



## Article

# Acquisition Method of User Requirements for Complex Products Based on Data Mining

Juan Hao <sup>1,2</sup>, Xinqin Gao <sup>1,2,\*</sup>, Yong Liu <sup>1,2</sup>  and Zhoupeng Han <sup>1,2</sup> 

<sup>1</sup> School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, China; haojuan1111@163.com (J.H.); liuyong@xaut.edu.cn (Y.L.); hanzp@xaut.edu.cn (Z.H.)

<sup>2</sup> Shaanxi Modern Equipment Green Manufacturing Collaborative Innovation Center, Xi'an University of Technology, Xi'an 710048, China

\* Correspondence: xaut66gxq@163.com

**Abstract:** The vigorous development of big data technology has changed the traditional user requirement acquisition mode of the manufacturing industry. Based on data mining, manufacturing enterprises have the innovation ability to respond quickly to market changes and user requirements. However, in the stage of complex product innovation design, a large amount of design data has not been effectively used, and there are some problems of low efficiency and lack of objectivity of user survey. Therefore, this paper proposes an acquisition method of user requirements based on patent data mining. By constructing a patent data knowledge base, this method combines the Latent Dirichlet Allocation topic model and a K-means algorithm to cluster patent text data to realize the mining of key functional requirements of products. Then, the importance of demand is determined by rough set theory, and the rationality of demand is verified by user importance performance analysis. In this paper, the proposed method is explained and verified by mining the machine tool patent data in CNKI. The results show that this method can effectively improve the efficiency and accuracy of user requirements acquisition, expand the innovative design approach of existing machine tool products, and be applied to other complex product fields with strong versatility.

**Keywords:** data mining; user requirements; topic model; rough set; innovative design



**Citation:** Hao, J.; Gao, X.; Liu, Y.; Han, Z. Acquisition Method of User Requirements for Complex Products Based on Data Mining. *Sustainability* **2023**, *15*, 7566. <https://doi.org/10.3390/su15097566>

Academic Editor: Ripon Kumar Chakraborty

Received: 13 April 2023

Revised: 28 April 2023

Accepted: 3 May 2023

Published: 5 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the development and application of cloud computing, industrial Internet, machine learning, and other new-generation technologies, big data in the manufacturing industry is becoming increasingly abundant, rapidly penetrating into all aspects of the product life cycle, and showing real-time, multi-source, heterogeneous, massive, high-speed, and other characteristics [1]. Therefore, knowledge base and resource base have become important basic data to promote product innovation design [2]. The product innovation model is gradually shifting from the traditional “passive discovery based on user requirements” to “active acquisition driven by data” [3]. For example, digital design based on industrial Internet and user requirements acquisition driven by review data can effectively compensate for the shortcomings of traditional research methods in terms of efficiency and accuracy [4].

Complex products have the characteristics of multi-level structure, a large number of parts, and complex processes [5]. Generally, demand information is obtained by visiting product development and maintenance experts, interviewing users, and other investigation methods. However, due to the short timeliness, the data cannot be processed timely and effectively, which is unable to satisfy the rapid demand for products in the current market. Additionally, the innovative design of complex products not only needs to consider the users' personal feelings and preferences, but also take into account the professional and technical aspects. The application of data mining technology in the field of complex product innovation design can transform online digital files, patents, and user evaluations into

design resources; improve the accuracy of users' personalized requirements; and shorten the development cycle [6,7]. Therefore, through the use of big data and data mining technology [8], manufacturers can not only collect a large number of product-related data in a short time, but also obtain user requirements through the analysis of data, which significantly improves the efficiency and accuracy of product innovation design [9].

However, the current research on user requirements acquisition based on data mining mainly focuses on the mining of online reviews [10]. Extracting product features from users' perceptual evaluation of product appearance and service and conducting emotional analysis is helpful for the innovative design of consumer goods such as mobile phones, computers, washing machines, and automobiles [11]. However, complex products are more focused on the precise acquisition of functions and performance. Users' subjective feelings and feedback cannot easily provide more valuable demand information from a technical perspective [12].

In this context, the acquisition of user requirements in the innovative design process of complex products faces two important issues:

1. There is a lot of product design knowledge in the product life cycle, such as technical reports, patent documents, maintenance records, and user evaluation [13]. The hidden design information is weak in structure, huge in total, and low in information utilization. How can we select and identify valid data and mine it?
2. In user-oriented product evaluation, it is difficult for users to grasp the relationship between demand importance and customer satisfaction with subjective experience [14]. How can we objectively reflect the importance of product requirements to determine user requirements?

Therefore, in order to improve the effectiveness and accuracy of complex product user requirements acquisition, this paper proposes a complex product user requirements acquisition method based on patent data mining that can comprehensively solve the above two main problems. The main contributions of this paper are as follows:

1. Patent data is an important carrier of research and development achievements [15]. The mining of complex product text patents can quickly obtain professional and technical knowledge. In this study, patent data were used to effectively extract available information and transform it into design knowledge. The knowledge base of product design was constructed. The Latent Dirichlet Allocation (LDA) topic model was used to identify product keywords. The similarity of product demand characteristics was calculated based on document–topic probability distribution, and the keywords with similar attributes were clustered. Then, the K-means algorithm was used to conduct the secondary clustering of subject words with similar characteristics, refine the key functional requirements categories, and realize the explicit design requirements.
2. The subject words extracted by patent data mining are often too specialized, which leads to a lack of regularity and the systematization of product functional requirements obtained by the clustering method, which is not conducive to innovative design identification. Rough set theory does not need any prior information to determine the importance, nor does it depend on people's subjective judgment. The calculation results are stable and can objectively and effectively determine the importance of design requirements. In this paper, rough set theory was used to calculate the importance of product innovation design requirements, and the purpose of rapid screening and sorting of requirements was realized. Combined with user satisfaction, importance–performance analysis (IPA) was carried out to obtain complex product innovation design decisions.

In summary, the current research on user requirements acquisition is mostly based on the analysis of review data, and the data mining analysis of patents is less important. The innovative design of complex products based on patent mining has been studied, but the user needs are ignored. This method not only avoids the semantic ambiguity and lack of professionalism caused by users' subjective feelings, but also provides an efficient mining

mechanism to discover the hidden design knowledge of patent data; in other words, it combines the objectivity of patent knowledge with the subjectivity of users. The ultimate goal of this method is to quickly and accurately obtain user requirements and improve the efficiency of complex product innovation design.

The organization of this paper is as follows: Section 2 briefly reviews the relevant works, indicating that the basic principles of this study are reasonable. Section 3 illustrates the framework of user requirements acquisition for complex products based on patent data mining. The key technologies of the method are described in detail in Section 4. Then, Section 5 verifies our research ideas through a case study of machine tool user needs acquisition. Finally, Section 6 summarizes the full text and prospects the future research work.

## 2. Literature Review

This study comprehensively reviews the related works of user requirements acquisition from two aspects of patent data mining methods and demand importance.

### 2.1. Patent Data Mining

Patent data can provide technical reference and prediction for the product design process, and are superior to other data types in terms of data volume, technology, and objectivity [16]. Therefore, patent-data-driven product innovation design has attracted the attention of academia. By retrieving patents related to target functions, Jia et al. used the analogy method to analyze product structure and obtain product design inspiration [17]. Bai et al. identified the hot innovation fields of smart grid at home and abroad by mining patent information [18]. Kim et al. analyzed humanoid robot patent data through various data mining techniques such as topic modeling, cross-impact analysis, association rule mining, and social network analysis, and proposed sustainable strategies and methods [19]. The process of patent data mining and analysis is less manually dependent, which can provide functional requirements and new directions in the field of technology for products. According to the different expression of patent information, the existing research is mainly divided into the following two categories:

One is data mining based on classification numbers. The classification number is the number of technical classifications of patents by the International Property Organization. Wu et al. constructed an enterprise-centered international patent classification (IPC) multi-level supply chain network to discover technological opportunities and support corporate R&D decisions [20]. Georg et al. provided available metadata in the form of IPC symbols, trying to guide topics to easily identifiable labels, enabling experts to quickly acquire advanced technologies [21]. Because the content contained in the classification number is relatively broad, it cannot provide more details for the design.

The other is data mining based on keywords. Keywords are important words extracted from patent titles, abstracts, and claims. Kim et al. extracted the vector of SAO structure from patent documents to identify various factors that need to be considered in patent infringement [22]. Srinivasan et al. searched and identified related patents based on knowledge similarity in patent texts of different products, and obtained keywords by combining network measurement methods [23]. Keywords are important information sources to interpret product technical solutions and innovation, which have strong reference and utilization values. However, there are a large number of auxiliary words and non-disciplinary words in the patent text, and the patent text has a large number of auxiliary words and words that do not reflect the subject content, which increases the difficulty of keyword extraction.

