*Journal of Network Intelligence* ©2025 ISSN 2414-8105
**Taiwan Ubiquitous Information** Volume **10**, Number **2**, May **2025**

# Research on Product Form Innovation Design Based on Multi-label LDA Clustering

Ting Wang*
School of Architecture and Engineering, Xuzhou Vocational College of Industrial Technology, Xuzhou 221140, P. R. China
717483823@qq.com

Zhou-Xiang Zhen
Faculty of Information Science and Technology, National University of Malaysia, Selangor 43600, Malaysia
619543699@qq.com

*Corresponding author: Ting Wang

Received June 4, 2024; revised January 6, 2025; accepted April 3, 2025.

---

Abstract. *Product form innovation is an effective way for enterprises to enhance their efficiency in the competitive market environment. Existing product form innovation design methods are more dependent on designers' accumulated experience and have deviations from users' needs. Aiming at the above problems, this paper proposes a product form innovation design method based on multi-label Latent Dirichlet Allocation (LDA) clustering. Firstly, the binomial distribution is introduced into the traditional LDA clustering model to increase the discriminative ability of word items. Secondly, the optimized LDA model is used to cluster the subject words of the product evaluation text, and time constraints are added to ensure the timeliness of the subject words. Then from the polysemy of words in the product technical documents, multiple cluster-centered vectors combined with the attention mechanism are adopted to extract the design elements in different contexts, and these elements are fused and enhanced with the embedded representations of the labels for dot product to get the final design elements. Finally, the similarity between the evaluation text subject words and the product form design elements is calculated to obtain the final input variables of the product form design, and the adaptation evaluation neural network is established to output the evaluation data given by users to the design scheme. The experimental outcome implies that the clustering performance and prediction accuracy of the design method are better than the comparison model, which verifies the efficiency of the method in this article.*
**Keywords:** Product form design; Multi-label embedding; LDA clustering; Attention mechanism; Similarity computation

---

1. **Introduction.** Product form innovation has become an important method for enterprises to seek survival and development in the fierce competition in the market [1]. Innovative design of product form is a development process of applying innovative ideas and technologies to product form, aiming at creating new products with unique appearance, function and user experience. In the face of increasingly fierce market competition and diversified, personalized consumer demand, the enterprise production mode is experiencing a shift from mass production to multi-species, small batch production change [2, 3]. The emergence of personalized customization mode strengthens the user-driven attribute in product design, and industrial design is no longer mass production and meeting the general needs of the public, but shifting to user-centered and adapting to the consumer's personalized demand-oriented characteristics [4, 5]. Personalized customization mode is

an application practice of innovative design of product form, which promotes the development of product design in a more diversified and innovative direction by meeting the individual needs of consumers. Therefore, how to closely combine the material function, emotional attributes and spiritual value of products in the process of product form design, and design functional, emotional and personalized products that meet consumer expectations is a very important research topic, which is also the core issue to be explored in this research.

1.1. **Related work.** Product form, as the external expression of the product itself, reflects the emotional needs of users and influences their purchasing behavior [6]. For example, Apple's iPhone conveys a high-end, fashionable and innovative brand image with its simple and streamlined design and high-quality material selection, which is in line with the emotional needs of users who pursue quality life. Shen and Wang [7] constructed a multivariate linear product form model from individual brand imagery. Goswami et al. [8] proposed an approach to reasoning about styling features driven by the cognitive demand for style imagery to extract the key semantic features of the product's styling style. Cheng and Ma [9] used ergonomics to improve and optimize the shape and size of the product for the comfort of hand-holding. Chou [10] applied the innovation principle and matrix of TRIZ theory to solve the product styling innovation design problem and optimize the shape and color of the product, and Kauffmann et al. [11] used the method of topological semantic analysis to extract the styling features, which reduces the main design features extraction and reduces the subjectivity of design. These researches reduce the subjectivity of human factors to a certain extent, and have a positive role in promoting the operation experience of products.

