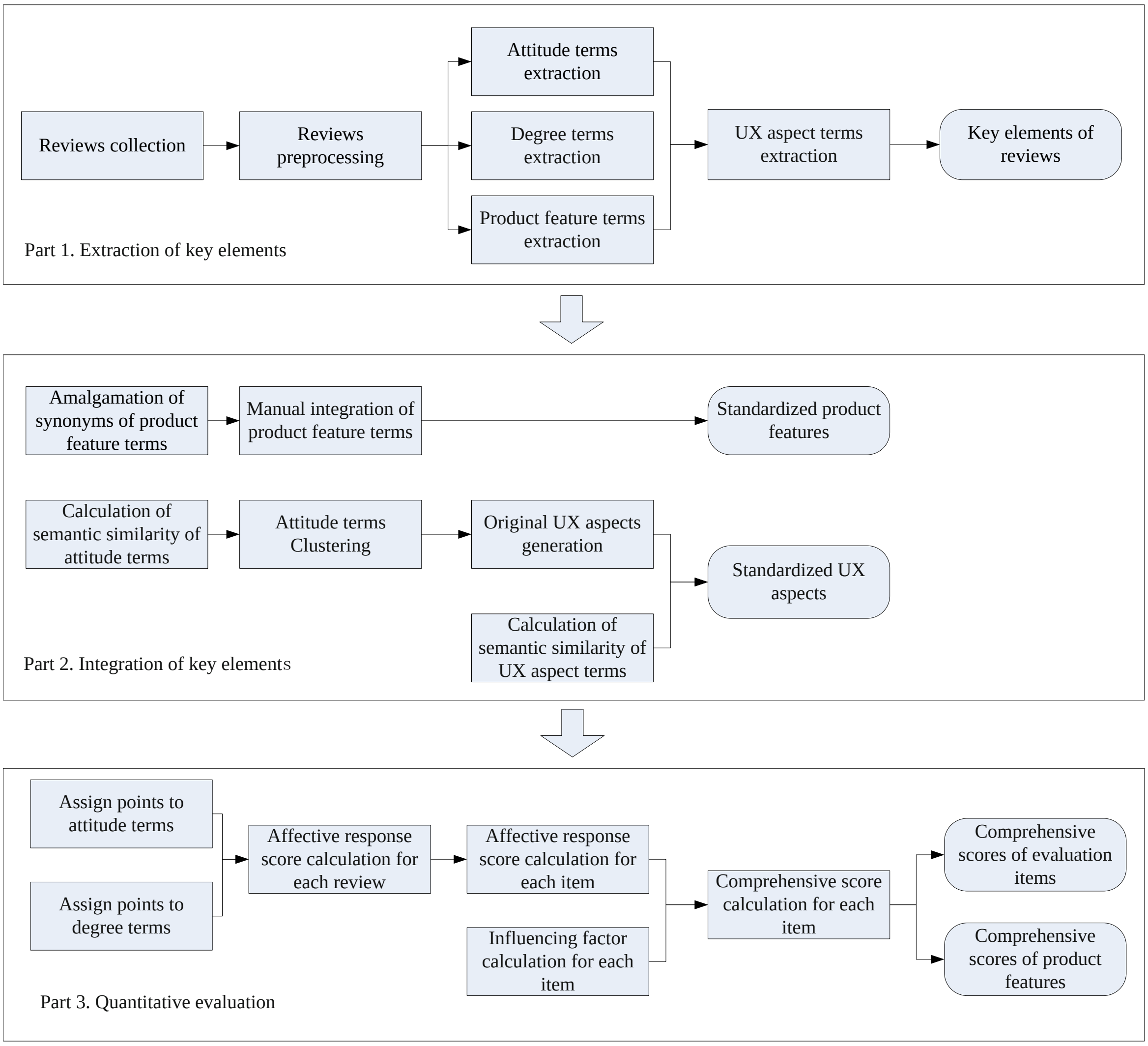

Reviews collection
Reviews preprocessing
Attitude terms extraction
Degree terms extraction
Product feature terms extraction
UX aspect terms extraction
Key elements of reviews
Part 1. Extraction of key elements
Amalgamation of synonyms of product feature terms
Manual integration of product feature terms
Standardized product features
Calculation of semantic similarity of attitude terms
Attitude terms Clustering
Original UX aspects generation
Standardized UX aspects
Calculation of semantic similarity of UX aspect terms
Part 2. Integration of key elements
Assign points to attitude terms
Assign points to degree terms
Affective response score calculation for each review
Affective response score calculation for each item
Influencing factor calculation for each item
Comprehensive score calculation for each item
Comprehensive scores of evaluation items
Comprehensive scores of product features
Part 3. Quantitative evaluation

Fig. 2 Architecture of proposed approach.

### *4.1.1. Attitude term extraction*

Since attitude term extraction is a kind of sentiment analysis, we utilize sentiment lexicon, a common sentiment analysis method, to extract and analyze customer attitude. Although some general sentiment lexicons are available, their accuracy is relatively low because the polarity of sentiment words may vary in different domains. Therefore, we construct a domain-specific sentiment lexicon to extract the sentiment words in the reviews and determine their polarity.

A seed sentiment lexicon is first constructed and then expanded by adding synonyms to form the sentiment lexicon. Previous work suggests that the words expressing attitude are mainly adjectives [37]. Therefore, $k$ adjectives with the highest term frequency are extracted directly from the review corpus as seed sentiment words. According to the type of product, seed words are manually divided into positive, neutral or negative sets. Then, using synonym lexicon and antonym lexicon, synonyms and antonyms of seed words are added to the corresponding word sets to form the final sentiment lexicon.

