Artificial Intelligence for Engineering Design, Analysis and Manufacturing

cambridge.org/aie

A personalized requirement identifying model for design improvement based on user profiling

Jing Li[1], Xinwei Zhang[1], Keqin Wang[1], Chen Zheng[2], Shurong Tong[1] and Benoit Eynard[3]

[1]Institute of Design Management, Northwestern Polytechnical University, Xi'an, China; [2]Mechanical Engineering School, Northwestern Polytechnical University, Xi'an, China and [3]Department of Mechanical Systems Engineering – UMR 7337 Roberval, Sorbonne Universités, Université de Technologie de Compiègne, Compiègne Cedex, France

Research Article

Cite this article: Li J, Zhang X, Wang K, Zheng C, Tong S, Eynard B (2020). A personalized requirement identifying model for design improvement based on user profiling. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **34**, 55–67. https://doi.org/10.1017/S0890060419000301

Received: 13 February 2018
Revised: 4 July 2019
Accepted: 16 July 2019
First published online: 27 November 2019

Key words:
Design improvement; personalized requirement; requirement identifying; user profiling

Author for correspondence:
Jing Li, E-mail: lijing2015@nwpu.edu.cn

Abstract

The personalization of products and services has become an inevitable trend in the manufacturing and service industry, but it is very difficult to identify users' personalized requirements accurately. This paper solves this problem by constructing an identifying model for personalized requirement based on user profiling. Firstly, the framework of the proposed model and the process of identifying the user's personalized requirements with this model are introduced, and then an experimental scheme for obtaining users' profiling data is designed. On this basis, an experiment is performed by investigating users' requirements for the computer to obtain the data, and the data are used for the analysis based on the proposed model. The analysis result shows that the model can reveal the difference among heterogeneous users well, find out the implicit requirements of users, and identify the gap between existing products and users' personalized requirements, which provides support to the subsequent improvement of product design.

Introduction

At present, the personalization of products and services has become an inevitable trend in the manufacturing and service industry, so it is vital for an enterprise to fully understand the requirements of the users and provide the corresponding products or service. The research data shows that 38.2% of the causes of design failure were related to the lack of accurate understanding of users' requirements, and the accurate requirements understanding contributed 21.2% to the design success (Hull *et al.*, 2011). The Harvard Business Review also notes that it is often difficult for users to find products or services that fully meet their requirements in the market (Davenport *et al.*, 2011). Therefore, it is very important to understand and build user requirements accurately and design it into the products. In the past, people usually analyzed and constructed the requirements of different user groups with methods based on market segmentation (He *et al.*, 2015; Sousa-Zomer and Miguel, 2017). But with a lack of accurate understanding of users' personalized characteristics and requirements, those methods cannot support the satisfaction of personalized requirements. Though mass customization has been widespread used to meet personalized requirements, it still has a problem with users' personalized requirements' accurate understanding and recognition. Tseng, a mass customization expert, emphasizes that understanding users' personalized requirements accurately is the key to the success of mass customization design (Tseng and Hu, 2014).

There are two reasons that lead to the main difficulties in the identifying of users' personalized requirements: (a) users are heterogeneous and the same requirement expression of different users often represents different product needs (Boukhari *et al.*, 2012); (b) in many cases, users clearly know their own needs, but the expressions are often ambiguous and inaccurate (Jiao and Chen, 2014) and contain implicit requirements. Aiming at the existing problems, this paper put forward a personalized requirement identification model based on users' profiling. It builds up users' profiling through their browsing history and on this basis obtains the understanding of users' personalized requirements. The main content of this article is as follows: Section "Related work" summarizes and analyzes the research status quo about the related problems, section "Model framework" presents a users' personalized requirement identification model framework based on users' profiling, section "Experiment design" designed the experiment process for getting the data as input for the model, section "Case analysis" takes computer buyers, for example, to carry out data collection and the analysis of the results and verified the effectiveness of the model and then discussed the value of main parameters in the model, section "Discussion" points out the contribution of the article and some further work can be done in the future.

Related work

Considering the individuation and the heterogeneity of customer requirements, the research on requirement analysis and identification is divided into three categories as follows:

Requirement acquisition method based on requirement template. In the study of mass customization, this method is widespread used to achieve the acquisition and understanding of requirements, such as the parameter-based product family classification tree template (Jiao *et al.*, 2003), requirement template based on the product BOM (Yu *et al.*, 2008), and so on. To make the customer express the requirements more intuitively, Shieh *et al.* (2008) introduced the graphic classification technology for the product requirement template construction. To furthermore reflect the personalization, Stormer provides different requirement templates to different users based on collaborative recommendation algorithm (Stormer, 2009), Miceli *et al.* put forward the dynamic acquisition system framework of users' requirements (Kreutler and Jannach, 2006; Miceli *et al.*, 2007) and proposed an idea that companies should present a personalized requirement template interface by interacting with users. At the Massachusetts Institute of Technology, professor Von Hippel discussed enabling users to select components with different parameters to form their personalized product configuration according to their personalized requirements by providing a component toolbox to users rather than analyzing customer needs directly (Von Hippel and Katz, 2002). These methods require users to express requirements from the perspective of the product domain, and a large number of parameter choices can bring confusion and even stress to users (Miceli *et al.*, 2007).

Requirement identification method based on fuzzy requirement analysis. In order to solve the problem of the first method, the second solution allows the users to express their needs directly, and then identifies the fuzzy requirements from the users more accurately with different analytical methods, such as the fuzzy-based framework, fuzzy hierarchy analysis, KJ method, fishbone diagram, and kano model (Bamford and Greatbanks, 2005; He *et al.*, 2015; Sousa-Zomer and Miguel, 2017). By analyzing the majority of users, the essence of the solution is to form an understanding of the user group and reason with that while the reasoning process is a lack of consideration about the personalized requirements of the user. To reflect the individuation of requirement better, some scholars considered some personalized information such as the demographic information of the users, the context information, and the intended use in the process of analysis of the requirement for the user (Greenyer *et al.*, 2015; Zhao and Yan, 2015), Mitsuo Nagamachi of Hiroshima University in Japan has proposed perceptual engineering theory which aims at identifying users' fuzzy perceptual needs as specific design requirements (Nagamachi, 2016). The methods of identification include joint analysis, discrete selection analysis, etc. (Jiao and Chen, 2014). Although the requirement for users to provide a lot of preference information may lead to resistance in the identification process, these studies provide good solutions to the heterogeneity problems in the process of identifying users' requirements.