Keyword mining is the process of extracting unknown knowledge from a large amount of unstructured text data and finally obtaining available knowledge [24]. The LDA [25] document–topic generation model is commonly used to identify the topic information hidden in a large-scale document collection or corpus. The model regards the document as a mixed distribution of multiple topics, and the topic as a mixed distribution of

different words, including the three-level structure of the document, the topic, and the word [26], forming a three-level structure map of the “product-category-demand”, which can maximize the independence and hierarchy of the membership relationship between the requirements categories when obtaining the product design requirements. However, the LDA topic model obtains co-occurrence word pairs at the single-text level to discover topics [27]. Patent text is relatively short after cleaning and denoising, and co-occurrence information is insufficient, which is not conducive to generating high-quality topics. Moreover, due to the similarity of product demand, there is a problem in that the same topic word belongs to different categories, the distribution of high-frequency words is relatively uniform, and the semantics are similar. A single method cannot accurately cluster the demand effectively. The K-means clustering algorithm is an iterative clustering analysis algorithm [28] which can effectively solve the semantic sparsity of the LDA model after single-text clustering and improve the quality of patent-knowledge-base-generated topics. In addition, the K-means algorithm is greatly disturbed by the initial clustering center, and it can well correlate the functional requirements by clustering on the premise of obtaining the topic probability distribution.

According to the review, it can be clearly seen that previous studies have used patent mining to obtain product design inspiration. However, previous studies rarely used patent mining to support user requirements acquisition. Therefore, this paper used the web crawler tool to construct the product patent knowledge base and extract keywords. Additionally, the combination of the LDA topic model and K-means clustering algorithm was used to obtain the text topic of keywords.

## 2.2. User Demand Acquisition

In the fierce market competition environment, on the one hand, product innovation design should quickly meet the diversified needs of users; on the other hand, it is necessary to use tools to quantify the importance of requirements. Qi et al., for online reviews, through attribute recognition and sentiment analysis, proposed a method based on joint analysis to determine attribute weights, and combined with the Kano model, proposed a product improvement strategy [29]. Iriland et al., combined with online product reviews, refined the quantitative analysis of product characteristics by naive Bayes and other methods, and proposed better decisions for the design requirements of folding chairs [30]. Kim et al. quantitatively analyzed the factors related to washing machine design in user reviews through linear regression modeling [31]. Chen et al. proposed a method combining demand extraction and evaluation, combining fuzzy sets, rough set, decision-making experiments, evaluation experiments, and analysis network process methods, and proposed a hybrid model for evaluating sustainable value demand [32]. Li constructed a visualization method of design requirements information, and proposed a qualitative adjustment strategy of design requirements information weight based on a strategic coordinate diagram. The above method can effectively allocate the demand weight, but the regularization and systematization of the design system need to be considered.

Rough set theory [33] is a mathematical tool for dealing with imprecise and incomplete information and knowledge, and has been widely used in knowledge discovery, data mining, decision support, and analysis [34]. Rough set theory can be used to define the set to describe the category of product functional requirements, so as to obtain the approximate set definition of functional requirements, study the dependency relationship between functional requirements, and obtain effective demand importance. IPA is a simple and effective user satisfaction evaluation method which is commonly used in market survey [35]. This method consists of performance and importance to form a two-dimensional matrix, and determines the priority of product function improvement by dividing the attributes into four regions: maintaining advantage, possibly excessive, low priority, and concentrated improvement [36]. Importance is an abstract concept; because users are limited by professional technology, it is difficult to provide a reasonable value of importance [37]. Based on

this, this paper uses rough set to determine the importance of functional requirements, and proposes RIPA to improve the objectivity of user requirements acquisition.

In summary, it was found that there is a close dependence between complex product design, product user needs, and product patent data. First of all, the product patent data accumulates the knowledge of complex product design. In the design process, the knowledge evolution itself is constantly improved and expanded, and the application of knowledge iteration can tap the professional user needs, effectively reduce the probability of product innovation failure, and guide the complex product to develop in a feasible direction. The above three concepts form a closed cycle in complex product design.

### 3. Acquisition Method of Data-Driven User Requirements

#### 3.1. Acquisition Process of Data-Driven User Requirements

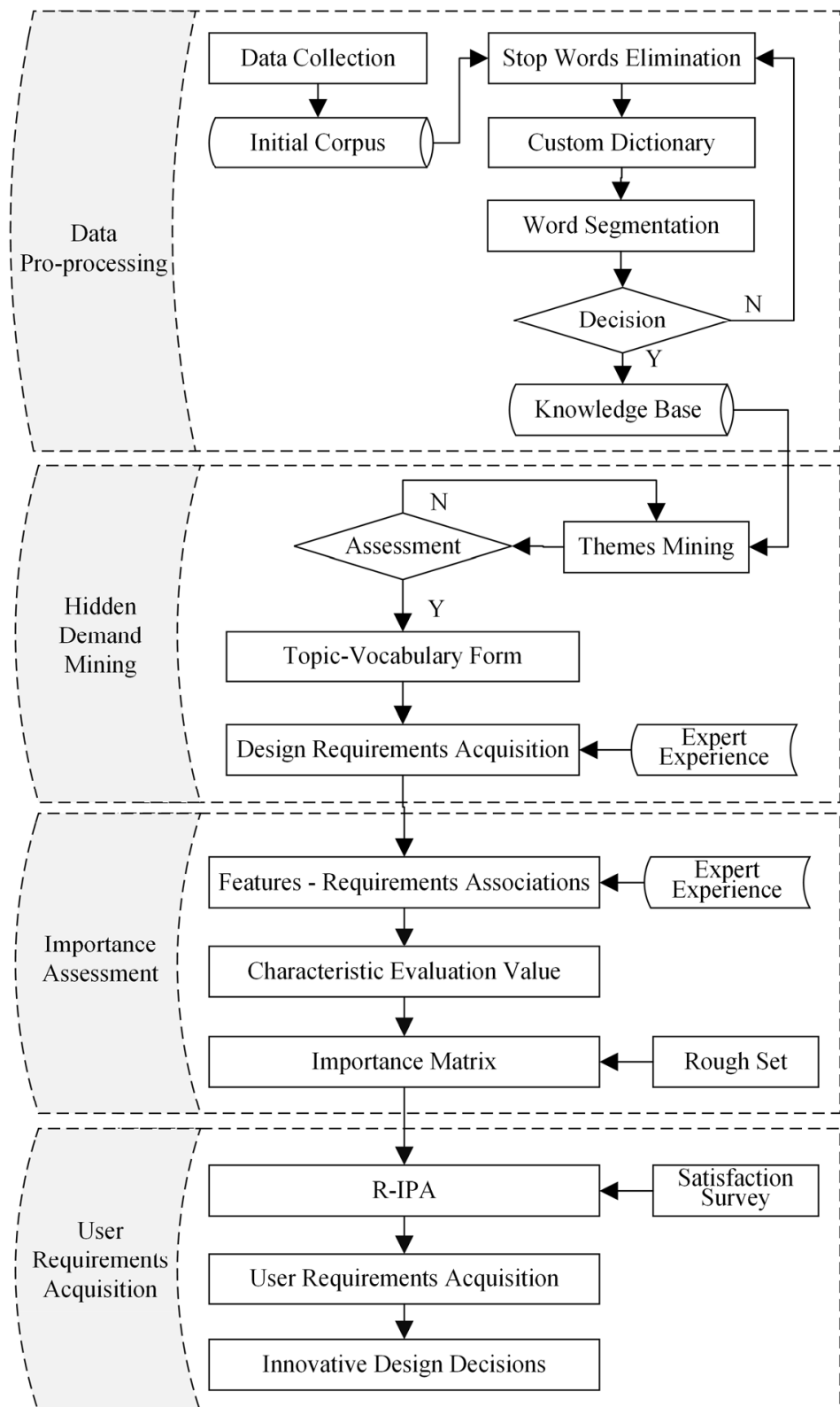
In order to meet the needs of complex product innovation design and improve the competitive advantage of product market, it is necessary to realize the effective use of patent knowledge and achieve the purpose of rapid and accurate acquisition of user needs. By analyzing the characteristics of complex product design, a method based on data mining is proposed, which can quickly complete the patent text data acquisition and knowledge extraction. Through users' satisfaction evaluation of the product requirements extracted from the patent text, the design knowledge can be transformed, and the professionalism and accuracy of the user requirements can be improved. This section describes the whole process of the method and briefly introduces each stage, as shown in Figure 1. For the two problems mentioned in this paper, this method can be divided into the following four stages. The first stage is information collection and data preprocessing, and the second stage is hidden demand mining. These two stages are the application of patent data mining technology in demand acquisition. The third stage is importance evaluation, and the fourth stage is user requirements acquisition. These two stages are the objective evaluation of user demand and the judgment of innovative design direction.