The attitude terms are extracted by checking whether the words in the review sentence are contained in the sentiment lexicon. Then, the negative words set is used to check whether there are negative words around the attitude words. Finally, the attitude of the user is determined according to the polarity of the attitude word and whether the negative word exists. Since the user's attitude is the user evaluation opinion on the product UX, if a review sentence does not contain attitude terms, we assume the review sentence is useless and exclude it.

#### *4.1.2. Product feature term extraction*

Different product features have different influences on UX. It is efficient and reasonable for enterprises to improve the quality of product UX by improving the critical features of a product rather than all of them. Therefore, we only extract critical product features and ignore non-critical features.

We assume that product features that appear frequently in the reviews are critical features, and therefore we construct the product feature set based on the term frequency in reviews. A previous study shows that product features are generally nouns and noun phrases [38]. Thus, all nouns and noun phrases in the reviews are extracted as candidate words. There are many non-product feature nouns that appear frequently in reviews, such as brand nouns, personal nouns, proper nouns, and so on. Based on experience, a removing word set is constructed to filter the candidate words. After that, a threshold is set to remove words with low frequency. The words whose term frequency is higher than the threshold are the critical product feature words. Many product features have multiple expressions, and thus synonyms are added to the word set; this ensures that some uncommon expressions of critical product features can also be extracted. Thus, the final product feature word set is constructed.

While this method is unable to extract the product feature terms with low term frequency, our purpose is not to exhaust all the product features but to extract the key ones, thus it has little or no impact on the UX evaluation results.

#### *4.1.3. UX aspect term extraction*

The UX aspect terms appear relatively infrequently in reviews, and therefore cannot be extracted based on term frequency. From the pilot study, we find that UX aspect terms have close semantic relationships with product feature terms and attitude terms, and hence we construct the extraction method based on the semantic relationship between terms.

SDP is a technique used to analyze the semantic relations among the language units of a sentence [39]. SDP describes a word through a semantic framework and is not affected by syntactic structure. To explore significant semantic relationships between different terms, we used LTP, an NLP tool, to perform SDP for the reviews in the corpus of the pilot study. LTP (Language Technology Platform) is a set of efficient and high-precision Chinese natural language processing platform developed by HIT Social Computing and Information Retrieval Research Center (HIT-SCIR). It has become the most influential Chinese processing platform. Taking two typical review sentences as an example, the SDP results show the most significant semantic relationships of UX aspect terms, as shown in Figure 3. From the SDP results, we find significant semantic dependencies between UX aspect terms and product feature terms and attitude terms. The most significant semantic relationship between UX aspect terms and product feature terms is "Feature", tagged "FEAT", in which UX aspect terms are semantically dependent on product feature terms. In the corpus, this type of relationship accounts for 82% of the semantic relationships between the two kinds of terms. There are two significant semantic relationships between UX aspect terms and attitude terms. One is "Experiencer", tagged "EXP", in which attitude terms are semantically dependent on UX aspect terms. The other is "Feature", tagged "FEAT", in which UX

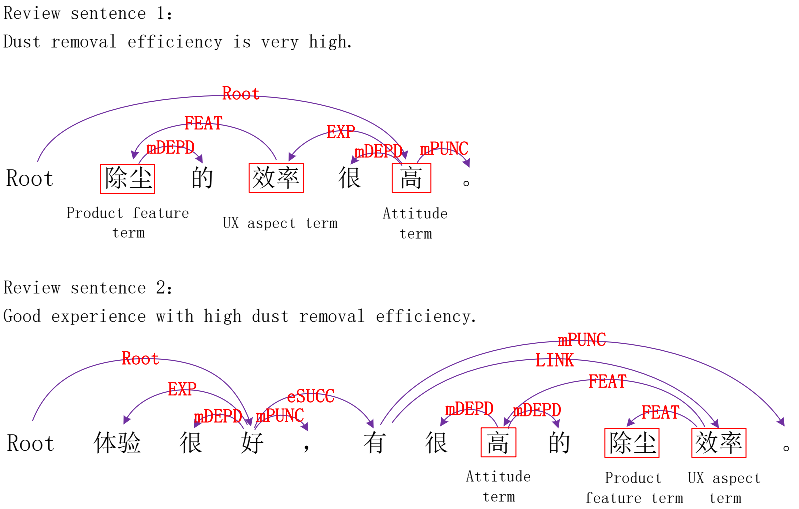


Fig. 3. Examples of SDP results of UX aspect terms.

aspect terms are semantically dependent on attitude terms. These two relationships account for 67% and 25% of the semantic relationships of the two kinds of terms, respectively. Therefore, we assume that a word is a UX aspect term if the above semantic dependency relations account for a high proportion of all its semantic relationships. Furthermore, UX aspect terms are extracted based on the semantic dependency of candidate words.

The words representing UX aspects are nouns. Exclude the product feature terms and removing words defined in Section 4.1.2, and the remaining nouns are selected as the candidates for UX aspect terms. List all semantic dependencies and related words of a candidate word $s$. Calculate the number of occurrences of words in the reviews that have FEAT relationship with $s$ and belong to the product feature term set $F$: $n\big(FEAT(s,F)\big)$. $n\big(EXP(A,s)\big)$ and $n\big(FEAT(s,A)\big)$ are calculated in a similar way. According to the method proposed in [36], we propose an index to judge whether a candidate word is a UX aspect term. The principle of index calculation is based on the probability that the candidate word has characteristic semantic relations with product feature terms and attitude words. The index of the semantic relationship between candidate word $s$ and product feature terms and attitude terms is defined as

$$index(s) = \ln\left(1+\frac{n\big(FEAT(s,F)\big)+1}{n(s)+1}\frac{n\big(EXP(A,s)\big)+n\big(FEAT(s,A)\big)+1}{n(s)+1}\right) \quad (1)$$

where $n(s)$ is the number of candidate words $s$ that appear in the corpus.