Personalized requirement identification method based on user profiling. With the rapid development of Internet and big data, many studies have tried to use the Internet user data to construct user profiling with the purpose of further understanding the users and identifying users' personalized needs more accurately and comprehensively. Sara *et al.* access to the user's demographic data, interest, browsing history, etc. through interacting with them and establish a multidimensional user profiling for the subsequent process of personalized information recommendation based on the data obtained (Ouaftouh *et al.*, 2015). Vu *et al.* build the click user profile based on the user's click record, build the query user profile based on the user's query record, and push the personalized query results to the user based on the similarity between the two documents (Vu *et al.*, 2017). Except for general information query, some scholars also use profiling for the personalized recommendation of the product or service information, such as recommending different products based on the user's browsing history (Hauser *et al.*, 2014), recommending personalized music based on the user's listening history (Chung *et al.*, 2009). In the process of matching the user profiling and the query statements, methods include subject model (Trusov *et al.*, 2016), hypergraph distance matching (Daoud *et al.*, 2010), latent semantic indexing (Kesorn *et al.*, 2009), dynamic programing and the Bayesian model (Hauser *et al.*, 2014), and Markov Monte Carlo method (Chung *et al.*, 2009) are adopted. Although these studies mainly focus on the user's personalized information query and recommendation, the user query statements are similar to the needs expressed by users in the requirements engineering and both of them are requirements expressed by users for the product or information. Therefore, these studies are good references for how to get the individuate understanding of requirements expressed by users based on user profiling.

The related works and limitations are summarized in Table 1.

Model framework

Based on the idea of perceptual engineering and the LDA (theme model) method, this paper gives a model framework that uses user profiling to understand users' personalized requirements:

(1) Users' requirements include product requirements, emotional requirements, etc. Different types of requirements can also be divided into some sub-requirements. The user's requirements statement implies the different types of requirements that the user wants to express. In this paper, the LDA theme model proposed by Blei *et al.* (2003) is used to analyze the requirements stated by users.

The process of stating a user's requirement is divided into two stages: firstly, the user should select the requirements for the statement (that is, the requirements category); secondly, the users should think about what vocabulary should be used to describe the requirements. If a user has a particular requirement for a part of the product, the user will use related vocabulary to state the requirement. Let's say that the set of user requirement statements documents is represented by M, and the number of words in the documents is N.

The process of generating a requirement statement document is: (a) extract the requirement class distribution θ_m of the requirement statement document m from the Dirichlet distribution of α; (b) extract the requirement class $Z_{m,n}$ of the nth word of m from the polynomial distribution θ_m of

https://doi.org/10.1017/S0890060419000301 Published online by Cambridge University Press

Table 1. Related works and limitations

Model and method for personalized requirement identifying	Limitations
Product family classification tree template based on parameters (Jiao *et al.*, 2003)	Too many kinds of parameters, complex input, and not intuitive enough
Demand template based on product BOM (Yu *et al.*, 2008)	
Graphic classification technology (Shieh *et al.*, 2008)	Too many kinds of parameters
Personalized template based on a collaborative algorithm (Stormer, 2009)	Considering the difference among users, but the parameters are still given from the perspective of product domain
Dynamic acquisition system framework (Kreutler and Jannach, 2006; Miceli *et al.*, 2007)	
Component toolbox (Von Hippel and Katz, 2002)	Not suitable for complex product
Fuzzy frame, fuzzy level analysis, KJ method, fishbone diagram, and kano model (Bamford and Greatbanks, 2005; He *et al.*, 2015; Sousa-Zomer and Miguel, 2017).	Insufficient consideration of user's heterogeneity
Perceptual engineering method (Jiao and Chen, 2014; Nagamachi, 2016)	Understanding requirements from a user's individual perspective but requires a lot of information from users
User profiling is matched with query statements for personalized information query (Chung *et al.*, 2009; Hauser *et al.*, 2009; Greenyer *et al.*, 2015; Ouaftouh *et al.*, 2015; Zhao and Yan, 2015; Vu *et al.*, 2017)	In the information query, the user's heterogeneity is considered and provides a reference for the understanding of the user's personalized demand

the requirement class; (c) extract the $Z_{m,n}$'s lexical distribution $\varphi_{z_{m,n}}$ from the Dirichlet distribution of β; (d) extract the final generated words $w_{m,n}$ from the polynomial distribution $\varphi_{z_{m,n}}$ of the vocabulary of requirement statement. According to the principle above, the generation process of requirements statement document set M can be represented by the following formula (Blei *et al.*, 2003):

$$p(M|\alpha,\beta)=\prod_{m=1}^{M}\int p(\theta_m|\alpha)\left(\prod_{n=1}^{N_m}\sum_{z_{mn}}p(z_{mn}|\theta_m)p(w_{mn}|z_{mn},\beta)d\theta_m\right)$$

where θ_m, $\varphi_{z_{m,n}}$ are unknown vectors. The distribution vector is estimated by Gibbs Sampling method, and the sampling formula is shown below (Guo *et al.*, 2016):

$$p(z_i=k|z_{-i},w)\propto\frac{n_{-i,m}^{k}+\alpha_k}{\sum_{k=1}^{K}(n_{-i,m}^{k}+\alpha_k)}\cdot\frac{n_{-i,k}^{w}+\beta_w}{\sum_{w=1}^{V}(n_{-i,k}^{w}+\beta_w)}$$

In the formula, $z_i = k$ means that the requirement class of the ith word is k, the form of i is (m, n) and it means that i is the nth term of the mth document; $-i$ means that the term with subscript i should be removed; $n_{-i,m}^{k}$ represents the number of words belonging to the requirement category k in document m; $n_{-i,k}^{w}$ represents the number of times that word w_i is in the requirement category k.

By sampling and training, the value of $p(\theta_m|\alpha)$ and $p(\varphi_{z_{m,n}}|\theta_m)$ that can maximize $p(M|\alpha, \beta)$ will be determined, that is, the distribution matrix of the requirement statement document–requirement class (theta) and the requirement class–requirement statement (phi) and then count the top N high-frequency words in each requirement class (twords).

(2) According to the results of theta, select requirement classes whose probability is in top T (such as top 3) from each user's requirement statement document as the user's complete requirement elements.

(3) Collate each user's browsing history of recent D days to form the user profiling. User's browsing history has unstructured features. Re-record each user's browsing history in a form, each column of the table represents a browsing record of the user, and the rows of the table represent the main parameters of the product to be viewed. For example, the main parameters of the computer include brand, price, color, standby time, naked machine weight, main selling points, and so on.