However, due to the accuracy of the above methods, it is easy to cause the advantageous factors to be ignored. Nowadays, the development of artificial intelligence is deep, how to use good intelligent methods to promote the further development of product morphology design research has become one of the hot research topics. Hsiao and Huang [12] use improved gray theory and sentiment clustering to analyze the deformed product shape, and carry out the morphology optimal design prediction based on BP neural network. Tsai et al. [13] mine the connection between design variables and optimization objectives through Pearson correlation coefficient, and construct a product morphology prediction model based on integrated learning method. Quan et al. [14] combined the deep learning method with innovative design process, and used BP neural network algorithm to realize the product design style migration. Nizinski et al. [15] based on the attention mechanism of the feature recognition and extraction of the product image, and used convolutional neural network algorithm to predict the optimal design of the product. In order to reduce the time cost of product form design, Li et al. [16] used VSM to represent the user evaluation document, by calculating the similarity between the documents and finally through the k-means method for clustering, to obtain the product form requirements design document, but the user satisfaction is not high. Kashkoush and ElMaraghy [17] used graph matching method to cluster product design documents with similar structure into one class, but the degree of document similarity is low. Liu et al. [18] relied on the Fourier transform algorithm, label coding the beginning and end of the product requirements document, and used the meanshift clustering method to cluster the user's emotional factors in the document in order to innovatively design the product form. Wan et al. [19] proposed a Latent Dirichlet Allocation (LDA) clustering algorithm based on association constraints to classify the online evaluation of products, to realize the summarization and aggregation of massive product evaluation, and to improve the efficiency of design, but the effect of the clustering is poor.

1.2. **Contribution.** It can be seen from the analysis of current research that the existing product form innovation design methods cannot satisfy the user needs well. Aiming at this issue, this article designs a product form innovation design method based on multi-label LDA clustering. Firstly, the binomial distribution is introduced to optimize the traditional LDA clustering model, increase the discriminative ability of word items, and parallelize it. Then, relied on this, this article clusters the keywords of product evaluation text, extracts the theme words from the text, and adds the distribution of theme words under the time constraint. Secondly, from the polysemy of words in technical documents, this article embeds the multi-label feature representation of the documents, introduces the cluster center vector as the query vector of the attention mechanism, and uses the attention mechanism many times to assign weights to the elements in the documents, so as to realize the automatic extraction of design elements. Finally, the similarity between the evaluation text subject words and product design elements is calculated to obtain the final input variables, and an adaptation evaluation neural network is established to output the evaluation data given by users to the design scheme.

## 2. Theoretical analysis.

2.1. **Definition of Product Form Innovation Design.** Product form design is a multidisciplinary and complex innovation system integrating knowledge from multiple fields, which is accumulated step by step from product design and morphology. From the design requirements and necessary design information, a series of innovative thinking methods and effective innovation tools have been explored [20], which transform the user's uncertain perceptions of the product into the product design cognition, and map them onto the product form design through image description, which is a process of continuous improvement and refinement, so as to realize the form design oriented to the user's needs [21].

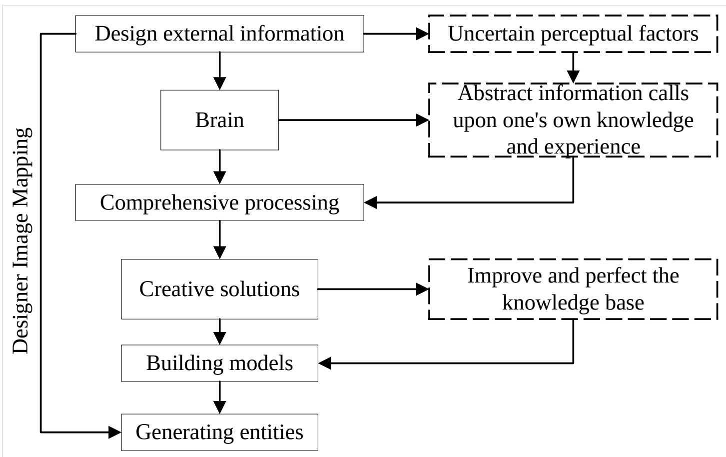


Figure 1. The basic process of product form design

The origin of a design project is the expression of design information and the availability of different knowledge bases in the design process. In innovative design, the designer's problem solving process is that the brain absorbs visual information from the outside and also transfers information from the outside world to the brain as image-graphic information, and then the brain carries out comprehensive processing, i.e., generates design ideas, and then creates a model through computer-assisted creation, and finally generates an entity [22], as indicated in Figure 1.