Then, whether the candidate word $s$ is a UX aspect term is determined by

$$Category(s) = \begin{cases} UX\ aspect\ term & if\ index(s) \geq \beta \\ unrealted\ word & if\ index(s) < \beta \end{cases} \quad (2)$$

where $\beta$ is a threshold.

The extraction result can be changed by adjusting the threshold. When $\beta$ is large, the precision is high, but the recall is low. On the contrary, when $\beta$ is low, the precision is low, but the recall is high.

### *4.2. Integration of product features and UX aspects*

A pair of a product feature and a UX aspect constitute an evaluation item. Thus, it is necessary to integrate these two kinds of terms to form a standardized evaluation index system.

Since a product feature can be represented by multiple terms, it is necessary to integrate the different terms that represent the same product feature into the most common terms. Since a consumer product usually does not contain too many product features, and the extraction method of product feature terms filters out many low-frequency terms, it is possible to manually integrate product feature terms without too much workload.

There are two term sets containing information about UX aspects. One is the UX aspects explicitly defined in the UX aspect term set, and the other is the UX aspects that are implicit in the attitude term set. The UX aspects implicitly expressed in the attitude terms need to be converted into UX aspect terms, and then the two types of UX aspect terms should be integrated into a standardized UX aspect set.

Due to the quantity and complexity of the attitude terms, they are first clustered according to semantic similarity. We calculate the semantic similarity between terms based on Hownet (http://www.keenage.com/). Hownet is a knowledge base used to describe the concepts represented by Chinese and English words and their relationships. Details of the semantic similarity algorithm for terms can be found in [40]. In the sentiment lexicon, each word in the positive (negative) set has an antonym in the negative (positive) set, and a pair of antonyms represents opposite opinions on the same UX aspect. Therefore, to obtain the UX aspects implicitly expressed in attitude terms, it is enough to only cluster the words that are in the positive set. After we calculate the semantic similarity between each word in the positive set, the spectral clustering algorithm [41] is used to cluster the terms. After clustering the words in the positive set, the cluster name is manually defined by suitable UX aspect terms according to the semantics of the words contained in each cluster and the UX aspects defined by the existing research. The cluster names are the original UX aspects. All terms unrelated to any UX aspect are treated as a cluster, and the cluster name is defined as “Undefinable aspect”; this is treated as a special UX aspect.

Then, combine UX aspect terms extracted from reviews with the original UX aspects. The semantic similarity between each extracted UX aspect term and initial UX aspect is calculated, and the threshold *T* is set. If the semantic similarity between a term and an original UX aspect is the largest, and the semantic similarity is greater than *T*, the term is classified as the UX aspect. If the semantic similarity between a term and all original UX aspects is less than *T*, the term is added as a new UX aspect. In this way, the final UX aspects are obtained, which can greatly represent the UX aspect terms extracted from the reviews and the UX aspects involved in the attitude terms.

### *4.3. Quantitative evaluation of product UX based on rough number*

Through the extraction and integration of the key elements of the review, each online product review sentence is converted into a quadruple. On this basis, the quantitative evaluation method for evaluating an item is proposed. First, assign values to attitude terms and degree terms, and obtain a single score for the affective response of each review data. Second, by comprehensively considering both the user affective response and user attention, the performance of each product feature in each UX aspect is quantitatively evaluated based on rough number. Finally, product improvement strategies are put forward based on the evaluation results.

#### *4.3.1. Affective response score of each review sentence*

To quantitatively measure the customers’ affective response, we assign points to attitude terms and degree terms. A 7-point SD scale is used to measure the affective response. The degree terms are divided into three levels according to their intensity, i.e., extremely high, slightly high, and general, and are given scores of 3, 2, and 1, respectively. For the attitude terms, positive, neutral, and negative terms are given scores of -1, 0, and 1, respectively. Assume that for a review sentence *R*, the score of the degree term is $D_r$, and the score of the attitude term is $A_r$. Then, the score of the affective response can be calculated as

$$S_r = D_r \times A_r \tag{3}$$

Thus, the affective response of each review sentence can be measured by an integer score of -3 to 3. A score of -3 means that the UX is extremely bad while a score of 3 means that the UX is extremely good.

#### *4.3.2. Quantitative evaluation based on rough number*

Based on the above steps, the affective response of each review sentence, that is, the customers’ affective response to the UX aspects of the product features in each review sentence, can be measured by a single score. However, the users’ affective response is fuzzy and uncertain, which is difficult to be reflected by a single score. Therefore, rough number [42] is used to express the single score in the form of interval number to reduce the influence of the fuzziness and uncertainty of the affective response on the evaluation results.

Assume that there are *n* product features ($F_i$, $i$=1,2,⋯,$n$) and a total of *m* UX aspects ($A_j$, $j$= 1,2, ⋯,$m$). Then, consider a pair of $F_i$ and $A_j$ as an evaluation item $I_{ij}$, and assume that the amount of relevant review data is *l*. If $l$ =0, the corresponding score is expressed as rough number $S_{ij}$= [0,0]. If $l \neq 0$, for a review score $x_{ijh}$ ($x_{ijh} \in$ {-3,-2,-1,0,1,2,3}, $h$=1,2, ⋯,$l$), calculate the lower approximation $\underline{Apr}(x_{ijh})$ and the upper approximation $\overline{Apr}(x_{ijh})$, and then calculate the lower limit $\underline{\lim}(x_{ijh})$ and the upper limit $\overline{\lim}(x_{ijh})$. The single score $x_{ijh}$ can be expressed as rough number $[\underline{\lim}(x_{ijh}), \overline{\lim}(x_{ijh})]$. Details of the computational formulas can be found in [42].