In the first stage, the patent text is selected as the data collection sample in the Web data source to construct the initial corpus. In order to reduce the interference information, the initial corpus is processed, the stop word list and the custom dictionary are constructed, the text redundancy is removed, and the data analysis effect and the acquisition accuracy are improved. The patent abstract text after cleaning and denoising is segmented to obtain the target data set and construct the product knowledge base.

In the second stage, the knowledge base is mined for text topics. Through the text clustering method combining LDA and K-means, high-frequency words are obtained from massive patent texts, the potential topics of texts are clustered, and the hierarchical relationship between topics and words is established. Continuous iteration until the clustering of topics obtains the best-quality evaluation in line with product features. Combined with the professional knowledge of the product field, the hierarchical relationship is established according to the similar attributes of product design topics to obtain functional requirements.

In the third stage, combined with expert experience, the association rules of innovative design characteristics demand categories are constructed, and the relative importance evaluation values between various requirements are obtained. Rough set theory is used to calculate the importance ranking of the design requirements.

In the final stage, according to the user satisfaction surveys of design requirements, the importance performance analysis is carried out to evaluate whether effective user needs can be extracted and transformed into knowledge to help managers understand user needs and determine effective innovative design decisions for complex products.



**Figure 1.** Data-driven user requirements acquisition process.

### 3.2. Information Collection and Preprocessing

Complex product patent texts contain a large number of words unrelated to product function and design knowledge, which affects data quality. Therefore, it is necessary



to preprocess the collected data and extract valuable design knowledge. The steps are as follows:

1. Obtain the initial corpus. The target product is determined from many complex devices, and the patent text is selected as the data collection sample in the Web data source to construct the initial corpus.
2. Eliminate the stop words. The information content of the preliminary collected product patent abstract text is mixed, and the value density is low. It contains some verbs, adjectives, adverbs, quantifiers, pronouns, and so on. It has a high frequency but does not reflect the subject content, such as “proposed, constructed, reliable, one, based”, and so on. In order to reduce the interference information and ensure the accuracy of the word segmentation results, according to the characteristics of the product, a stop words list is built and saved in the TXT text format of UTF-8 using natural language processing software jieba to filter text information, removing redundant words, and improving data analysis results.
3. Construct a custom dictionary. Since word segmentation software is a word segmentation mechanism based on popular vocabulary, a misclassification of professional terms may occur. The text data used in this paper is the patent text of complex products, which contains a large number of professional terms. Therefore, in order to improve the accuracy of the word segmentation, this paper constructs a custom dictionary. Taking a gun drill machine as an example, before using the custom dictionary, the word segmentation software divides “ball screw, coupling sleeve” into “ball, screw, coupling, sleeve”. The result loses the meaning contained in the professional terms. Therefore, adding the professional vocabulary to the custom dictionary before word segmentation not only avoids the lack of key needs, but also improves the accuracy of the acquisition. Additionally, it is more in line with professional background knowledge.
4. Conduct word segmentation and acquire available knowledge base. The patent abstract text, after cleaning and denoising, is segmented to obtain the available corpus. If the vocabulary does not meet the professional requirements, the initial corpus needs to be reprocessed. Constantly update the stop-words dictionary and the custom dictionary to achieve the purpose of accurate word segmentation. Follow the above steps to iterate, obtain the target data set, and finally obtain the available knowledge base.

### 3.3. Text Clustering Algorithm and Requirement Topic Extraction

The collected patent text is preprocessed to obtain the available knowledge base for topic mining. Firstly, we set the number of topics;  $K$  is the number of topics artificially set before model training. Clustering results vary with the number of topics. Then, LDA topic modeling is performed, assuming that there are  $|M|$  patent texts in the available corpus:  $M = \{m_1, m_2, \dots, m_{|M|}\}$ .  $T = \{t_1, t_2, \dots, t_K\}$  is used to represent the topic information obtained by LDA topic model training. By continuously training and mining the topics hidden in the patent text data, the probability distribution of the text on the topic and the probability distribution of words on the topic are obtained. The relationships between all variables in the model are expressed as follows:

$$p(w_i, z_i, \theta_i, \phi | \alpha, \beta) = \prod_{i=1}^N p(w_{i,j} | \phi_{z_{i,j}}) p(z_{i,j} | \theta_i) \cdot p(\theta_i | \alpha) p(\phi | \beta) \quad (1)$$

Among them,  $p(\alpha)$  is a prior distribution with parameter  $\alpha$  for LDA sampling to generate the corresponding demand category distribution of patent text.  $p(\beta)$  is a prior distribution with parameter  $\beta$ , and the distribution of demand words corresponding to demand categories is generated by sampling in the model [24]. During the implementation of LDA, we try again and again to select the appropriate  $K$ . Gibbs sampling is a simple implementation form of the Monte Carlo algorithm, and the probability distribution of the text topic–word and document–topic path can be obtained through Gibbs sampling. Therefore, each document  $m_i$  has a probability distribution on the topic text set  $T$ , that is,

$T^i = \{t_1^i, t_2^i, \dots, t_K^i\}$ . The topic distribution vector of the patent text can be expressed as a text–topic two-dimensional matrix:  $D = \{M, T\} = \{T_1, T_2, \dots, T_{|N|}\}$ . The two-dimensional matrix is used as the input of K-means clustering.

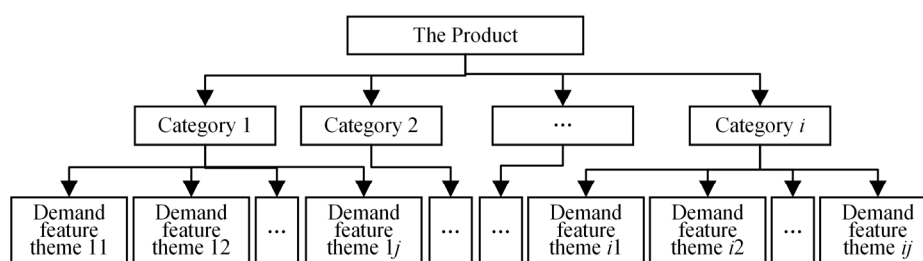
Among the  $n$  high-frequency words obtained by the LDA topic model,  $k$  samples are randomly selected as the initial clustering center  $C = (c_1, c_2, \dots, c_k)$ . We find the nearest cluster center  $c_r$  for each sample  $x_i$  and assign it to the topic  $u_i$  specified by  $c_r$ . The average method is used to calculate the cluster centers after reclassification.

$$C = \sum_{i=1}^n \left[ \min_{r=1,2,\dots,k} d(n_i, c_r)^2 \right] \quad (2)$$

The clustering center constitutes the nodes of the product demand theme tree. A different number of nodes will have different effects on the clustering results. The number of nodes that meet the range of the structural number of the target product can be selected as the number of clustering centers. In Formula (2), the chord distance between the high-frequency word  $n_i$  and the clustering center  $c_r$  is calculated as:

$$d(n_i, c_r) = \frac{\sum_{k=1}^n P_{ik} P_{rk}}{\sqrt{\sum_{k=1}^n P_{ik}^2} \sqrt{\sum_{k=1}^n P_{rk}^2}} \quad (3)$$

The sum of distances between all high-frequency words and their associated cluster centers is minimized, and the high-frequency words within each class form a branch of the text topic tree. However, the text topic tree cannot directly reflect the product–demand relationship. Therefore, the product–demand tree is constructed according to the three levels of product, demand category, and demand feature theme, as shown in Figure 2.



**Figure 2.** The Requirements Tree.

Precision, recall, and  $F$ -measure values are used to evaluate the quality of the clustering results [38]. The calculation rules are as follows:

$$\begin{cases} P = precision(i, j) = \frac{N_{ij}}{N_i} \\ R = recall(i, j) = \frac{N_{ij}}{N_j} \\ F(i, j) = \frac{2 \times P \times R}{P + R} \end{cases} \quad (4)$$

$N_i$  represents the number of samples with characteristic  $i$  in the corpus.  $N_j$  represents the number of samples with feature  $j$  in the clustering results.  $N_{ij}$  represents the number of correct samples in the clustering results.

By setting different numbers of topics, the matching degree between the clustering topics and the product functional requirements is adjusted. Finally, a quality evaluation that meets industry requirements and product characteristics is obtained.



### 3.4. Demand Importance Based on Rough Set Theory

In this paper, rough set theory was introduced into the demand acquisition problem. The partition of any concept family on domain  $E$  is called a relation system about  $E$ . Let  $P$  be the product demand category, which is a set of equivalence relations on  $E$ . The relation system can be expressed as  $U = (E, P)$ . Let  $[n]_P$  be the equivalent class of  $P$  containing the functional requirements category,  $n_i$  represent the sub-attribute of the product functional requirements, and  $PN_i$  represent the rough number of the functional requirements.