Correspond the vocabulary of each user's complete requirements to rows in step 3 and take the number that is repeated most often as the value of that term. It is important to note that there are two classes of vocabulary for users to describe requirements. One class includes words that describe specific product features, such as price, color, size, etc., and they can take corresponding values directly from user profiling. The other includes emotional imagery words that do not specify the characteristics of a particular product, such as "advanced", "grade", "beautiful", etc., these kinds of terms are processed by the quantitative measurement methods of perceptual engineering (Nagamachi and Imada, 1995; Guo *et al.*, 2015): first extract the emotional intention terms that represent the emotional needs of the user from the user requirement statement and form the vocabulary of emotional intention; then list the characteristics of the product and the components of the products according to the characteristics of the product; finally, determine the relationship between the emotional intention terms and the characteristics of the product and parts with back propagation neural network (Guo *et al.*, 2015) and the statistical analysis method based on multiple regression analysis (Su *et al.*, 2004), etc. In this paper, the statistical analysis method of Su *et al.* (2004) was adopted, the corresponding relationship between the emotional intention vocabulary and the product characteristics was obtained by conducting questionnaires and carrying out regression analysis on the results. Then, replace the emotional intention vocabulary stated by users with the corresponding product features and get the corresponding value of the product features from user profiling.

(5) Go through all the words obtained in step 2, and assign the values as in step 4, then end with all the words assigned.

https://doi.org/10.1017/S0890060419000301 Published online by Cambridge University Press

Table 2. Main parameters of PRUP model

Procedure	Parameter
Step 1	Requirement class K, Number of high-frequency words in each requirement class N The superparameter α and β of the requirement class distribution
Step 2	The probability rank of the requirement class is T
Step 3	The time span D of browsing history
Step 4	The correspondence table of emotional intention terms and product characteristics

The main parameters in the model are shown in Table 2, and the determination of each parameter needs to be confirmed according to the specific application field. The process of using this model to identify users' personalized requirements is shown in Figure 1. The model classifies the requirements through LDA. On the semantic level, similar words are classified into one category. Each requirement category contains the requirements that the user may concern but did not mention; therefore, the imperfection of user requirement document can be solved well by understanding user requirements according to the words in each category. Through the processing of perceptual engineering, the ambiguity of user requirement statement can be solved well and at the same time, the problem of heterogeneity in the expression of user requirement can be solved well by using the user's browsing history to assign values to the requirement statement.

Experiment design

To understand the user's personalized requirements through user profiling, the user's browsing data need to be obtained first for building user profiling. Trusov *et al.* (2016) obtained website browsing information of more than 45,000 households from a leading global information and measurement company, but in view of the privacy of users and the business secrets of the enterprise, accurately and comprehensively.

It is often difficult to obtain the user's browsing record obtaining relevant data through the experiment is more often adopted, such as Daoud *et al.* (2010) picked 10 users as experimental subject and extracted the 10 users' search log within 3 months and then provided users with personalized retrieval results based on the search log. Hauser *et al.* (2009) invited some people to visit the experimental shopping site designed by himself and recorded the visitors' purchase probability in real time to deduce the relationship between the page style of the website and the purchase probability of different users. The relationship between users' browsing behavior and requirements can be well reflected by simulating users' web browsing, shopping web browsing, and other behaviors based on certain requirements through good experimental design. And a reference for the in-depth understanding of users' personalized requirements can be obtained from the relationship. Therefore, this paper also obtains the user's browsing and requirement data by experiment. The process is as follows:

(1) Select the person who has the purchase habit of browsing various related products and inquiring products online and then shopping as the experimental subject.
(2) Suppose the experiment subject now needs to buy a product based on life or job needs. Ask the experimenters to determine the approximate requirements first and then learn about related product information in shopping sites and search engines according to shopping habits. This step does not need to be done quickly because we need the experimenter to try to simulate the normal shopping state like browsing the shopping site in free time and checking out the products recommended by others after discussing with them.
(3) After the second step, the subjects are more specific about their requirements. Now, we ask the experimenters to describe their own requirements.
(4) Record the requirement description of each subject and compile the relevant browsing history for each person then form the user profiling which provides a basis for the use of user profiling to understand personalized requirements. There are many kinds of browsing records that can be used to understand users' personalized requirements, including product browsing history, web browsing history, and users' comment history. However, within a small time span, a single user has a very limited comment history which can be ignored, only the user's product browsing record and web browsing history need to be collected.

Case analysis

Experiment process

Determine the parameters required for each step according to the design in section "Model framework":

(1) A total of 24 people, including students, white-collar workers, and business people, were selected as the subjects.
(2) To set the computer to be the purchase object. At the same time, according to the feedback time span from subjects, we set the time needed by this step to be 7 days.
(3) When collecting users' requirement descriptions, we give some hints to get the experimenters to describe the requirements as comprehensively as possible such as:
    (a) What is the most common use of your computer? (for ordinary games, movie entertainment, or learning materials, running software, etc.? Or there are a variety of uses, mainly for what, and then for what, etc.)
    (b) What are your requirements for computers? (a few general words can be used, such as looking advanced, running smoothly, looking full and round or looking steady, preferring bright color, etc. In addition to these qualitative descriptions, more specific requests can be proposed such as cannot be XX color, the request of the memory size, the price should be in a certain interval, preferring XX brand, and so on.)

    These are just simple tips and the experimenters can describe the requirements according to their real needs.
(4) To record users' requirement description and their browsing history.

Data preprocess

The requirement description documents for 24 users need to be processed by the "jieba" tool to excise the meaningless adverbs