Thus, the total score for $I_{ij}$ can be calculated based on the rough number of each single score. According to [42], the lower limit is

$$\underline{\lim}(x_{ij}) = \frac{1}{l}\sum_{h=1}^{l} \underline{\lim}(x_{ijh}) \tag{4}$$

The upper limit is

$$\overline{\lim}(x_{ij}) = \frac{1}{l}\sum_{h=1}^{l} \overline{\lim}(x_{ijh}) \tag{5}$$

The total score can be expressed by rough number as

$$S_{ij} = [\underline{\lim}(x_{ij}), \overline{\lim}(x_{ij})] \tag{6}$$

The mean value of $S_{ij}$ can be calculated as

$$\mathrm{M}(\mathrm{S_{ij}}) = \frac{\underline{\mathrm{lim}}(\mathrm{x_{ij}})+\overline{\mathrm{lim}}(\mathrm{x_{ij}})}{2} \tag{7}$$

For $I_{ij}$, $M(S_{ij})$ is the final score of the affective response.

It is obvious that different product features and UX aspects have different impacts on UX. However, this phenomenon is ignored by most existing UX evaluation methods, which treat all product features and UX aspects as equal. We assume that an evaluation item has higher user attention and thus greater impact on UX if it has a larger number of related reviews. To evaluate the quality of a UX aspect of a product feature, we should not only consider the affective response, but also consider the user attention. Therefore, the relative number of reviews is taken as the influencing factor of each affective response score. The influencing factor of $w(S_{ij})$ can be calculated as

$$w(S_{ij}) = \frac{N(I_{ij})}{N_{max}} \tag{8}$$

where $N_{max}$ is the number of reviews obtained by the product feature that receives most review sentences. $N(I_{ij})$ is the amount of review data related to $I_{ij}$.

Thus, the comprehensive score of $I_{ij}$ is calculated as

$$CS(I_{ij}) = w(S_{ij}) \times M(S_{ij}) \tag{9}$$

The comprehensive score of product feature $F_i$ is calculated as

$$CS(F_i) = \sum_{j=1}^{m} CS(I_{ij}) \tag{10}$$

### *4.3.3. Developing product improvement strategies*

The product features are ranked based on the comprehensive score. The high-ranked product features have good UX quality, and the design of these product features should be maintained in new products. The low-ranked product features have poor UX quality, and the score of each aspect should be analyzed in detail to indicate the specific problems of the product features. In the new products, design improvements should be made to address the issues corresponding to the specific UX aspect. Furthermore, the rough number $[\underline{\lim}(x_{ij}), \overline{\lim}(x_{ij})]$ of each evaluation item also needs to be analyzed. If $\underline{\lim}(x_{ij}) < 0$ and $\overline{\lim}(x_{ijh}) > 0$, it means that customers have different opinions on the evaluation item, thus various design schemes should be considered in new product development to meet the diversified needs of users.

## 5. Case study

In this section, a case study was implemented to illustrate the proposed approach. Household appliances are common consumer products, and thus customers are familiar with the product features so that they can provide valuable reviews. Therefore, we chose a wireless vacuum cleaner (Puppy T10 young) as the target product. We chose JD.com, one of the largest e-commerce platforms in China, as the source of reviews. JD.com claims that only consumers who have purchased products can post reviews, which guarantees the authenticity of the reviews.

### *5.1. Extracting critical elements*

A total of 817 reviews were collected through a web crawler. Then the review data was pre-processed. In this case study, we used LTP for sentence segmentation, word segmentation, POS tagging, and SDP. We filtered out review sentences that had too few Chinese characters (less than five characters), and finally obtained 3277 review sentences.

Then, we extracted the key elements in the reviews are extracted. For attitude terms, adjectives with a frequency greater than five were used as seed words to construct the sentiment lexicon. For product feature terms, 1% of the frequency of the highest frequency product feature term was used as the threshold to filter out the low-frequency term. For UX aspect terms, the threshold $\beta$=0.4 was set to maximize the *F* score. Finally, a total of 378 attitude terms, 58 degree terms, 102 product features terms, and 21 UX aspect terms were extracted. The review sentences without attitude terms were discarded, leaving 2,765 review sentences.

Two PhD students in industrial design manually annotated the key elements in the reviews. The annotation scheme is introduced in Section 3. Each review sentence was annotated separately by the two annotators. Conflicts of review annotation were resolved through discussion. The result of manual annotation served as the baseline for element extraction. Three widely utilized classification evaluation metrics were employed to evaluate the extraction performance of the method, including recall, precision, and *F* score. The results are shown in Table 1.

As seen from Table 1, the proposed element extraction method achieves acceptable results. In terms of the extraction of attitude terms and degree terms, a relatively higher performance is achieved. For product feature terms, the recall rate is relatively low. This is mainly because the product feature terms with low term frequency are not extracted. The results show that the product feature term with the highest term frequency is “dust removal” and the term frequency is 604. Product feature terms whose term frequency is higher than 1% of the maximum term frequency, i.e., the term frequency is higher than 6, are extracted. Therefore, the low recall for product feature terms has little negative impact on the final evaluation results.