To participate in product design and development, functional requirements need to be selected according to different design characteristics. Therefore, industry experts define design feature–demand association rules based on experience and technical characteristics, associate product design features with functional requirements, and assign relative importance evaluation values between functional requirements. The evaluation value represents the degree of correlation between design requirements and design features. The larger the scale value, the greater the correlation between design features and design requirements.

Product design characteristics are defined as decision-making subjects  $S = \{S_1, S_2, \dots, S_u\}$ , corresponding to  $u$  design characteristics, respectively. The evaluation value assigned by decision-maker  $S$  adopts the scale of 1–9, indicating the relative importance of each demand. Among them, 9 indicates that the relative importance is very high. As the scale value decreases, the impact of the product design requirements decreases. The equivalent importance of design requirements is represented by 5, and 1 indicates that the relative importance is very low.

In rough set theory, the decision set  $S_i$  can be represented by the rough number  $PN_i$ , and the importance evaluation value is transformed into the rough value. The rough upper limit and rough lower limit of  $PN_i$  are defined as  $\overline{\lim}(S_i)$  and  $\underline{\lim}(S_i)$ . The rough number with upper and lower limits can better reflect the uncertainty of product demand while maintaining the objectivity of the original design.

The importance of product requirements is expressed in rough set theory, as shown in Formula (5).

$$\omega_i = \frac{\frac{1}{m} \sum_{j=1}^m \frac{1}{u} \sum_{v=1}^u PN_i(S_{ijv})}{\sum_{i=1}^m \frac{1}{m} \sum_{j=1}^m \frac{1}{u} \sum_{v=1}^u PN_i(S_{ijv})} = \frac{\sum_{j=1}^m \sum_{v=1}^u [\underline{\lim}(S_{ijv}), \overline{\lim}(S_{ijv})]}{\sum_{i=1}^m \sum_{j=1}^m \sum_{v=1}^u [\underline{\lim}(S_{ijv}), \overline{\lim}(S_{ijv})]} \quad (5)$$

In the formula,  $m$  is the number of product requirements;  $u$  is the number of decision makers; the evaluation value  $S_{ijv}$  of the  $i$ th requirement relative to the  $j$ th requirement is given by the  $v$ th decision maker;  $PN_i(S_{ijv})$  is the rough value of the  $i$ th demand relative to the  $j$ th demand; and  $\overline{\lim}(S_{ijv})$  and  $\underline{\lim}(S_{ijv})$  are the upper and lower bounds of the rough set boundary, respectively. The rough set is a set of upper and lower boundary numbers.

According to Formula (5), the demand importance matrix and the degree of correlation between product requirements are obtained, which can determine the multiple innovative design directions of complex products.

### 3.5. Importance Performance Analysis

The core of IPA method is the determination of importance and performance value. In the implementation process, the performance of each demand feature of the product is scored by the user. By drawing the analysis diagram composed of importance based on rough set and performance value based on user evaluation, the design requirements attributes are divided into four categories: low priority, excessive performance, maintaining advantages, and key improvement. Accordingly, the priority and management strategy of attribute improvement are determined, and the overall user satisfaction is improved. The steps of importance performance analysis are as follows:

1. Collect users' performance evaluation values of the design requirements, and assign any number between 0–1 according to users' satisfaction with the design requirements.

2. For each design requirement, calculate the average according to all user rating values as the vertical axis input of the IPA diagram. For the  $i$ th design requirement, average the sum of the satisfaction ratings of  $n$  users to obtain the average performance of a single design requirement, as shown in Formula (6).

$$Y_t = \sum_{i=1}^n Y_{ti} / n \quad (6)$$

In Formula (6),  $n$  is the number of users, which indicates the satisfaction evaluation value of user  $t$  to design requirements  $i$ .

3. Use the design requirements' importance based on rough set in Section 3.4 as the horizontal axis input of the IPA diagram. Reasoning the relationship between design requirements data reasonably avoids subjective interference.
4. Build a biaxial coordinate system based on the average score of all users for a single feature, forming an R-IPA diagram, as shown in Figure 3. According to the region where the scattered points representing the demand for innovative design in the R-IPA figure are located, analyze the direction of the product innovation, and make reasonable design and development decisions.

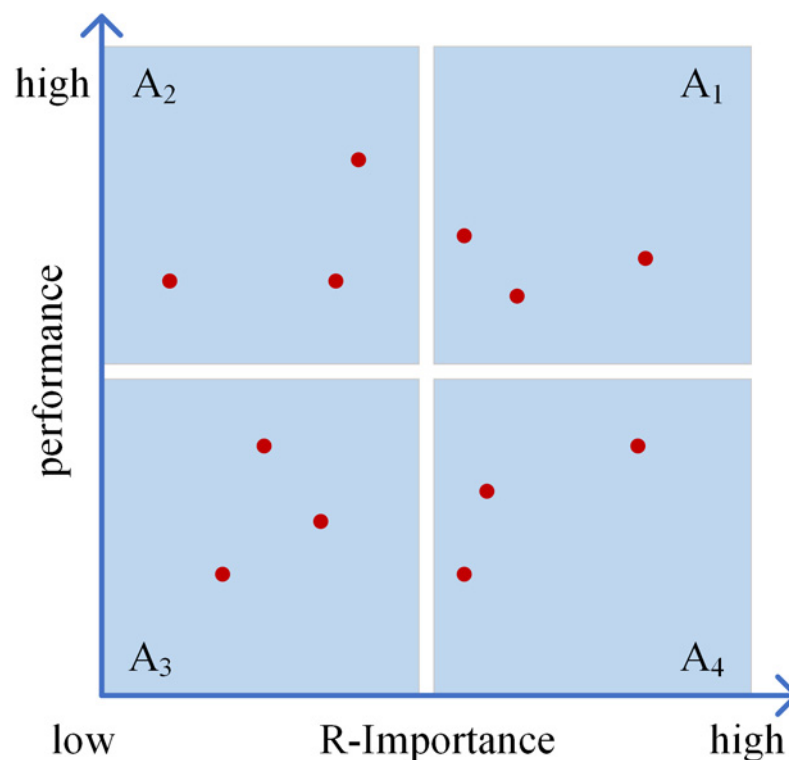


Figure 3. The R-IPA figure.

In Figure 3,  $A_1$  is the region that maintains the advantage, where the demand importance is relatively high, the user satisfaction is high, and the performance is excellent.  $A_2$  is a region that may be excessive, where the user satisfaction is high, and it is the user-excited innovation demand, but its importance is relatively low, and it may be excessive for the function of the product itself.  $A_3$  is a low-priority area, where user satisfaction and importance are generally low.  $A_4$  is the concentrated improvement area, where the average performance is low, and belongs to the basic demand of products.

#### 4. Case Study

With the rapid development of aviation, aerospace, shipbuilding, automobile, and other fields, the demand for the personalized customization of CNC machine tools is increasing. In order to shorten the design cycle of the drilling machine and improve the reliability of the acquisition of design requirements, this paper uses the user demand acquisition process of the drilling machine as an example to verify the user requirements acquisition method of complex products based on data mining.

##### 4.1. Information Collection and Preprocessing

In order to ensure the reliability of machine tool information sources, the related patents in the CNKI patent database were selected as the data collection objects to obtain the initial corpus. The CNKI database has been widely used in related research [39,40]. This paper used the patent name of the drilling machine as the retrieval condition, and the application date was from 1 January 2000 to 1 January 2023. After screening, 296 related patent texts were obtained, with a total of 142,300 characters.

According to the actual research needs, the deactivated dictionary and the custom dictionary were constructed and updated iteratively. These dictionaries can filter out words unrelated to the drilling machine and prepare for the construction of the design knowledge base.

##### 4.2. Topic Clustering of Requirements Information

We used Python 3.6 and the Gensim module to train the corpus. The prior parameters  $\alpha$  and  $\beta$  in the LDA model were set as  $\alpha = 50/K$  and  $\beta = 0.01$  according to reference [41], and then Gibbs sampling was used to continuously train the iterations until convergence. For  $M = 296$  abstract texts, we set the number of topics  $K$  to 10, 20, 30, 40, 50, and 60, respectively. After model training, it was found that  $K = 50$  has the best clustering effect, and the text clustering vocabulary has obvious thematic significance. Because of the small difference in probability values, the top 15 topics were selected, and the top 8 words with the highest probability were selected for analysis under each topic, as shown in Figure 4.