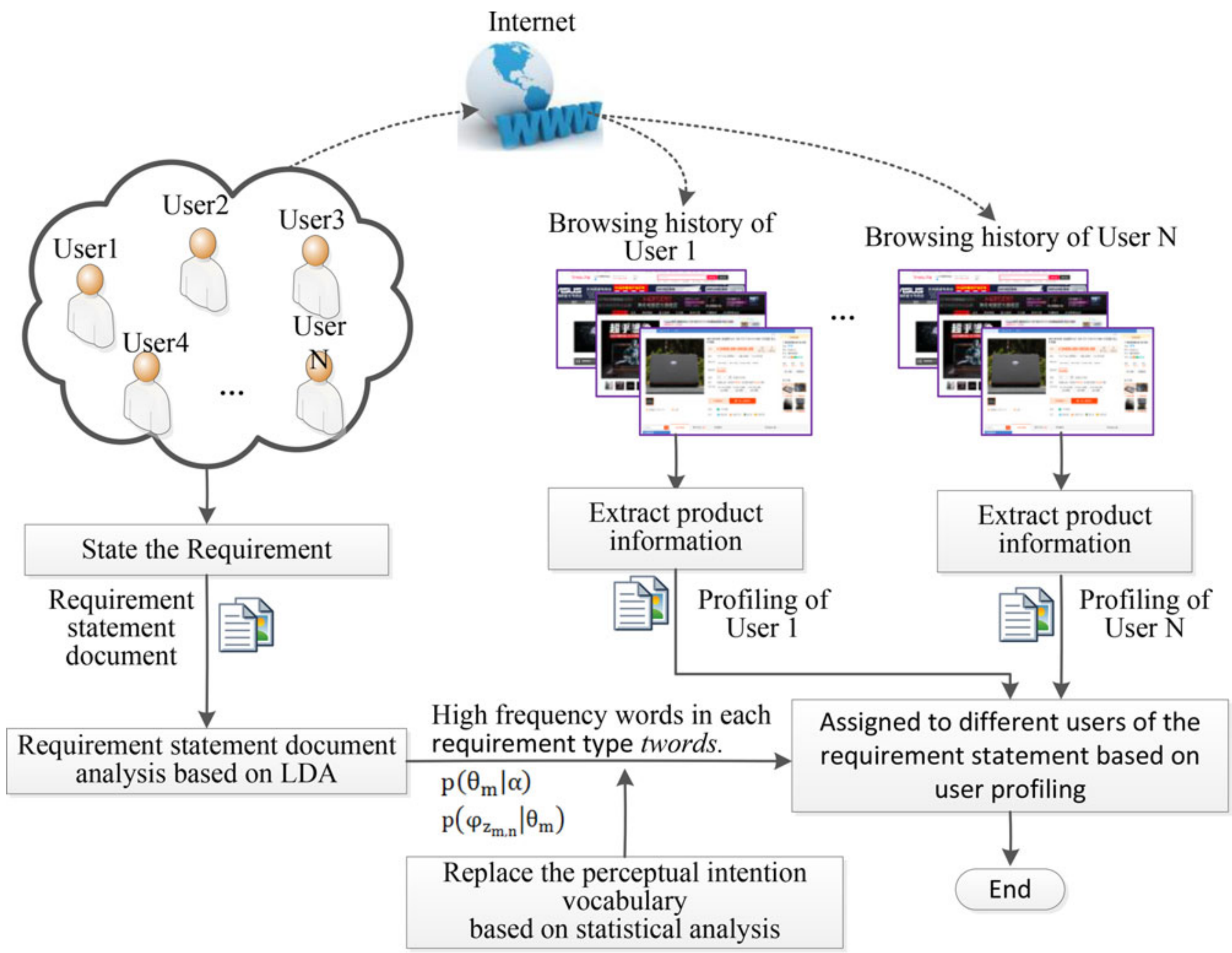

Fig. 1. The process of personalized requirement identifying.

and carry out word segmentation and finally get a corpus including 2773 terms instead of being used for LDA analysis directly.

Users' browsing data should be processed through the following steps before being put into use:

Firstly, web browsing history is unstructured and contains a variety of contents so it needs to be simplified: re-record the browsing history according to the title and keywords of the web page. For example, if the title of the web page content is about "How to select graphics card" or " How about the XX laptop" or "the difference between i5 and i7" or " brand X and brand Y which is better" etc., then record the reference objects represented by the title, namely, the four browsing records should be recorded separately as "graphics card, brand XX, CPU, brand X, brand Y"; If the title is only about tips of selecting a computer and the purchase strategy, etc. rather than containing specific objects, then this record should be ignored.

Secondly, the browsing history of shopping websites should also be processed. In order to get a more significant analysis result, the continuous data need to be changed into discrete data when recording the user's browsing history and the main processing method is: for the price, note the range 1–499 as the price P1, 500–999 as the price P2, 1000–1499 as the price P3, and so on; for the thickness, note the range 0–9.9 mm as H1, 10–14.9 mm as H2, 15–19.9 mm as H3, 20–24.9 mm as H4, 25 mm and above as H5; for the weight of the naked machine, note the range 0–1 kg as W1, 1–1.49 kg as W2, 1.5–1.99 kg as W3, 2–2.49 kg as W4, 2.5–2.99 kg as W5, 3 kg and above as W6. See Appendix 1 for specific record examples.

Data analysis

Acquisition of the result

According to step 1 of the model framework, the LDA model is extracted from the requirement statement documents after the word segmentation, and the algorithm of LDA is implemented with *c* #. Most of the existing references (Guo *et al.*, 2016) set the parameters in the LDA model as: $\alpha = (50/K)$, $\beta = 0.01$. K represents the number of implied requirement classes, and the study revealed that setting K to 7 can maintain the balance of model accuracy and leanness well (Trusov *et al.*, 2016). The parameter setting in this paper follows the above researches. After the parameter setting is completed, the model begins to run. However, in our experiment, the difference among the seven implicit requirement types obtained when $K = 7$ was not significant and the different requirement types were not differentiated. After assigning to K, the value ranged 5–8, we found that the result is significant when $K = 6$. The results of high-frequency

Table 3. High-frequency words in the requirement class

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Top1	Process	Price	Game	Requirement	Computer	Operation
Top2	Smooth	Configuration	Lenovo	Internal memory	Quest	Software
Top3	Aspect	Weight	Note	Choose	Do	Unable
Top4	Hard disk	Speed	Brand	Endurance	Looking	Use
Top5	Graphic card	Screen	Office	Entertainment	Use for	Time
Top6	Processor	Purchase	Consider	Laptop	Color	Work
Top7	Color	Size	Apple	Preference	Basic	Look
Top8	Can	Product	SSD	Hope	Good looking	Video
Top9	Demand	Be used	Learn	Battery	Sense	Like
Top10	Is	Price level	Laptop	As	With	Movie
The implicit requirement class	The performance of CPU	Cost performance	Brand and application scenarios	Endurance	Looking	Running software and playing video

Table 4. The probability of each requirement class in each user's requirement document

	The probability of class 1	The probability of class 2	The probability of class 3	The probability of class 4	The probability of class 5	The probability of class 6
User 1	0.1147	0.0996	0.0919	0.1602	0.0996	0.2284
User 2	0.2725	0.1082	0.0796	0.1724	0.1367	0.1296
User 3	0.0994	0.1877	0.2187	0.1371	0.0700	0.1485

word (twords) in each requirement class are shown in Table 3, the probability results (theta) of each requirement class in each user's requirement document are shown in Table 4. For the sake of simplicity, the data from the 24 participants were no longer listed, only three users' data were listed.

According to Tables 3 and 4, user 1 pays more attention to the performance of CPU, endurance, running software, and playing video, user 2 prefers better CPU, longer running time, and better appearance, user 3 puts emphasis on price performance, brand, running software, and playing video. According to step 5 in the model framework, the detailed requirements based on user profiling of the users are obtained, as shown in Appendix 2 (for the sake of simplicity, only the detailed requirements of the first user are listed). According to the perceptual engineering experiment mentioned in section "related work", five perceptual image terms were extracted based on users' requirement statements which are beautiful (or pretty, good looking), fancy (or classy, top grade), portable (carry), price performance, respectively. Then interview the subjects about the five words and ask them to state the specific product characteristics represented by each perceptual image term, the product characteristics are shown in Appendix 1. For the sake of simplification, in this case, simple statistic methods are applied to the survey results, and the corresponding product characteristics of each image term are shown in Table 5. The subjects in this interview were the same people as the subjects in the experiment. When the model is applied to other product areas, the relationship between perceptual image terms and product characteristics can be determined through a large amount of research and regression analysis first and then it can be used in the subsequent analysis as known conditions.