Table 1

Results of element extraction

| | Recall | Precision | *F* score |
|---|---|---|---|
| Attitude terms | 0.837 | 0.896 | 0.865 |
| Degree terms | 0.962 | 0.942 | 0.952 |
| Product feature terms | 0.730 | 0.922 | 0.815 |
| UX aspect terms | 0.685 | 0.742 | 0.712 |

For UX aspect terms, the performance of the extraction method is relatively poor. It is found from the extraction results that many unrelated words have the semantic dependency adopted by the extraction method. For instance, some review sentences said that “the surface of *product feature term* is *attitude term*”. The word “surface” has a FEAT relationship with a product feature term and an EXP relationship with an attitude term and was not filtered out by removing words. The index of “surface” was then calculated as 0.53, and thus was extracted as a UX aspect term. This leads to low precision. Meanwhile, there are some unusual semantic dependencies of UX aspect terms in sentences, especially those with grammatical problems. For instance, a review sentence said that “the dust removal function is satisfactory, especially the efficiency”. The UX aspect term “efficiency” has no semantic dependency with the product feature term and the attitude term. This makes it possible that some low-frequency terms cannot be extracted, which decreases recall. This problem leaves some space for developing more sophisticated algorithms to improve the performance of the extraction methods.

Table 2

UX aspects, corresponding attitude terms and UX aspect terms

| UX aspect | Positive attitude terms | Negative attitude terms | UX aspect terms |
|---|---|---|---|
| Efficiency | Efficient, fast, quick | Inefficient, slow, dilatory | Efficiency, speed |
| Hygiene | Clean, neat, hygienic | Dirty, shabby, unhygienic | Hygiene, cleanliness |
| Dependability | Secure, dependable, durable | Dangerous, trustless, fragile, shoddy | Dependability, solidness, Durability |
| Learnability | Easy, simple, legible | Difficult, complicated, ambiguous | Complexity, perspicuity |
| Comfortability | Comfortable, relief, suitable | Discomfort, afflictive, awkward | Comfortability, suitability |
| Attractiveness | Attractive, fascinating, interesting, novel | Boring, tiresome, mediocre | Attractiveness, enjoyment |
| Undefinable aspect | Good, excellent, superior, advanced | Bad, terrible, inferior, backward | |

Table 3

Scores of degree terms

| Level | Score | Degree terms |
|---|---|---|
| High | 3 | Very, quite, highly, really, super, especially, extremely |
| Slightly high | 2 | Slightly, more or less, nearly, little, a bit, in some degree |
| General | 1 | Modest, moderate, ordinary or not mentioned |

Table 4

Scores of attitude terms

| Sentiment polarity | Score | Attitude terms |
|---|---|---|
| Positive | 1 | Good, excellent, superior, advanced, attractive, safe, clean |
| Neutral | 0 | Common, ordinary, mediocre, moderate, conventional |
| Negative | -1 | Bad, terrible, inferior, backward, boring, tiresome, mediocre |

Table 5

Rough numbers of evaluation items

| | Dust removal | Cleaning head | Handle | Shape | Dust bag |
|---|---|---|---|---|---|
| Efficiency | [1.52, 2.58] | [1.22, 2.36] | [1.00, 1.00] | [0, 0] | [0.94, 2.11] |
| Hygiene | [0.97, 2.12] | [1.06, 2.23] | [-0.90, -0.54] | [1.04, 2.33] | [-1.23, -0.72] |
| Dependability | [1.05, 2.50] | [1.44, 2.71] | [-2.48, -1.42] | [3.00, 3.00] | [0.94, 2.11] |
| Learnability | [2.15, 2.78] | [1.72, 2.15] | [0.92, 2.16] | [0.86, 2.02] | [1.94, 2.32] |
| Comfortability | [1.01, 2.03] | [-0.78, 1.36] | [-2.70, -1.68] | [1.79, 2.52] | [-0.61, 1.11] |
| Attractiveness | [0, 0] | [-1.75, -1.25] | [0, 0] | [2.34, 2.69] | [0, 0] |
| Undefinable aspect | [1.14, 1.98] | [1.64, 2.33] | [-2.11, -0.64] | [1.66, 2.02] | [-0.45, 1.04] |

Table 6

Five product features with the worst UX quality and their comprehensive scores

| | Handle | Battery | Noise | Wheel | Weight |
|---|---|---|---|---|---|
| Efficiency | 0.005 | -1.960 | 0.000 | -0.457 | -0.219 |
| Hygiene | -0.070 | -0.018 | 0.000 | -0.100 | 0.000 |
| Dependability | -1.03 | -0.660 | -0.090 | -0.360 | 0.025 |
| Learnability | 0.118 | 0.055 | 0.000 | 0.044 | 0.000 |
| Comfortability | -1.58 | -0.157 | -0.240 | 0.183 | -0.460 |
| Attractiveness | 0.000 | 0.000 | 0.000 | 0.748 | 0.000 |
| Undefinable aspect | -0.346 | -0.155 | -0.698 | -0.961 | -0.233 |
| Comprehensive score | -2.903 | -2.895 | -1.368 | -0.903 | -0.887 |

### *5.2. Integration of product features and UX aspects*

The product feature terms were integrated. Synonyms in product feature term set were merged firstly. Then, according to the semantics of product feature terms and the design knowledge of vacuum cleaner, the terms were further combined manually. Finally, 42 product features are obtained, including 21 components, 13 functions and 8 attributes.