2.2. **LDA clustering model.** LDA is a document clustering model that associates documents and words with latent topic variables, representing each document as a mixture of topics. It has been shown to outperform traditional clustering models in many document analysis tasks [23]. LDA is capable of identifying subtopics of a technology domain consisting of many patents and representing each patent in a series of topic distributions. The LDA model is implied in Figure 2.

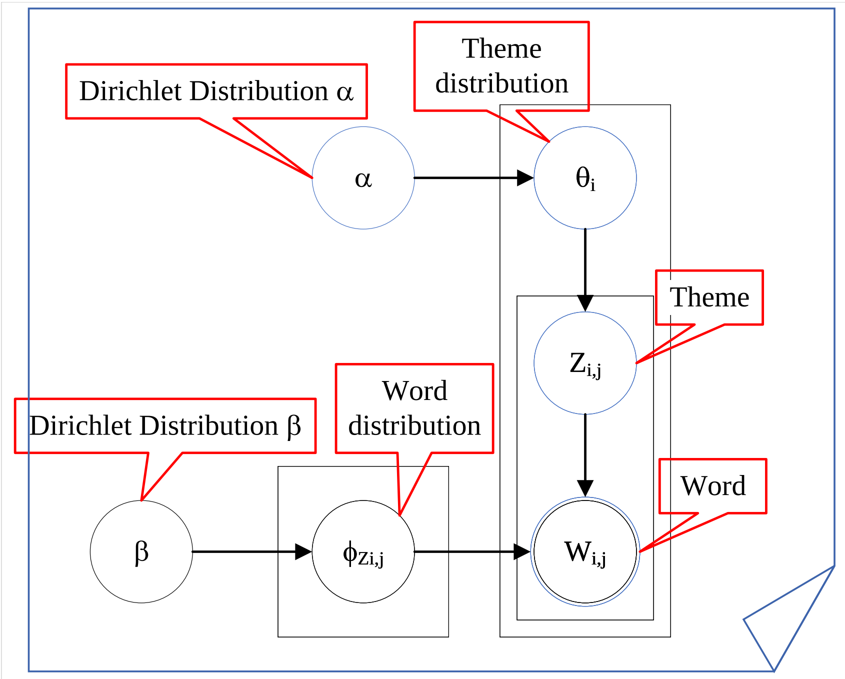


Figure 2. LDA model

In the LDA model, a document is generated as follows: firstly, the topic distribution $\vartheta_i$ of document $i$ is sampled from the Dirichlet distribution $\alpha$; secondly, the topic $y_{ij}$ of the $j$-th word of document $i$ is sampled from the multinomial distribution $\vartheta_i$ of topics; then, the distribution $\phi_{y_{ij}}$ of the word corresponding to topic $y_{ij}$ is sampled from the Dirichlet distribution $\beta$; and lastly, the word $V_{ij}$ is sampled from the multinomial distribution of words $\phi_{y_{ij}}$. The joint probability distribution of occurrence of each word in the generated document is as follows.

$$P(v_j \mid y_l) = \sum_{l=1}^{L} P(w_j \mid z_l)P(z_l \mid d_i) \tag{1}$$

By solving the above model through probabilistic derivation, the main structure of the document is obtained, and $\vartheta_i$ and $y_{ij}$ are derived, similar to the initial paper of the PLSA and LDA models, and the unknown parameters are estimated by the EM algorithm. The method of estimating the unknown parameters of LDA is Gibbs sampling [24], which is an algorithm used to obtain a series of samples of observations approximately equal to a specified multidimensional probability distribution.