Table 5. Corresponding relation table of perceptual image terms

Beautiful	Color, shape/outline (specific product model can be used for presentation)
Fancy	Color, shape/outline (same as above), screen resolution, screen type
Portable	Size (including thickness, screen size), weight, standby time
Texture	Color, shape/outline (same as above), material, color
Price performance	Price, CPU, graphics card, internal memory, hard disk, standby time

Result analysis

The requirement statements of the 3 users are stated in Appendix 3. Take Appendices 1–3 together, we can see that:

(1) After analysis, more detailed and complete user requirements are obtained. For example, after analysis, we found that "running smoothly" mentioned by user 1 means a specific level of

https://doi.org/10.1017/S0890060419000301 Published online by Cambridge University Press

the parameters of hard disk, graphics card, and CPU; "the standby time should be long enough" implies that user 1 prefers the standby time to be longer than 9 h; "the color needs to be a little darker" means that user 1 prefers silver or gray. Seventy-nine percent of the subjects (19 people) agreed that the results were consistent with the real needs they did not express after reading the analysis results.

It is important to note that, without LDA analysis, the user's preferences for different parameters can also be obtained directly from the browsing records. However, it is not possible to feed all parameters back to the users for verification when there are so many product parameters. One of the goals of LDA analysis is to filter out the user's implicit concerns and feed the parameters valued by users back to them.

(2) The first section points out that due to the heterogeneity of the users, the specific meanings of different users' same requirement statements are different. For example, in Appendix 3, user 1 and user 2 have both proposed "running smoothly", fast speed but the results of the analysis about CPU are different; user 1, 2, and 3 all mentioned that the standby time needs to be long enough but the analysis result shows that the specific standby time wanted by the three people is slightly different. It is shown that this model can reflect the individuation of heterogeneous user requirements well.

(3) As can be seen, the analysis results are not exactly consistent with the user requirement statements, where there will be conflicts. For example, user 2 stated that it must be a solid-state hard disk, but the analysis results show that the user needs a 500 GB normal hard disk or a 256 GB solid-state hard disk. Besides the model cannot fully understand users' requirements, the most important reason for the conflictions is that products in the market cannot fully meet users' requirements at present and this causes the users to browse some products that do not conform to their requirements. For example, after inquiring the users, we found that user 2 wants a computer with SSD but the products with SSD and other conforming configurations may equip with a plastic shell.

This is the difference between this model and the personalized product recommendation based on browsing history. Personalized product recommendation based on browsing history simply recommends existing products that are similar to the user's requirements. Although the personalized product recommendation is able to recommend different products according to different users' browsing history, it is possible that there is no product in the market that fully meets users' requirements just as Harvard business review notes that in a market, it is often difficult for users to find products or services that fully meet their needs (Davenport *et al.*, 2011). The method mentioned in this paper can find the gap between existing products and users' personalized requirements and provide the foundation for providing completely personalized products in the future.

The value analysis of the main parameters

The main parameter settings in the model are further analyzed and the results are as follows:

(1) During the first step of LDA analysis, 10 words were selected from each implied requirement type. Then select the top 5, top 15 and top 20 words, respectively, and carry out the analysis according to Section "Data analysis". After that, we can get the percentage of users satisfied with the analysis results of different parameter settings and the results are shown in Figure 2. The results show that the analysis results can reflect the user requirements better with a greater value of N. As a result, more people are satisfied with the analysis results. But the number of users who are satisfied with the results has been growing more and more slowly after the size of N reaching a certain degree. This is because as the ranking of the high-frequency words decreased, the words will appear in the users' statements with the lower frequency and their correlation with the user's real requirements will also decrease. At the same time, the increase in N results in the increase in parameters in the analysis results, which increases the user's cognitive load. So the value of N best be taken between 10 and 20.

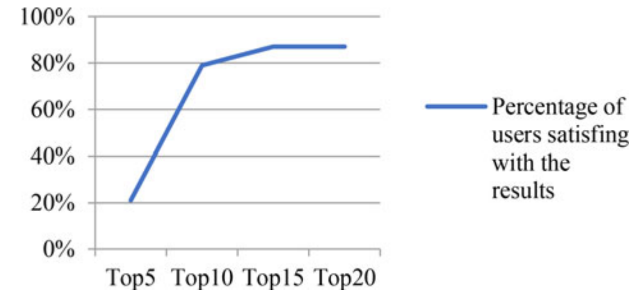

Fig. 2. The percentage of users who are satisfied with the analysis results with different numbers of words.

In addition to the percentage change of the whole, this article also analyzed the individuals, mainly about the change of the users' satisfaction with the analysis results with the change of N's value. Figure 3 shows the three representative change curves, when N takes 5, some users think that the analysis results did not understand their requirements well so the satisfaction degree of the analysis result is 0; with the increase in N, some users are more satisfied with the analysis results while some users think that the analysis results did not change significantly so the satisfaction increase is also slower. In general, with the increase in N, users' satisfaction with the analysis results will increase and some users' attitudes changed from dissatisfaction to satisfaction, which is why the curve in Figure 1 rises.

(2) The top three classes of requirements are selected for understanding user requirements. Each user's satisfaction with the analysis results varied when the top 5 or top 7 requirements classes are chosen respectively. The change of some users' satisfaction is shown in Figure 4. It can be seen that as the number of selected requirement classes increases, users' satisfaction is likely to rise but the upward trend may be not obvious, and it is also likely to decline. The reason is that the requirements classes with low ranking may be subordinate to this user with a very small probability, so the accuracy of user requirements identified based on this class will significantly reduce. Therefore, only the requirements with the value of K ranks in the first 50% need to be selected.

(3) The browsing records within 7 days during the test period are selected to construct user profiling in step 3. Some users' satisfaction with the analysis results is shown in Figure 5 when the browsing records of the last 2 days, the last 4

https://doi.org/10.1017/S0890060419000301 Published online by Cambridge University Press

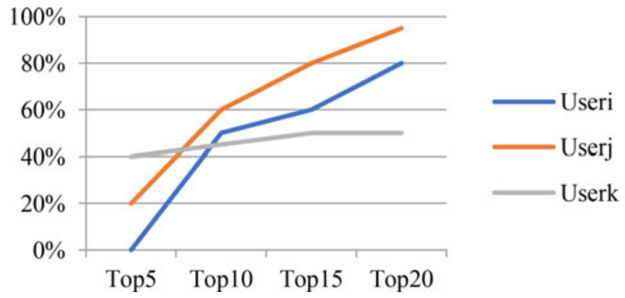

Fig. 3. The change of satisfaction with the analysis results of the individual users with the change of *N*.