Then, the UX aspect terms were integrated. The similarity between attitude words in the positive set was calculated firstly. Seven clusters were obtained by spectral clustering, and the cluster names were defined as original UX aspects. Attitude terms in the negative set were mapped to the corresponding UX aspect according to their antonyms in the positive set. Then, the original UX aspects and UX aspect terms were combined based on semantic similarity. All UX aspect terms were categorized into the original UX aspects, resulting in seven UX aspects. Table 2 presents the UX aspects along with some of their corresponding attitude terms and UX aspect terms.

### *5.3. UX quantitative evaluation*

Scores were assigned to degree terms and attitude terms, and the comprehensive score of each review sentence was calculated. The scores of some terms are shown in Tables 3 and 4.

The rough number of each evaluation item was calculated according to the steps defined in Section 4.3.2. Take “hygiene” (UX aspect) of the “cleaning head” (product feature) as an example. There were six review sentences related to this evaluation item, and their scores were 2, 2, 3, 1, 2, and 0. Then the rough numbers of these scores were [1.4, 2.24], [1.4, 2.24], [1.67, 3], [0.5, 2], [1.4, 2.24], and [0, 1.67], respectively. Thus, the rough number of this evaluation item was [1.06, 2.23]. The mean value of the rough number was 1.645. Finally, the rough numbers of all evaluation items were calculated, as shown in Table 5. Limited by the length of the paper, only the rough numbers of 10 product features with the largest amount of relevant review data are listed.

The number of reviews related to each item can also be counted. According to the statistical results, the item with the largest number of review sentences was "efficiency" of "dust removal", with a data volume of 195. Then, the influencing factor of each evaluation item was calculated based on the number of review sentences. After the comprehensive scores of each item and each product feature were obtained according to the method in Section 4.2.3, the product features were ranked according to their comprehensive scores. Five product features with the worst UX quality are listed in Table 6. As seen from Table 6, handle, battery, etc. are the product features with the worst UX quality.

We can then analyze the product features with poor UX quality in detail. Taking the wheel as an example, according to the comprehensive score of each UX aspect, its efficiency, dependability, and hygiene are relatively poor. In the development of new products, design methods such as quality function deployment (QFD), should be implemented to improve the design of the wheel in these aspects. The rough number of the hygiene of wheel is [-2.12, 0.26], where -2.12 < 0 and 0.26 > 0. This indicates that customers have different opinions on the hygiene of the wheel, and thus multiple design schemes should be considered in new product development to meet the needs of different customers. After a comprehensive analysis of the evaluation results, a detailed list of product features to be improved in new product development and specific improvement directions are provided in terms of UX aspects, as shown in Table 7. In addition, the comprehensive scores of product features and evaluation items define the priority of improvement, which can solve conflicts during new product development.

## 6. Discussion

In this paper, a user experience evaluation method based on online product reviews is proposed. Our work contributes to both theory and practice. For theory, we regard large-scale online product reviews as the information source of UX evaluation and put forward a systematic UX evaluation method. The proposed method is a data-driven evaluation method. It does not simply rely on big data, but combines big data with traditional methods. In the process of method construction, we combine traditional methods with big data through two aspects of efforts. On the one hand, we use the classic questionnaire method to guide the construction of the new method's framework and the definition of key elements in reviews. On the other hand, according to this framework and the definition of key elements, we adopted the appropriate NLP technology to transform unstructured review data into structured data and integrate the data into the evaluation process. We argue that integrating big data into traditional evaluation methods can provide effective guidance for big data analysis and obtain more reliable insights. This approach can be applied to product design evaluation and further research in the field.

Table 7

List of product features to be improved

| Sequence Number | Product feature | Comprehensive score | UX aspect | Comprehensive score of UX aspect | Remark |
|---|---|---|---|---|---|
| 1 | Handle | -2.903 | Comfortability | -1.58 | |
| | | | Dependability | -1.03 | |
| | | | Hygiene | -0.070 | |
| 2 | Battery | -2.895 | Efficiency | -1.960 | |
| | | | Dependability | -0.660 | |
| | | | Comfortability | -0.157 | |
| | | | Hygiene | -0.018 | |
| 3 | Noise | -1.368 | Comfortability | -0.240 | |
| | | | Dependability | -0.090 | |
| 4 | Wheel | -0.903 | Efficiency | -0.457 | |
| | | | Dependability | -0.360 | |
| | | | Hygiene | -0.100 | Disagreement |
| … | … | … | … | … | … |

By combining traditional UX evaluation methods and big data analysis, the proposed method can benefit from the advantages of both. By taking the online product review data as the data source, the proposed method obtains a more realistic and objective voice of users than traditional UX evaluation methods. This also enables us to obtain richer information. For example, the proposed method measures the impact of different product features and UX aspects on UX based on the number of reviews, which is ignored by the traditional questionnaire method.

Compared with the existing research on product design improvement based on review mining, the proposed method has also made some progress. Existing research usually analyze only the positive or negative opinions of users about a product feature [32, 33], but since a product feature has many attributes, such a general opinion is of little value to guiding product design improvement. Benefiting from the guidance of UX theory and UX classical evaluation methods, the proposed method introduces UX aspect into the analysis of review data. The results of the UX aspect evaluation can provide a more detailed direction for improving product features.

For practice, the proposed method can inspire enterprises in the development of new products for enterprises. The product features to be improved, as well as the direction and priority of future improvements, can be obtained from the evaluation results. For new products development, the evaluation results can be used for QFD to generate design schemes. In addition, companies can easily obtain online reviews of competing products, so that they can use this method to evaluate them and gain advantages in the highly competitive market.