It can be seen from Figure 4 that there were repeated words in different themes, such as “workbench” in theme 1 and 3, indicating that a word belongs to different product demand categories according to different classifications. This is the defect of LDA topic clustering. According to the meaning of 15 topics and the distribution probability of keywords, the words representing similar attributes were extracted. Combined with expert opinions, 15 types of demand characteristics were selected as follows: TP1—Workbench, TP2—Feed, TP3—Appearance, TP4—Positioning, TP5—Tool storage, TP6—Tool change, TP7—Guide rail, TP8—Cuttings, TP9—Servo, TP10—Tool setting, TP11—Shield, TP12—Induction, TP13—Connection, TP14—Collection, and TP15—Installation. For example, the top-ranking keywords in topic 5 were Tool library, Cutter, Drill pipe, Spring, Manipulator, Location, Clamping, and Claw. These words are related to tools and storage, reflecting the product requirements of drilling tools and their storage methods. Therefore, topic 5 was named Tool library storage.

K-means clustering was performed according to the topic distribution shown in Figure 4. Based on the significance and probability distribution of each topic, it was aggregated into eight demand categories. According to the requirements reflected and represented by the vocabulary, eight categories were named. They were DR1—Motion function requirement, DR2—Chip removal requirement, DR3—Control and detection requirement, DR4—Protection function requirement, DR5—Appearance requirement, DR6—Operation performance requirement, DR7—Support function requirement, and DR8—Tool library requirement. The product requirements tree was built as shown in Figure 5.

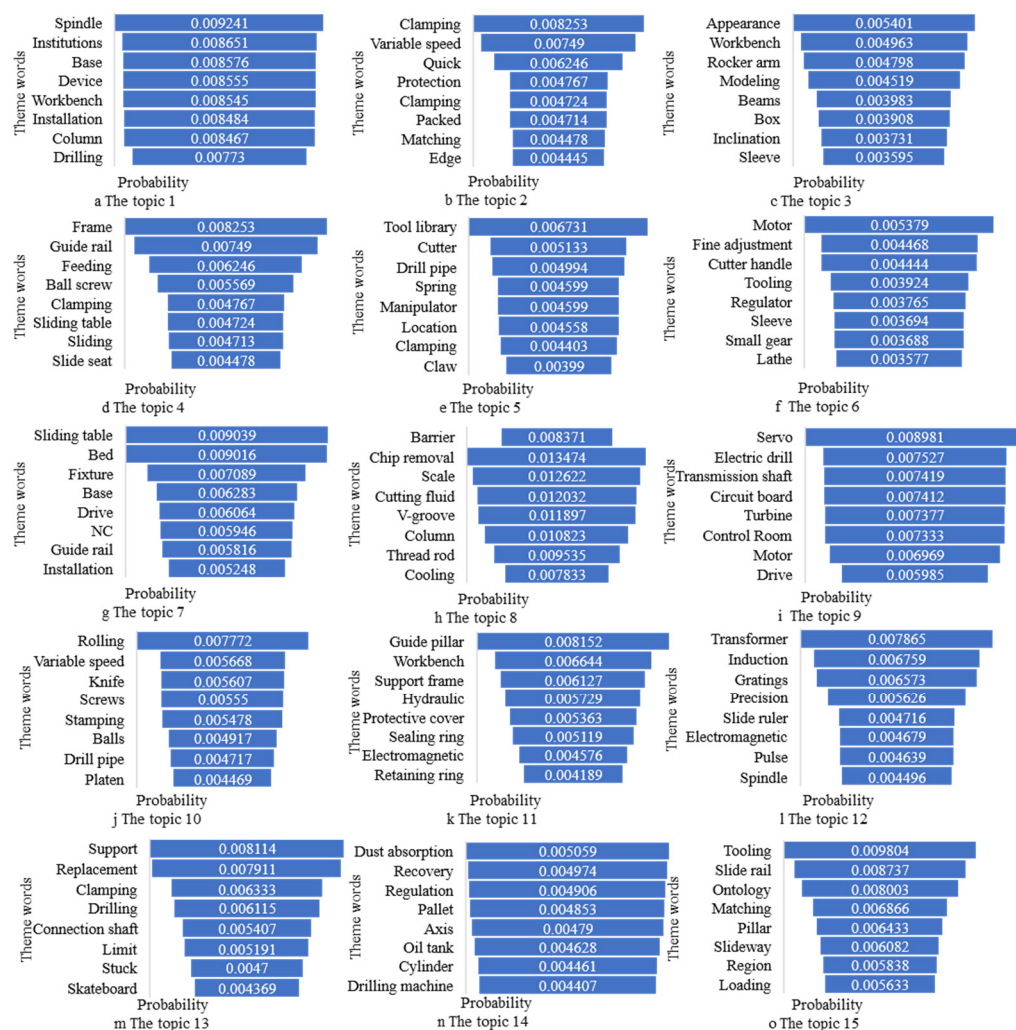


Figure 4. The theme words.

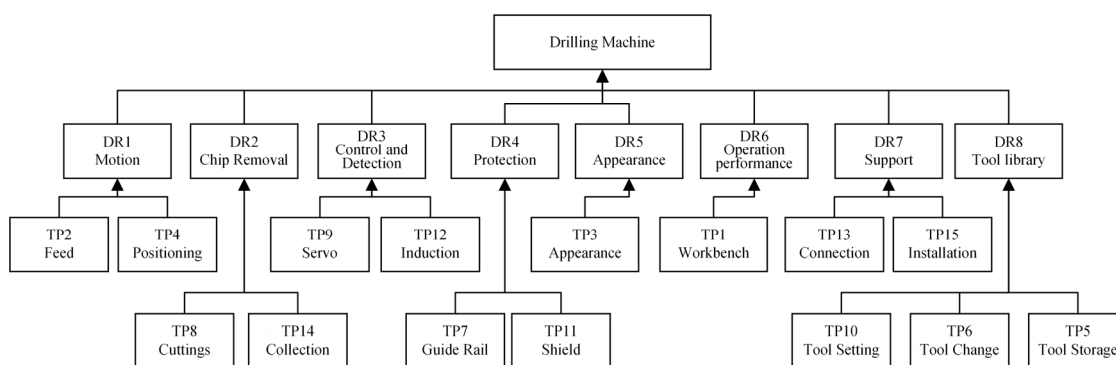
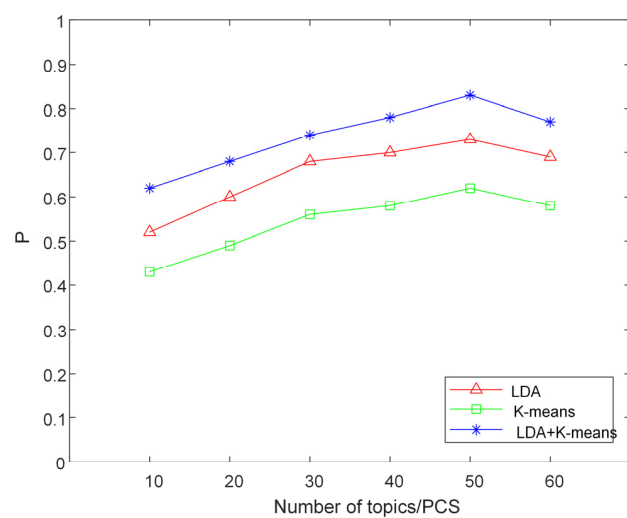
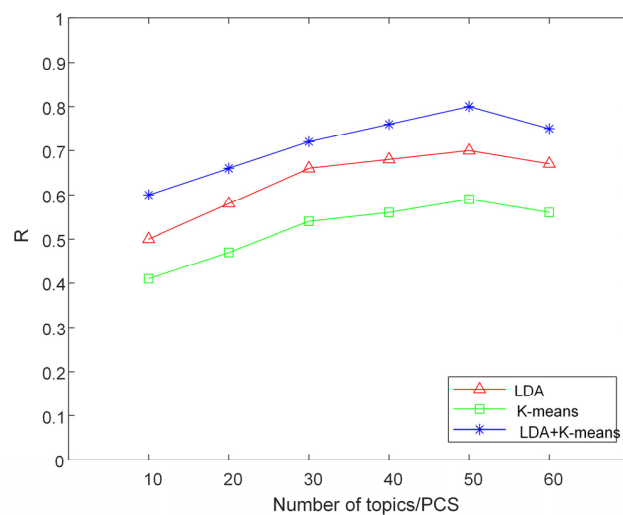


Figure 5. The product requirements tree.