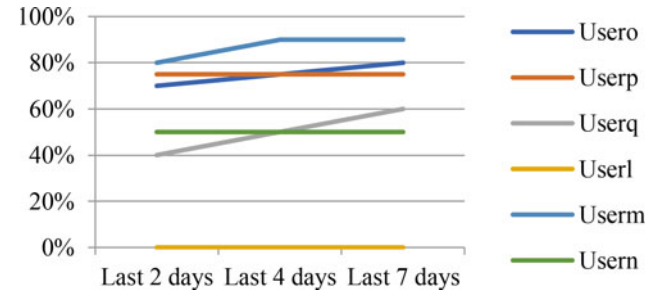

Fig. 5. The changes in the individual users' satisfaction with the analysis results with the change of the time span of the browsing records.

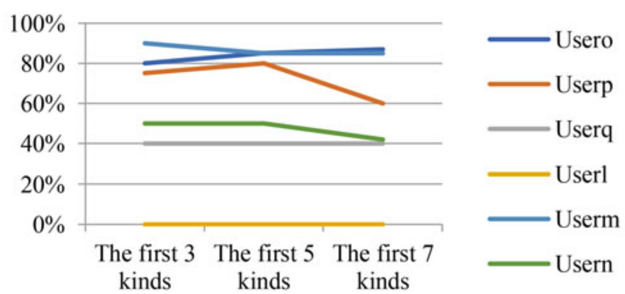

Fig. 4. The changes in the individual users' satisfaction with the analysis results with the change of the number of requirements classes selected.

days, and the last 7 days are selected. It can be seen that the longer the time taken, the higher the user's satisfaction with the analysis results. But as the timeline gets longer, the rate of change in the satisfaction is getting smaller and smaller. The reason is that the more users' browsing data were gathered, the more accurate the user's personalized requirements will be and the more satisfied the user will be with the analysis results. So the curve will go up steadily. At the same time, the user's requirements may be not specific at the beginning and they will become clear gradually with the continuous online browsing. The newer the browsing history is, the better it can reflect the user's personalized requirements. So the time span of the browsing history is not a thing that bigger is better. The appropriate time span can be selected according to different products and the corresponding browsing habits of users. Or the time can be used as a parameter in user profiling and different weights can be set according to the time. The closer the time is, the greater the weight will be.

Discussion

How to understand users' requirements better has been a major research issue in requirement engineering. As users paying more and more attention to personalization, more and more researches have begun to focus on understanding users from an individual perspective. Most approaches focus on achieving personalization from the process of obtaining requirements such as providing the personalized interface or get an in-depth understanding of the content of the requirements through interacting with users from time to time. These methods provide good solutions for identifying the requirements of heterogeneous users. However, there are still some problems, mainly about increasing the burden of users, including cognitive burden and time burden. Under the background of the rapid development of the Internet, users are producing more and more data which includes the users' personalized information and the data makes it possible to understand users' personalized requirements better without increasing the users' burden.

The identifying model for personalized requirement based on user profiling (PRUP) makes use of the Internet data generated by users to understand user's requirements. Through the analysis of the experimental data, it can be known that the method can interpret the incomplete and fuzzy requirements of the user's statement as specific parameters that meet the user's real thoughts and interpret the same statement vocabulary from different users as different product characteristic parameters according to the user's heterogeneity. And the gap between existing products and user requirements in the market also can be found, which makes up for the lack of product recommendation methods based on browsing history. At the same time, it can be known from the parameter analysis results that the selection of some important parameters in the proposed model has a significant impact on the analysis results. Therefore, in the process of using the method to identify the personalized requirements, the value of the parameters needs to follow certain principles. For example, for the number of words *N* of each requirement class, the requirement class *K*, and the time span selected to extract user browsing records, the value of these parameters is not as large as possible. The balance between user satisfaction and model complexity, the characteristics of different products and user browsing habits should be considered when assigning parameters.

The main characteristic of the PRUP method is the use of user-generated data to understand the requirements from an individual perspective. Industry 4.0 proposes to use the highly flexible product add-on service production mode to meet the personalized requirements of consumers. It can be seen that PRUP provides a reference for discovering the personalized requirements of consumers and better achieving the goals of Industry 4.0. At the same time, the PRUP method can be extended to many application areas when the personalization of products and services has become a trend in the manufacturing and service industries. For example, this way of identifying personalized requirements is a step forward than the personalization of segmentation groups in mass customization, providing better support for increasing the satisfaction of mass customization to the requirements; the data resources generated by users are effectively utilized to provide support for enterprises to improve the design in the PRUP method. Essentially, it introduces external resources to support rapid innovation activities of products, providing an effective way for enterprises to implement open innovation; design for sustainability (DFS) begins with a focus on technical aspects of sustainability, such as green design, eco-design, etc., and now has recognized the crucial importance of the role of users, thus the emotionally durable design, design for sustainable behavior, community innovation design are getting attention (Vezzoli *et al.*, 2018). The PRUP method focuses on community data and emotional requirements that users cannot accurately express, which provides a method and technology support for the transition from technology to users of DFS, etc.

Certainly, understanding the user's personalized requirements requires not only obtaining the browsing history of the user but also grasping more dimensional information of the user. Based on the proposed PRUP framework, personal information, such as professional background, knowledge level, work tasks, historical behavior information, such as related product usage history, product usage logs in computer, usage preferences, and external environmental information, such as new technologies, competing product information, laws and regulations, social culture, etc. could be added into the framework in the future. Personal information can help engineers better identify user heterogeneity. For example, differences in the professional background may make the words expressed by users have different meanings. Historical behavior information of the user and external environmental information can help engineers find an implicit requirement that users are difficult to express clearly. For example, through the product usage log, it may reveal how the current product cannot satisfy the user's requirements, thus forecasting future product functions; the relevant new technologies on the market may represent the user's implicit expectations of the product, and so on. Adding these information to the PRUP will help engineers to understand the user more deeply and identify the personalized requirements of the user more accurately.

Conclusion

This paper presents an identifying model of users' personalized requirements. The model aims to identify the personalized components of the user's requirements by collecting the user's browsing history so as to realize a profound understanding of the user's requirements. In order to verify the validity of the proposed model, an experiment was designed to obtain the data for running the model and then this paper presents the process of processing and analyzing the input data and obtaining the understanding of user requirements based on the analysis results. Compared with the previous methods, the main contributions of this paper are as follows:

First, analyzing the requirements stated by users with the LDA model can extract the implicit requirements of users without any increase in the user's burden so as to get a more comprehensive understanding of users' requirements.

Second, the model's understanding of user requirements is based on the user's browsing history, with the same user requirement statements, the analysis results of different browsing history are also different which can better reflect the requirement difference among heterogeneous users.

Third, the model can compare the objective browsing history with the subjective requirements presented by users to identify the gap between existing products and users' personalized requirements and provides some suggestions for the improvement of product design.