The proposed method can overcome the disadvantages of the questionnaire survey method, such as labor-intensive, time-consuming, and low- reliability. However, it also has some disadvantages such as processing workload and a long delay after product launch. Therefore, these two methods are complementary and can be combined to support new product development in practice.

There are some limitations in our research. First, in the integration of product features, the relationship between product features is not considered. For complex products, there are many relationships between product features, such as the superior-subordinate relationship between components, and the relationship between product components and product functions. By introducing domain ontology, it is possible to clearly describe the relationship between product features and then obtain more accurate evaluation results. Second, there is still room for improvement in the algorithm for review elements extraction. Applying machine learning algorithms such as CRF and SVM may improve the extraction performance.

## 7. Conclusion

The UX evaluation of existing products is significant for new product development. High-level big data analysis is considered to be an important condition for enterprises to realize product success [43], and our research is an attempt to adopt big data analysis in UX evaluation research. Based on online product reviews, this paper constructs a quantitative evaluation method for product UX. On the basis of defining the key elements of reviews, the product UX evaluation is completed through key element extraction, key element integration and quantitative evaluation. The utility of the method is verified by a case study of a vacuum cleaner evaluation.

The novelty of this paper is twofold. First, we combine online product reviews with UX evaluation and put forward a quantitative UX evaluation method based on online product reviews. This method takes advantage of both traditional questionnaires and large-scale online product reviews. Second, based on the number of reviews and rough number, a quantitative evaluation method was developed to determine the product features and UX aspects that need to be improved. Using rough number can effectively reduce the influence of fuzziness and uncertainty of reviews on evaluation results. This quantitative evaluation method is generally not available in the existing research on online product reviews. Compared with the existing analysis methods, such a quantitative analysis can be more effective in assisting product design.

Several further works on this topic can be explored. First, although our research is based on Chinese reviews, we believe that the proposed method is a general method that can be used for all languages. However, some steps still inevitably need to be adjusted. How to apply this method in various languages is worth exploring. Second, we only take the review texts as the information source. Other elements in online product reviews could also be included, such as ratings, tags, emoji, pictures and so on. Introducing these elements into the UX evaluation may make the evaluation results more accurate and may also bring more inspiration.

## Acknowledgement

This work was supported by the National Key Research and Development Program of China under Grant No. 2017YFB1104205; and the 111 Project under Grant No. B13044.

## Reference

[1] R. Rousi, Formidable bracelet, beautiful lantern: studying multi-sensory user experience from a semiotic perspective, in: *Proceedings of the 8th international conference on Design Science at the Intersection of Physical and Virtual Design*, Springer-Verlag, Helsinki, Finland, 2013, pp. 181–196.

[2] P. van Schaik and J. Ling, Modelling user experience with web sites: Usability, hedonic value, beauty and goodness, *Interacting with Computers* **20** (2008), 419-432.

[3] M. Hassenzahl, The thing and I: understanding the relationship between user and product, in: *Funology: from usability to enjoyment*, Kluwer Academic Publishers, 2005, pp. 31–42.

[4] F. Pucillo and G. Cascini, A framework for user experience, needs and affordances, *Design Studies* **35** (2014), 160-179.

[5] E.L.-C. Law, P. van Schaik, and V. Roto, Attitudes towards user experience (UX) measurement, *International Journal of Human-Computer Studies* **72** (2014), 526-541.

[6] M. Zarour, M. Alharbi, and E. Park, User experience framework that combines aspects, dimensions, and measurement methods, *Cogent Engineering* **4** (2017) ,1-25.

[7] W. He, Improving user experience with case-based reasoning systems using text mining and Web 2.0, *Expert Systems with Applications* **40** (2013), 500-507.

[8] J. Jin, Y. Liu, P. Ji, and H. Liu, Understanding big consumer opinion data for market-driven product design, *International Journal of Production Research* **54** (2016), 3019-3041**.**

[9] J.R. Lewis, Usability: Lessons Learned … and Yet to Be Learned, *International Journal of Human–Computer Interaction* **30** (2014), 663-684.

[10] J. Nielsen, Finding usability problems through heuristic evaluation, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, Monterey, California, USA, 1992, pp. 373–380.

[11] J. Rieman, M. Franzke, and D. Redmiles, Usability evaluation with the cognitive walkthrough, in: *Conference Companion on Human Factors in Computing Systems*, Association for Computing Machinery, Denver, Colorado, USA, 1995, pp. 387–388.

[12] M. Schrepp, A. Hinderks, and J.r. Thomaschewski, Construction of a Benchmark for the User Experience Questionnaire (UEQ), *International Journal of Interactive Multimedia and Artificial Intelligence* **4** (2017), 40-44.

[13] M. Schrepp, T. Held, and B. Laugwitz, The influence of hedonic quality on the attractiveness of user interfaces of business management software, *Interacting with Computers* **18** (2006), 1055-1069.

[14] B. Laugwitz, T. Held, and M. Schrepp, Construction and Evaluation of a User Experience Questionnaire, in: *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work*, Springer-Verlag, Graz, Austria, 2008, pp. 63–76.

[15] M. Minge, M. Thüring, I. Wagner, and C.V. Kuhr, The meCUE Questionnaire: A Modular Tool for Measuring User Experience, in: *Advances in Ergonomics Modeling, Usability & Special Populations*, 2017, pp. 115-128.

[16] S. Borsci, S. Federici, S. Bacci, M. Gnaldi, and F. Bartolucci, Assessing User Satisfaction in the Era of User Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a Function of Product Experience, *International Journal of Human-Computer Interaction* **31** (2015), 484-495.