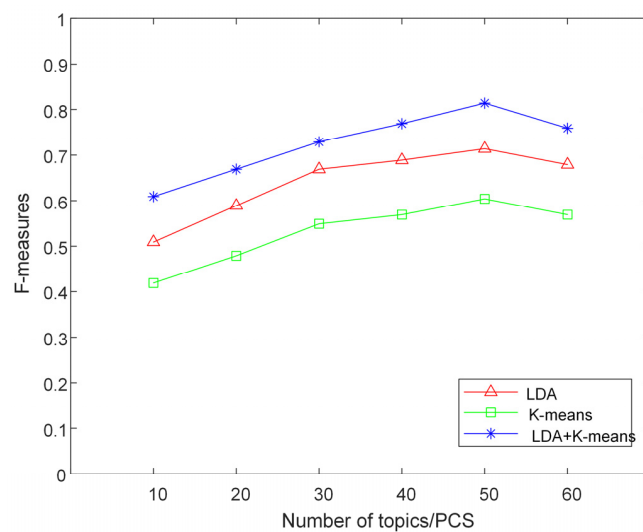
In order to verify the effectiveness of the LDA + K-means algorithm method, the LDA clustering in reference [42] and the K-means clustering algorithm in reference [43] were used for topic mining. Three sets of experiments were completed in the same environment. Figure 6 shows the comparison of the change in precision, recall, and F-measures by using LDA, K-means, and LDA + K-means. The abscissa is the number of topics, and the ordinates of Figure 6a–c represent the precision, recall, and F-measures of clustering, in turn.



(a)



(b)



(c)

**Figure 6.** The comparison of three algorithms. (a) The precision. (b) The recall. (c) The *F*-measures.

It can be seen from the figure that the precision, recall, and *F*-measures obtained by the proposed method are the highest. In addition, in the implementation process of the algorithm, when the number of topics is set to 10, the effect is the worst. With the increase in the number of topics, the clustering effect becomes better and better. When the number of topics is 50, the effect is the best. Then, with the increase in the number of topics, the clustering effect shows a downward trend. Therefore, it is most reasonable to set the number of topics *K* to 50.

Table 1 shows the comparison of precision, recall, and *F*-measures obtained by different clustering methods when the number of topics is set to 50. Compared with the LDA and K-means algorithm methods, the *F*-measures of LDA + K-means algorithm are increased by 13.21% and 34.77%, respectively, which verifies the effectiveness of the proposed algorithm. Through analysis, it can be seen that because there are many feature words about drilling machine products in the patent texts, the LDA model extracts topic words, obtains probability distribution, and then performs K-means clustering, which can effectively reduce the interference of the initial clustering centers and improve the accuracy of the drilling machine abstract text mining. K-means clustering uses iterative solution to effectively solve the semantic sparsity after LDA text clustering and improve the recall rate of drilling machine summary text mining.

**Table 1.** The comparison of clustering algorithms.

Algorithm	Precision	Recall	<i>F</i> -Measures
LDA	0.74	0.70	0.719
K-means	0.62	0.59	0.604
LDA + K-means	0.83	0.80	0.814

#### 4.3. Requirements Importance Determination

Due to the different descriptions of demand from various innovative design characteristics, the determination of demand weight in drilling machine design encounters conflicts, and the key points are difficult to grasp. Six professors of mechanical design and six engineers of a machine tool factory were invited to participate as experts in the field. According to the expert experience and the characteristics of deep hole processing technology, four innovative design characteristics—adaptability, reliability, economy, and environmental protection—related to the demand were chosen. Association rules are shown in Table 2.

**Table 2.** Product innovative design characteristics–demand association rules.

Innovative Design Features	Correlation Basis
Adaptability	High degree of automation, strong adaptability to processing objects
Reliability	Maintaining cutting accuracy, low failure, feed rate, spindle speed
Economy	Low energy consumption, low maintenance cost, easy disassembly and assembly
Environmental protection	Environmentally friendly, reconfigurable, and recyclable materials

Let  $S = \{S1, S2, S3, S4\}$  correspond to four innovative design characteristics, respectively. According to Figures 4 and 5 and Table 2, eight types of requirements were evaluated. Taking the economic design characteristics as an example, the evaluation values of each demand module are shown in Table 3. The evaluation value represents the degree of correlation between the design requirements modules under the economic design characteristics. The highest degree of correlation is expressed by 9, and the lowest degree of correlation is expressed by 1.



**Table 3.** Evaluation value of design characteristics.

Economy	DR1	DR2	DR3	DR4	DR5	DR6	DR7	DR8
DR1	0	6	9	7	9	9	8	6
DR2	4	0	7	6	8	8	6	5
DR3	2	4	0	4	6	6	5	2
DR4	3	5	6	0	7	7	7	5
DR5	2	3	4	4	0	5	5	2
DR6	2	3	4	4	5	0	6	2
DR7	3	5	4	4	5	4	0	3
DR8	7	6	8	6	9	9	8	0

The importance evaluation values of the four design characteristics were determined and converted into rough values. For example, the evaluation value of the correlation degree of DR2 relative to DR8 in the four design characteristics is {6, 6, 9, 5} and converted into rough values is {[5, 6], [5, 6], [8, 9], [4, 5]}. Then, the element  $c_{28} = ([5, 6] + [5, 6] + [8, 9] + [4, 5])/4 = [3, 4]$ .

According to Formula (5), the importance weight of each demand category  $\omega_i = [0.18, 0.14, 0.11, 0.12, 0.10, 0.11, 0.09, 0.16]$  was determined, and the importance matrix of innovative design requirements for drilling machine tools was constructed, as shown in Table 4.

**Table 4.** Importance Matrix of Innovative Design Requirements.

Innovative Design Requirements	DR1	DR2	DR3	DR4	DR5	DR6	DR7	DR8	$PN_i$	$\omega_i$
DR1	[0, 0]	[6, 7]	[7, 8]	[6, 7]	[7, 8]	[7, 8]	[7, 8]	[6, 7]	[46, 53]	0.177419
DR2	[4, 5]	[0, 0]	[5, 6]	[4, 5]	[6, 7]	[5, 6]	[5, 6]	[3, 4]	[32, 45]	0.137993
DR3	[2, 3]	[4, 5]	[0, 0]	[4, 5]	[5, 6]	[4, 5]	[5, 6]	[3, 4]	[27, 34]	0.109319
DR4	[4, 5]	[5, 6]	[4, 5]	[0, 0]	[5, 6]	[5, 6]	[4, 5]	[4, 5]	[31, 38]	0.123656
DR5	[2, 3]	[4, 5]	[3, 4]	[4, 5]	[0, 0]	[4, 5]	[4, 5]	[3, 4]	[24, 31]	0.098566
DR6	[2, 3]	[4, 5]	[4, 5]	[4, 5]	[4, 5]	[0, 0]	[5, 6]	[3, 4]	[26, 33]	0.105735
DR7	[2, 3]	[4, 5]	[4, 5]	[3, 4]	[3, 4]	[3, 4]	[0, 0]	[3, 4]	[22, 29]	0.091398
DR8	[4, 5]	[5, 6]	[6, 7]	[5, 6]	[6, 7]	[7, 8]	[7, 8]	[0, 0]	[40, 47]	0.155914

It can be seen from Table 4 that the order of importance of drilling machine innovative requirements is DR1 > DR8 > DR2 > DR4 > DR3 > DR6 > DR5 > DR7.

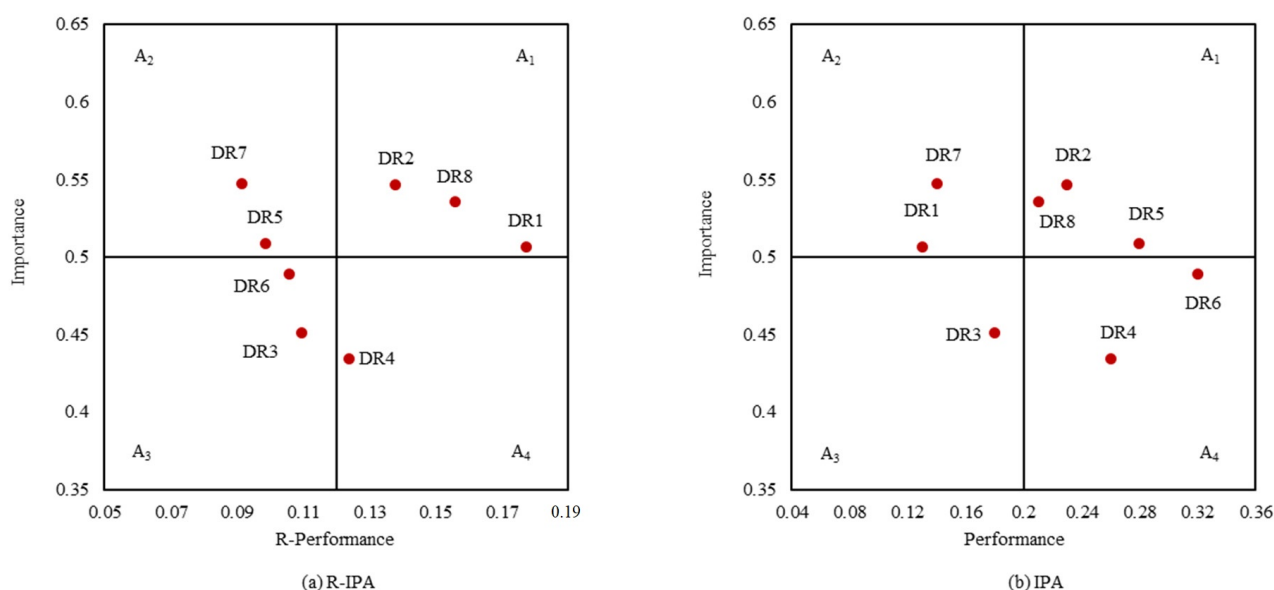
#### 4.4. Analysis and Discussion

In this experiment, 50 users of a certain type of drilling machine were collected to conduct a satisfaction survey on eight types of innovative design requirements. According to the survey results, the average performance of each type of user requirements was calculated by Formula (6), as shown in Table 5.