In addition, the paper also analyzes the main parameters of the model and discusses the scope and feasibility of the model parameter values.

However, the model that is being built is still in a preliminary stage and there is much work to be done to push the model into the usage stage. The work needed mainly includes developing a system based on PRUP model which allows users to enter their own requirements on the system interface and can capture the user's browsing history to build user profiling based on it automatically; In addition, in view of the complexity of Chinese, the result includes many meaningless words after LDA analysis, such as "aspect", "look", etc. which means that there is room for further improvement in the segmentation algorithm adopted by the model.

Financial support. This work was supported by the National Natural Science Foundation of China (71572147), Humanity and Social Science Foundation of Ministry of Education of China (17YJC630059), the Social Science Foundation of Shaanxi Province of China (2017S21), the Fundamental Research Funds for the Central Universities of China (3102018jcc014).

References

Bamford DR and Greatbanks RW (2005) The use of quality management tools and techniques: a study of application in everyday situations. *International Journal of Quality & Reliability Management* **22**, 376–392.

Blei D, Ng A and Jordan M (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research* **3**, 993–1022

Boukhari I, Bellatreche L and Jean S (2012) An ontological pivot model to interoperate heterogeneous user requirements. *Lecture Notes in Computer Science* **7610**, 344–358.

Chung TS, Rust RT and Wedel M (2009) My mobile music: an adaptive personalization system for digital audio players. *Marketing Science* **28**, 52–68.

Daoud M, Tamine L and Boughanem M (2010) A personalized graph-based document ranking model using a semantic user profile. *18th International Conference on User Modeling, Adaptation, and Personalization*, Big Island, HI, USA, June 20th–24th, 2010.

Davenport T, Mule LD and Lucker J (2011) Know what your customers want before they do. *Harvard Business Review* **89**, 84–92.

Greenyer J, Haase M, Marhenke J and Bellmer R (2015) Evaluating a formal scenario-based method for the requirements analysis in automotive software engineering. *ACM 10th Joint Meeting on Foundations of Software Engineering*, Bergamo, Italy, August 30th–September 4th, 2015.

Guo F, Liu W, Cao Y, Liu F and Li M (2015) Optimization design of a web-page based on Kansei Engineering. *Human Factors and Ergonomics in Manufacturing & Service Industries* **26**, 110–126.

Guo L, Li Y, Mu D, Yang T and Li Z (2016) A LDA model based topic detection method. *Journal of Northwestern Polytechnic University* **34**, 697–701.

Hauser JR, Urban GL, Liberali G and Braun M (2009) Website morphing. *Marketing Science* **28**, 202–223.

Hauser JR, Liberali G(G) and Urban GL (2014) Website morphing 2.0: switching costs, partial exposure, random exit, and when to morph. *Management Science* **60**, 1594–1616.

He L, Ming X, Li M, Zheng M and Zhitao X (2015) Understanding customer requirements through quantitative analysis of an improved fuzzy Kano's model. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* **231**, 699–711.

Hull E, Jackson K and Dick J (2011) *Requirements Engineering*. London: Springer.

Jiao J(R) and Chen C-H (2014) Customer requirement management in product development: a review of research issues. *Concurrent Engineering* **14**, 173–185.

Jiao J, Tseng MM and Du X (2003) Identifying need patterns for customization and personalization. *Integrating Manufacturing Systems* **14**, 387–396.

Kesorn K, Liang Z and Poslad S (2009) Use of granularity and coverage in a user profile model to personalise visual content retrieval. *IEEE 2009 Second International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services (CENTRIC)*, Porto, Portugal, September 20th–25th, 2009.
Kreutler G and Jannach D (2006) Personalized needs elicitation in web-based configuration systems. In Thorsten B and Gerhard F (eds), *Mass Customization: Challenges and Solutions*. Berlin, Germany: Springer Press, pp. 27–42.
Miceli G, Ricotta F and Costabile M (2007) Customizing customization: a conceptual framework for interactive personalization. *Journal of Interactive Marketing* **21**, 6–25.
Nagamachi M (2016) *Kansei/Affective Engineering*, 1st edn. Boca Raton, FL: CRC Press.
Nagamachi M and Imada AS (1995) Kansei Engineering: an ergonomic technology for product development. *International Journal of Industrial Ergonomics* **15**, 1–1.
Ouaftouh S, Zellou A and Idri A (2015) User profile model: a user dimension based classification. *IEEE 10th International Conference on Intelligent Systems: Theories and Applications (SITA)*, Rabat, Morocco, October 20th–21st, 2015.
Shieh MD, Yan W and Chen CH (2008) Soliciting customer requirements for product redesign based on picture sorts and ART2 neural network. *Expert Systems with Applications* **34**, 194–204.
Sousa-Zomer TT and Miguel PAC (2017) A QFD-based approach to support sustainable product-service systems conceptual design. *The International Journal of Advanced Manufacturing Technology* **88**, 701–717.
Stormer H (2009) Improving product configurators by means of a collaborative recommender system. *International Journal of Mass Customisation* **3**, 165–178.
Su J, Jiang P, Zhu B and Li H (2004) Research on Kansei Engineering and its application to product design. *Journal of Xi'an Jiaotong University* **38**, 60–63.
Trusov M, Ma L and Jamal Z (2016) Crumbs of the cookie: user profiling in customer-base analysis and behavioral targeting. *Marketing Science* **35**, 405–426.
Tseng MM and Hu JS (2014) *Mass customization*. Berlin/Heidelberg: Springer.
Vezzoli C, Ceschin F, Osanjo L, Mugendi MK and Diehl JC (2018) *Design for Sustainability: An Introduction*. London: Springer.
Von Hippel E and Katz R (2002) Shifting innovation to users via toolkits. *Management Science* **48**, 821–833.
Vu T, Willis A, Kruschwitz U and Song D (2017) Personalised query suggestion for intranet search with temporal user profiling. *ACM 17th Proceedings of the 2017 Conference on Human Information Interaction and Retrieval*, Oslo, Norway, March 7th–11th, 2017.
Yu W, Wang J, Wei X and Li R (2008) A new product configuration approach based on BOM (bill of material). *Mechanical Science and Technology for Aerospace Engineering* **27**, 446–450.
Zhao J and Yan X (2015) A service process model for outsourcing human resources management system: based on massive customized requirements research perspectives. *The Open Cybernetics and Systemics Journal* **9**, 1496–1501.