[17] J. Chai, Y. Ge, Y. Liu, W. Li, L. Zhou, L. Yao, and X. Sun, Application of Frontal EEG Asymmetry to User Experience Research, in: *11th International Conference on Engineering Psychology and Cognitive Ergonomics - Volume 8532*, Springer-Verlag, 2014, pp. 234–243.

[18] L.E. Nacke, M.N. Grimshaw, and C.A. Lindley, More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game, *Interacting with Computers* **22** (2010), 336-343.

[19] L. Feng and W. Wei, An Empirical Study on User Experience Evaluation and Identification of Critical UX Issues, *Sustainability* **11** (2019), 2432.

[20] J. Park, S.H. Han, H.K. Kim, Y. Cho, and W. Park, Developing Elements of User Experience for Mobile Phones and Services: Survey, Interview, and Observation Approaches, *Human Factors in Ergonomics & Manufacturing* **23** (2013), 279-293.

[21] M. Hassenzahl, S. Diefenbach, and A. Göritz, Needs, affect, and interactive products – Facets of user experience, *Interacting with Computers* **22** (2010), 353-362.

[22] N. Crilly, J. Moultrie, and P.J. Clarkson, Seeing things: consumer response to the visual domain in product design, *Design Studies* **25** (2004), 547-577.

[23] J. Park, S.H. Han, J. Park, J. Park, J. Kwahk, M. Lee, and D.Y. Jeong, Development of a web-based user experience evaluation system for home appliances, *International Journal of Industrial Ergonomics* **67** (2018), 216-228.

[24] A.N. Tuch, R. Trusell, and K. Hornbæk, Analyzing users' narratives to understand experience with interactive products, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 2013.

[25] M. Winckler, C. Bach, and R. Bernhaupt, Identifying User Experience Dimensions for Mobile Incident Reporting in Urban Contexts, *IEEE Transactions on Professional Communication* **56** (2013), 97-119.

[26] S. Moller, K.-P. Engelbrecht, C. Kuhnel, I. Wechsung, and B. Weiss, A taxonomy of quality of service and Quality of Experience of multimodal human-machine interaction, in: *2009 International Workshop on Quality of Multimedia Experience*, 2009, pp. 7-12.

[27] C.-P. Lin, Learning Online Brand Personality and Satisfaction: The Moderating Effects of Gaming Engagement, *International Journal of Human-Computer Interaction* **25** (2009), 220-236.

[28] I. Lee, G.W. Choi, J. Kim, S. Kim, K. Lee, D. Kim, M. Han, S.Y. Park, and Y. An, Cultural dimensions for user experience: cross-country and cross-product analysis of users' cultural characteristics, in: *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 1*, BCS Learning &amp; Development Ltd., Liverpool, United Kingdom, 2008, pp. 3–12.

[29] R. Safi and Y. Yu, Online product review as an indicator of users' degree of innovativeness and product adoption time: a longitudinal analysis of text reviews, *European Journal of Information Systems* **26** (2017), 414-431.

[30] S.N. Ahmad and M. Laroche, Analyzing electronic word of mouth: A social commerce construct, *International Journal of Information Management* **37** (2017), 202-213.

[31] J. Qi, Z. Zhang, S. Jeon, and Y. Zhou, Mining customer requirements from online reviews: A product improvement perspective, *Information & Management* **53** (2016), 951-963.

[32] Z. Li, Z.G. Tian, J.W. Wang, and W.M. Wang, Extraction of affective responses from customer reviews: an opinion mining and machine learning approach, *International Journal of Computer Integrated Manufacturing* **33** (2019), 670-685.

[33] J. Jin, P. Ji, and C.K. Kwong, What makes consumers unsatisfied with your products: Review analysis at a fine-grained level, *Engineering Applications of Artificial Intelligence* **47** (2016), 38-48.

[34] J. Jin, P. Ji, and R. Gu, Identifying comparative customer requirements from product online reviews for competitor analysis, *Engineering Applications of Artificial Intelligence* **49** (2016), 61-73.

[35] N. Li, X. Jin, and Y. Li, Identification of key customer requirements based on online reviews, *Journal of Intelligent & Fuzzy Systems* **39** (2020), 3957-3970.

[36] B. Yang, Y. Liu, Y. Liang, and M. Tang, Exploiting user experience from online customer reviews for product design, *International Journal of Information Management* **46** (2019), 173-186.

[37] B. Pang and L. Lee, Opinion Mining and Sentiment Analysis, *Foundations & Trends in Information Retrieval* **2** (2008), 1-135.

[38] S.C.J. Lim, Y. Liu, and W.B. Lee, A methodology for building a semantically annotated multi-faceted ontology for product family modelling, *Advanced Engineering Informatics* **25** (2011), 147-161.

[39] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, and A. Hussain, Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing: A Novel Approach, *Cognitive Computation* **7** (2015), 487-499.

[40] W. Zhang, H. Xu, and W. Wan, Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis, *Expert Systems with Applications* **39** (2012), 10283-10291.

[41] M. Schindler, O. Fox, and A. Rausch, Clustering source code elements by semantic similarity using Wikipedia, in: *Proceedings of the Fourth International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering*, IEEE Press, Florence, Italy, 2015, pp. 13–18.

[42] L.-Y. Zhai, L.-P. Khoo, and Z.-W. Zhong, A rough set based QFD approach to the management of imprecise design information in product development, *Advanced Engineering Informatics* **23** (2009), 222-228.

[43] Z. Xu, G.L. Frankwick, and E. Ramirez, Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective, *Journal of Business Research* **69** (2016), 1562-1566.