**Table 5.** The average performance of user requirements.

User Requirements	Average Performance	User Requirements	Average Performance
DR1	0.506403	DR5	0.508715
DR2	0.546499	DR6	0.488646
DR3	0.450823	DR7	0.547021
DR4	0.434433	DR8	0.506403

According to the importance value of Table 4 and the average performance value of Table 5, the IPA diagram was drawn and compared with the analysis results of the traditional IPA method, as shown in Figure 7.



**Figure 7.** Results Comparison.

According to the comparison in Figure 7a,b, it can be found that the quadrants of DR1—Motion function, DR5—Appearance, and DR6—Operation performance of the requirements category are significantly different. In Figure 7b, due to the users' focus on the comfort of the product appearance experience and operation performance, the given importance is higher, resulting in higher demand priority. However, the users lack professional technical knowledge and ignore the motion function of the product, resulting in lower demand priority. However, if the function is poorly performed in long-term use, it causes strong dissatisfaction. The demand priority in Figure 7a is more objective and rational. From searching machine tool patents in CNKI, the top three patents were found to be transmission, chip removal, and tool library. This is basically consistent with the analysis results in Figure 7a, which verifies the effectiveness and feasibility of the proposed method.

It can be seen from Figure 7a that eight categories of innovative design demand are distributed in the four quadrants of the R-IPA figure and divided into four groups. A<sub>1</sub> includes three types of requirements: DR1—Motion function, DR2—Chip removal processing, and DR8—Tool library. This indicates that the users have urgent requirements for the motion function, chip removal processing, and tool magazine function of this type of drilling machine, and the importance of these three requirements is the highest. Therefore, it is necessary to maintain a high design priority for the functions of feed, positioning, chip removal, collection, tool setting, tool changing, and tool storage. For example, it has or improves the function design of automatic tool changer, tool changer door, multi-tool storage, fast chip-to-chip tool changer, and workpiece spraying device, so as to reduce processing preparation time, improve processing efficiency, and effectively protect and save resources. A<sub>2</sub> may include two types of requirements, DR5—Appearance and DR7—Support function, indicating that the users have high demand for machine tool appearance, installation, and connection, but the design importance is low and belongs to the users' excited demand. However, it does not affect the use function of the machine tool, which belongs to the over-design characteristic, and only needs a slight improvement. The low-priority A<sub>3</sub> group includes two types of requirements, DR3—Control detection and DR6—Operation performance, and the user satisfaction and importance are low, indicating that the user does not have a strong demand for the innovative design of Servo, Induction, and Workbench. The two types of requirements in patents are not highly concerned, so there is no need to invest too much research and development. The A<sub>4</sub> centralized improvement group includes DR4—Protection function, indicating high importance. Designers pay more attention to the protection of machine damage during machine operation and the safety

protection of operators. Although users only consider their own factors, such design needs also need to be taken seriously.

## 5. Discussion

This study combines patent text data to mine professional product design knowledge so as to extract the user requirements of complex products. Obtaining the technical characteristics and user requirements of specific complex products from patent data provides richer research perspectives and more sufficient evidence for identifying more promising innovative products. The proposed research framework is based on the collection–mining–acquisition–evaluation process, which is a dynamic system process. Since this study is based on the analysis of existing patent text data, the innovation prospects have certain limitations over time. However, after constantly updating the data, the method provides new ideas for enterprises and R&D personnel in the field of complex products.

In terms of research methods and techniques, in order to illustrate the feasibility of the methods and techniques used in this study, it is necessary to compare it with and analyze other methods. In the existing literature, parallel code-phase acquisition (PCA) [44], the best–worst method (BWM) [45], fuzzy analytical hierarchy process (AHP), and the fuzzy analytical network process (ANP) [46] are often used to collect information from a technical perspective. Among them, PCA is more often used in signal processing. However, this study is more suitable for mining patent text data and using natural language processing methods. The best–worst method (BWM) can obtain consistent results with less comparison information, and the fuzzy analytical hierarchy process (AHP) and the fuzzy analytical network process (ANP) are used to determine subjective weights. However, the amount of research data in the text is large, and the objective evaluation results are emphasized. Rough set theory can assign weights to design requirements more objectively. Therefore, this study applies these methods and techniques to improve the comprehensiveness and accuracy of user requirements acquisition.

## 6. Conclusions

This study proposes a method of user demand acquisition based on patent data mining which can provide innovative ideas for enterprises and R&D personnel in the field of complex products.

The contributions of this study are as follows. From a technical point of view, this paper extracts the knowledge of complex product innovation design by mining patent text data, and proposes the method of user demand acquisition combined with user satisfaction analysis. It saves data collection time, realizes the combination of objective knowledge and subjective feelings, and solves the problems of strong subjectivity and insufficient professionalism in traditional user surveys. From the perspective of method, according to the similar attributes of design knowledge, the LDA + K-means topic model clustering method is proposed to mine and analyze the topic words of patent text data, which improves the efficiency and accuracy of knowledge acquisition. Based on rough set theory, according to product design characteristics–demand association rules, the importance ranking of demand categories is carried out, and the rationality of importance ranking is verified by user IPA, which provides decision making for product innovation design. From an application perspective, the method of user requirements acquisition for complex products based on data mining is applied to mine and analyze the patent data of drilling machines in CNKI, which provides a new method for the user requirements acquisition of machine tools, and can be applied to the user demand acquisition of other complex products with strong versatility.

Despite the contributions, this study still has some limitations and needs further study. First of all, the user requirements in this study were extracted based on patent text data, which has strong objectivity and insufficient access to user experience. Moreover, patents are time sensitive and suitable for product innovation in the short term. In future research,

we will try to further explore and analyze the data accumulated in the product life cycle, so as to maintain the sustainability of innovative ideas.