Appendix 1

The format of the browsing history

Main points of the record	Browsing history 1	Browsing history 2	……
Website	JD		
Brand	Lenovo	……	……
Model	小新潮7000		
Price	P10		
Color	Silver, red, etc.		
System	Windows 10		
CPU	Dual core, Intel I7-7500U		
Internal memory	8 GB, DDR4 2400		
Maximum size of memory supported	8 GB		
Size of hard disk	1 TB + SSD 128 GB (mixed hard disk)		
Graphics cards	Discrete graphics, 2 GB		
CD-ROM	Without		
Size of screen	14 inches		
Screen ratio	16:9		
Screen resolution	FULL HD (1920 × 1080)		
Type of screen	LCD		
Bluetooth	Bluetooth4.1		
WLAN	WLAN		
IR	Without		
Usb2.0	USB2.0 ×1		
HDMI	HDMI ×1		
USB3.0	USB3.0 ×1		
Type-C	Type-C ×1		
Webcam	HD, 720 pixels		
Card reader	SDSMMMCMS		
Battery	3 Core lithium ion battery		
Standby time	Longer than 9 h		
Thickness	H3		
Naked machine weight	W3		
Producing area	Main land of China		
Main features	Light, narrow bezel, backlit keyboard, dual hard disks, charge 80% in an hour, aviation 5 series aluminum alloy, glass bead sandblasting texture technology, wide-angle flip of 178°, authorized edition of Office		

https://doi.org/10.1017/S0890060419000301 Published online by Cambridge University Press

Appendix 2

Personalized requirements of user 1 based on user profiling

User 1					
Requirement 1	Parameter	Requirement 2	Parameter	Requirement 3	Parameter
Process	-	Need	-	Run	-
Smoothly	-	Internal memory	8 GB	Software	
Aspect	-	Choose	-	Can not	-
Hard disk	SSD/mixed hard disk	Endurance	9 h and above	Use	-
Stuck	Discrete graphics	Entertainment	-	Time	9 h and above
CPU	Dual core, I7	Laptop	-	Work	-
Color	Silver or gray	Preference	-	Look	-
Can	-	Hope	-	Video	-
Requirement	-	Battery	Lithium ion battery with 3 or 4 cores	Like	-
Is	-	As	-	Movie	-

Appendix 3

User 1: The main usage of my computer is to search the information, run the software, and play music and video for entertainment. It will also be used as a video communication tool to chat with others. My main requirements of the computer are running software and entertainment. My request for the computer is: firstly, it should run smoothly and have 8 GB internal memory; secondly, the computer should be light and thin and easy to carry because as a student I will use it frequently and carry it to other places. At the same time, the standby time should be as long as possible and the performance of the battery needs to be good. Then the looking of it needs to be nice and the metallic shell is better. Finally, the color should be dark and looking nice. I do not want jumping color because it looks lack of science and technology feeling. I prefer a product whose price is between 5 k and 6 k, and the Lenovo and Dell are brands that I prefer because I used products of the two brands and they were all good.

User 2: My last computer had been used for 3 years, although it can run smoothly now it cannot meet the requirements of my work. As a professional woman, my job is to deal with a large number of documents and use some drawing software according to the requirements of my work. And I will also carry my computer on my business trip so the computer is very important to me. Therefore, I need a more suitable computer. This time I want a computer that is not too big, looks stylish, the body is light and thin, relatively safe, and the most important thing is the standby time must be long enough so that it is not only convenient to carry, but also it will not cause troubles by the lack of electricity during the business trip. And its performance should be proper, the graphics card and internal memory should be good enough and the reaction speed should be high enough. What's more, a proper CPU is essential and the hard disk needs to be an SSD rather than a normal one. In this way, the graphics processing power will be improved, the operation will be faster, and it can improve the efficiency and save time. Finally, the interface of the computer should be rich and easy to use, which can satisfy the diversified transmission requirements in the work.

User 3: I want a laptop in white or rose gold and in a full and round shape. I hope that it can run smoothly without stuck. The storage of the laptop should be large enough and the standby time should be long enough and it needs to be light and thin so that it is portable. What's more, it should be able to run large software like CAD smoothly and look top grade. I prefer an apple laptop because the products of apple left me with a good impression. The price of the laptop should better be between 5 k and 10 k. Although the price of the same product in Taobao is lower than that in JD and there are many gifts from the Taobao seller, there are so many fakes in Taobao and electronic products should be high quality so I choose to purchase it on JD.

Jing Li is currently an assistant professor at the School of Management of Northwestern Polytechnical University – NPU in Xi'an, China. She obtained her PhD on Advanced Mechanics from Universite de Technologie de Compiegne – UTC, France, in 2015. Her PhD work dealt with the user-customized configuration design method for industrial robots. Previously, she obtained her bachelor's degree in management from NPU in 2009. Now, she is a researcher in requirement engineering, product design management, eco-design, and sustainable manufacturing.

Xinwei Zhang is an associate professor with the School of Management of Northwestern Polytechnical University, NPU in Xi'an, China. He received his BS degree in Management Information Systems in 2005, his MS degree in Management Science and Engineering from Northwestern Polytechnical University in 2008, and his PhD in Industrial Engineering from INSA Toulouse, France in 2012. His research areas are quality management and requirements engineering.

Keqin Wang is an associate professor at the School of Management, Northwestern Polytechnical University, China. He obtained his PhD on Mechanical System and Material from Universite de Technologie de Troyes-UTT, France and on Management Science and Engineering from Northwestern Polytechnical University, China. His research interests include design management, product knowledge engineering, and sustainable design.

Chen Zheng is currently an associate professor at the School of Mechanical Engineering of Northwestern Polytechnical University – NPU in Xi'an, China. Formerly, he was a post-doctoral researcher at the UMR UTC/CNRS 7337 Roberval research group of the Universite de Technologie de Compiegne – UTC, France. He obtained his PhD on Advanced Mechanics from UTC, France, in 2015. His PhD work dealt with the design and integration of multi-disciplinary interfaces in mechatronic systems engineering. Previously, he obtained the double MSc degrees from Xi'an Jiaotong University and UTC.

Professor Shurong Tong is currently the leader of Design Management institution of Northwestern Polytechnical University, NPU in Xi'an, China. She received her BS degree in Mechanical Manufacturing Engineering in 1983, her MS degree in Mechanical Manufacturing Engineering from NPU in

https://doi.org/10.1017/S0890060419000301 Published online by Cambridge University Press

1986, and her PhD in Aerospace Manufacturing Engineering from NPU in 1996. Her research areas are product development management and quality management.

Professor Benoît Eynard is currently the leader of Integrated Product-Process Engineering team of Roberval research institute at the Université de Technologie de Compiègne – UTC, France. In 2007, he has joined UTC as a full professor for leading the Department of Mechanical Systems Engineering until 2012. Previously, he was an assistant professor at the Université de Technologie de Troyes where he managed the MSc program on Information Technology for Mechanical Engineering. Since 2013, he is the general chairman of the academic division on Factories of the Future of the French Society of Mechanical Engineering.

https://doi.org/10.1017/S0890060419000301 Published online by Cambridge University Press