The distribution of the word $v_i$ over the topic $y$ is as follows.

$$\delta_{y,v_i} = \frac{n_{v_i,y} + \beta}{\sum_{i=1}^{L} n_{v_i,y} + \beta} \tag{2}$$

The distribution of document $d$ over subject $y$ is as follows.

$$\vartheta_{d,y} = \frac{n_d^z + \alpha}{\sum_{y=1}^{L} n_d^y + \alpha} \tag{3}$$

The Gibbs sampling formula is derived after a series of training derivation process.

$$p\left(y_i = l | \overrightarrow{y}, \overrightarrow{v}\right) \propto \frac{n_{d,i}^z + \alpha_l}{\sum\limits_{y=1}^{L} n_{d,i}^y + \alpha_l \sum\limits_{i=1}^{V} n_{v,y,i} + \beta_t} \tag{4}$$

3. **LDA clustering model improvement.** The basic LDA clustering model does not take into account the degree of differentiation of lexical items, the text is sparse, and the direct removal of lexical items that do not have a degree of differentiation of the topic will have a certain impact on the model, so we propose a new model that introduces the binomial distribution into the basic LDA model to increase the ability to discriminate between the lexical items.

The words are first sampled, and the same word belonging to different topics is treated as independent of each other, and the discriminative power of the lexical items is denoted by $\gamma$.

If topic $y$ contains lexical item $v$, then $\gamma_{v,y} = 1$; if it does not, then $\gamma_{v,y} = 0$. It can be seen that $\gamma$ is a parameter obeying a binomial distribution: $\gamma \sim B(L, \mu_v)$. Where $\mu_v$ represents the distribution of word $v$ over all topics. This article assumes $\mu_v \sim Beta(a, b)$. Because the parameters $\alpha$ and $\beta$ have been used, the parameters $a$ and $b$ are used in this paper to denote the parameters in the beta distribution.

$$p(\mu_v \mid a, b) = \frac{1}{B(a, b)} \mu_v^{a-1} (1 - \mu_v)^{b-1} \tag{5}$$

Assuming that the number of times a term $v$ is counted in $L$ topics is $n_v^{(1)}$ and the number of times it does not occur is $n_v^{(0)} = L - n_v^{(1)}$, the conditional probability that the term $v$ occurs in the next topic is $p(\gamma = 1 \mid v)$, which can be derived from the Bayesian [25] formula.

$$p(\mu_v \mid \Upsilon, a, b) = Beta\left(\mu_v \mid a + \sum_{y=1}^{L} \Upsilon_{v,y} = 1, b + \sum_{y=1}^{L} \Upsilon_{v,y} = 0\right) \tag{6}$$

where $\Upsilon = \{\Upsilon_{v,1}, \Upsilon_{v,2}, ..., \Upsilon_{v,L}\}$.

$$\mu_v = p\left(\gamma_{v,y} = 1 \mid \gamma\right) = \frac{\sum_{y=1}^{L} \gamma_{v,y} = 1 + a}{L + a + b} \tag{7}$$

If an item $v$ occurs uniformly across topics, a larger value of $\mu_v$ does not discriminate much between topics; on the contrary, if an item $v$ is concentrated in one or a small number of topics, the value of $\mu_v$ for this item is smaller. Therefore, $v = 1/\mu_v$ is used to represent the discriminative power of a word item, and more weight is assigned to words with high discriminative power to minimize the influence of words with low discriminative power on the clustering. The regularization of $v$ using $v \in [0, 1]$ is expressed as follows.

$$v = \frac{(L - n_v^{(1)})(a + 1)}{(L - 1)(a + n_v^{(1)})} \tag{8}$$

After introducing $\mu_v$, the improved LDA model first Gibbs samples the topic distribution: $\vartheta_d \sim Dir(\alpha)$; then selects the document topic: $y_d \sim Multi(\vartheta_d)$; and finally samples a word: $v \sim p(v \mid y_d, \beta)$, and $v$ obeys $v \sim p(v \mid a, b)$. The model is then used as an example of the LDA.