**Author Contributions:** Conceptualization, X.G.; Methodology, J.H. and X.G.; Formal analysis, J.H.; Investigation, J.H.; Writing—original draft, J.H.; Writing—review & editing, Y.L. and Z.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, grant number 51575443, and the Key Scientific Research Program of Shaanxi Provincial Education Department, China, grant number 20JY047.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zhan, Y.; Tan, K.H.; Huo, B. Bridging customer knowledge to innovative product development: A data mining approach. *Int. J. Prod. Res.* **2019**, *57*, 6335–6350. [\[CrossRef\]](#)
2. Zhang, Z.; Peng, Q.; Gu, P. Improvement of user involvement in product design. *Procedia CIRP* **2015**, *36*, 267–272. [\[CrossRef\]](#)
3. Li, Y.; Sha, K.; Li, H.; Wang, Y.; Dong, Y.N.; Feng, J.; Zhang, S.; Chen, Y. Improving the elicitation of critical customer requirements through an understanding of their sensitivity. *Res. Eng. Des.* **2023**, *16*, 1–20. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Iwasaki, K.; Kuriyama, Y.; Kondoh, S.; Shirayori, A. Structuring engineers' implicit knowledge of forming process design by using a graph model. *Procedia CIRP* **2018**, *67*, 563–568. [\[CrossRef\]](#)
5. Li, J.; Nie, Y.; Zhang, X.; Wang, K.; Tong, S.; Eynard, B. A framework method of user-participation configuration design for complex products. *Procedia CIRP* **2018**, *70*, 451–456. [\[CrossRef\]](#)
6. Tao, F.; Cheng, Y.; Zhang, L.; Nee, A.Y.C. Advanced manufacturing systems: Socialization characteristics and trends. *J. Intell. Manuf.* **2017**, *28*, 1079–1094. [\[CrossRef\]](#)
7. Peng, Y.; Huang, X.; Zhao, Y. An Overview of Cross-media Retrieval: Concepts, Methodologies, Benchmarks and Challenges. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *99*, 2372–2385. [\[CrossRef\]](#)
8. Salminen, J.; Rao, R.G.; Jung, S.G.; Chowdhury, S.A.; Jansen, B.J. Enriching Social Media Personas with Personality Traits: A Deep Learning Approach Using the Big Five Classes. In Proceedings of the International Conference on Human-Computer Interaction, Copenhagen, Denmark, 19–24 July 2020.
9. Li, H.; Mi, S.; Li, Q.; Wen, X.; Qiao, D.; Luo, G. A scheduling optimization method for maintenance, repair and operations service resources of complex products. *J. Intell. Manuf.* **2020**, *31*, 1673–1691. [\[CrossRef\]](#)
10. Guo, Q.; Xue, C.; Yu, M.; Shen, Z. A new user implicit requirements process method oriented to product design. *J. Comput. Inf. Sci. Eng.* **2019**, *19*, 11. [\[CrossRef\]](#)
11. Xie, N.; Chen, D.; Fan, Y.; Zhu, M. The acquisition method of the user's Kansei needs based on double matrix recommendation algorithm. *J. Intell. Fuzzy Syst.* **2021**, *41*, 2. [\[CrossRef\]](#)
12. Han, X.; Li, R.; Wang, J.; Qin, S.; Ding, G. Identification of key design characteristics for complex product adaptive design. *Int. J. Adv. Manuf. Technol.* **2018**, *95*, 1215–1231. [\[CrossRef\]](#)
13. Cong, Y.; Yu, S.; Chu, J.; Su, Z.; Huang, Y.; Li, F. A small sample data-driven method: User needs elicitation from online reviews in new product iteration. *Adv. Eng. Inform.* **2023**, *56*, 101953. [\[CrossRef\]](#)
14. Zhou, Q.; He, L. Research on customer satisfaction evaluation method for individualized customized products. *Int. J. Adv. Manuf. Technol.* **2019**, *104*, 3229–3238. [\[CrossRef\]](#)
15. Choi, J.; Lee, J.; Yoon, J. Anticipating promising services under technology capability for new product-service system strategies: An integrated use of patents and trademarks. *Comput. Ind.* **2021**, *133*, 103542. [\[CrossRef\]](#)
16. Liu, L.; Li, Y.; Xiong, Y.; Cavallucci, D. A new function-based patent knowledge retrieval tool for conceptual design of innovative products. *Comput. Ind.* **2020**, *115*, 103154. [\[CrossRef\]](#)
17. Jia, L.; Wu, C.; Zhu, X.; Tan, R. Design by analogy: Achieving more patentable ideas from one creative design. *Chin. J. Mech. Eng.* **2018**, *31*, 10. [\[CrossRef\]](#)
18. Bai, Y.; Chou, L.; Zhang, W. Industrial innovation characteristics and spatial differentiation of smart grid technology in China based on patent mining. *J. Energy Storage* **2021**, *43*, 103289. [\[CrossRef\]](#)
19. Kim, J.; Lee, J.; Kim, G.; Park, S.; Jang, D. A Hybrid Method of Analyzing Patents for Sustainable Technology Management in Humanoid Robot Industry. *Sustainability* **2016**, *8*, 474. [\[CrossRef\]](#)
20. Wu, Y.; Ji, Y.; Gu, F. Identifying firm-specific technology opportunities in a supply chain: Link prediction analysis in multilayer networks. *Expert. Syst. Appl.* **2023**, *213*, 119053. [\[CrossRef\]](#)

21. Pölzlbauer, G.; Auer, E. Applied patent mining with topic models and meta-data: A comprehensive case study. *World Pat. Inf.* **2021**, *67*, 102065. [\[CrossRef\]](#)
22. Kim, S.; Yoon, B. Patent infringement analysis using a text mining technique based on SAO structure. *Comput. Ind.* **2021**, *125*, 103379. [\[CrossRef\]](#)
23. Srinivasan, V.; Song, B.; Luo, J.; Subburaj, K.; Elara, M.R.; Blessing, L.; Wood, K. Does analogical distance affect performance of ideation? *J. Mech. Des.* **2018**, *140*, 71101. [\[CrossRef\]](#)
24. Zhang, Z.; Guo, J.; Zhang, H.; Zhou, L.; Wang, M. Product selection based on sentiment analysis of online reviews: An intuitionistic fuzzy TODIM method. *Complex. Intell. Syst.* **2022**, *8*, 3349–3362. [\[CrossRef\]](#)
25. Blei, D.; Ng, A.Y.; Jordan, M.L. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
26. Leng, B.; Zeng, J.; Yao, M.; Xiong, Z. 3D object retrieval with multitopic model combining relevance feedback and LDA model. *IEEE Trans. Image Process.* **2015**, *24*, 94–105. [\[CrossRef\]](#)
27. Lienou, M.; Maitre, H.; Datcu, M. Semantic annotation of satellite images using latent dirichlet allocation. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 28–32. [\[CrossRef\]](#)
28. Likas, A.; Vlassis, N.; Verbeek, J.J. The global k-means clustering algorithm. *Pattern Recogn.* **2003**, *36*, 451–461. [\[CrossRef\]](#)
29. Qi, J.; Zhang, Z.; Jeon, S.; Zhou, Y. Mining customer requirements from online reviews: A product improvement perspective. *Inf. Manag.* **2016**, *53*, 951–963. [\[CrossRef\]](#)
30. Ireland, R.; Liu, A. Application of data analytics for product design: Sentiment analysis of online product reviews. *CIRP J. Manuf. Sci. Technol.* **2018**, *23*, 128–144. [\[CrossRef\]](#)
31. Kim, H.; Noh, Y. Elicitation of design factors through big data analysis of online customer reviews for washing machines. *J. Mech. Sci. Technol.* **2019**, *33*, 2785–2795. [\[CrossRef\]](#)
32. Chen, Z.; Ming, X.; Zhang, X.; Yin, D.; Sun, Z. A rough-fuzzy DEMATEL-ANP method for evaluating sustainable value requirement of product service system. *J. Clean. Prod.* **2019**, *228*, 485–508. [\[CrossRef\]](#)
33. Ziarko, W. Variable precision rough set model. *J. Comput. Syst. Sci.* **1993**, *46*, 39–59. [\[CrossRef\]](#)
34. Lzak, D. *Rough. Sets and Bayes Factor*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 202–229.
35. Martilla, J.A.; James, J.C. Importance-Performance Analysis. *J. Mark.* **1977**, *1*, 77–79. [\[CrossRef\]](#)
36. Mikulić, J.; Prebežac, D. Accounting for dynamics in attribute-importance and for competitor performance to enhance reliability of BPNN-based importance–performance analysis. *Expert. Syst. Appl.* **2012**, *39*, 5144–5153. [\[CrossRef\]](#)
37. DiPietro, R.B.; Levitt, J.A.; Taylor, S.; Nierop, T. First-time and repeat tourists' perceptions of authentic Aruban restaurants: An importance-performance competitor analysis. *J. Destin. Mark. Manag.* **2019**, *14*, 100366. [\[CrossRef\]](#)
38. Cao, B.; Liu, X.F.; Liu, J.; Tang, M. Domain-aware Mashup service clustering based on LDA topic model from multiple data sources. *Inf. Softw. Tech.* **2017**, *90*, 40–54. [\[CrossRef\]](#)
39. Jiang, Y.; Li, M.; Dennis, A.; Liao, X.; Ampaw, E.M. The Hotspots and Trends in the Literature on Cleaner Production: A Visualized Analysis Based on Citespace. *Sustainability* **2022**, *14*, 9002. [\[CrossRef\]](#)
40. Chen, W.; Shi, X.; Fang, X.; Yu, Y.; Tong, S. Research Context and Prospect of Green Railways in China Based on Bibliometric Analysis. *Sustainability* **2023**, *15*, 5773. [\[CrossRef\]](#)
41. Blei, D.; Ng, A.; Jordan, M. Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems 14*. In *Proceedings of the Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, Vancouver, BC, Canada, 3–8 December 2001*.
42. Yang, Q. LDA-based Topic Mining Research on China's Government Data Governance Policy. *Soc. Secur. Adm. Manag.* **2022**, *3*, 2.
43. Triayudi, A.; Haerani, R. Data Mining K-Means Algorithm for Performance Analysis. *J. Phys. Conf. Ser.* **2022**, *2394*, 1. [\[CrossRef\]](#)
44. Tang, C.; Wen, T.; Liang, Z.; Xu, X.; Mou, W. Fast acquisition method using modified PCA with a sparse factor for burst DS spread-spectrum transmission. *ICT Express* **2022**, (in press). [\[CrossRef\]](#)
45. Nguyen, H.T.; Safder, U.; Kim, J.; Heo, S.; Yoo, C. An adaptive safety-risk mitigation plan at process-level for sustainable production in chemical industries: An integrated fuzzy-HAZOP-best-worst approach. *J. Clean. Prod.* **2022**, *10*, 339. [\[CrossRef\]](#)
46. Mukherjee, P.; Pattnaik, P.K.; Al-Absi, A.A.; Kang, D.-K. Recommended System for Cluster Head Selection in a Remote Sensor Cloud Environment Using the Fuzzy-Based Multi-Criteria Decision-Making Technique. *Sustainability* **2021**, *13*, 10579. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.