After a series of derivation transformations, the Gibbs sampling of the new model can be obtained the updated Equation.

$$p(y_i = l \mid \vec{y}_{-i}, \vec{v}) \propto \frac{(L - n_{v_i}^{(1)})(a+1)}{(L-1)(a + n_v^{(1)})} \cdot \frac{n_{d,?}^{y} + \alpha}{\sum_{y=1}^{L} n_{d,?}^{y} + L\alpha} \cdot \frac{n_{v,z,?}^{i} + \beta_t}{\sum_{i=1}^{V} n_{v,y,?}^{i} + \beta_t} \tag{9}$$

At the initial moment, a random assignment of topic $y_i$ is performed for each word item of the document. Then, the number of words $v$ under $y_i$ and the number of words in document $i$ containing the words in $y_i$ are counted. Each round estimates the probability that each word is assigned to each topic according to equation $p(y_i = l \mid \vec{y}_{-i}, \vec{v})$, and then samples new topics using this probability distribution. If the distribution of the words in each topic is updated and converges, the parameters $\phi$ and $\vartheta$ of the model are output.

$$\vartheta_{d,y} = \frac{n_{y,d} + \alpha}{\sum_{Y=1}^{L} n_{y,d} + L\alpha} \tag{10}$$

$$\phi_{v,v_i} = \frac{n_{v_i} + \beta}{\sum_{i=1}^{V} n_{v_i,y} + V\beta} \cdot \frac{(L - n_{v_i}^{(1)})(a+1)}{(L-1)(a + n_v^{(1)})} \tag{11}$$

## 4. Research on product form innovation design based on multi-label LDA clustering.

**4.1. User requirement mining based on LDA clustering.** Intending to the issue that users are not satisfied with the current product design methods, this paper first uses the optimized LDA model to cluster the subject words of the product evaluation text, and then uses multiple cluster centroid vectors combined with the attention mechanism to extract the product design elements in different contexts, and then fuses and enhances these elements with the embedded representations of the labels and performs dot product to obtain the final design elements. Finally, the similarity between the subject words and design elements is calculated to obtain the final input variables, and an adaptive evaluation neural network is established to output the evaluation data given by the users to the design solutions to provide designers with design references. The flow of the suggested method is implied in Figure 3

Every valid product review contains a theme that not only describes the user's attitude or feelings about the product, but also indirectly describes the product's performance and features. If the same theme appears in different evaluation texts, it means that the theme is the most popular part of the product. In this paper, we use the above optimized LDA model to cluster the keywords of the evaluation texts and extract the theme words from the texts. The specific steps are as bellow.

(1) For topic $y_l (l = 1, 2, \ldots, L)$, generate the multinomial distribution parameter $\phi_l \sim Dir(\beta)$ as the lexical distribution $p(v \mid y_l)$ for topic $y_l$.

(2) For text $W_m (m = 1, 2, \ldots, M)$, generate the multinomial distribution parameter $\vartheta_m \sim Dir(\beta)$ as the lexical distribution $p(v \mid y_l)$ of the topic $y_l$.

(3) For vocabulary $\overline{\omega}_{mn} (m = 1, 2, \ldots, M; n = 1, 2, \ldots, N_m)$ of $W_m$, topic $y_{mn} \sim Mult(\vartheta_m)$ is generated first, and then vocabulary $\overline{\omega}_{mn} \sim Mult(\phi_{y_{mn}})$ is generated.

where $L$ denotes the number of given topics; $M$ denotes the number of texts; and $\alpha$ and $\beta$ are both given hyperparameters of the Dirichlet distribution.

Because the evaluation of products is time-sensitive, this paper adds time constraints on the basis of the improved LDA model, and before calculating the distribution probability that a word $v$ belongs to a topic $y$, it first determines whether it belongs to a certain point in time $t_i$, and then decides whether or not to carry out the topic-weighted distribution. The topic-word distribution under time constraint $\phi^t_{v,v_i}$.

$$\phi^t_{v,v_i} = \frac{n_{v_i} + \beta}{\sum_{i=1}^{V} n_{v_i,y} + V\beta} \cdot \frac{(L - n^{(1)}_{v_i})(a+1)}{(L-1)(a + n^{(1)}_{v})} q^t(v) \tag{12}$$

where $T$ is the number of time points, $L$ is the number of topics, and $n_{v_i}$ is the total number of words in the document corpus. Where the time constraint factor $q^t(v)$ is calculated as follows.

In the current topic $y_l$, the number of words with the same timestamp as $v$ is $n_1$, and the number of words with different timestamps is $n_2$. The weight $q^t(v)$ of a word $v$ belonging to topic $y_l$ is calculated as implied in Equation (4), where $0.5 < \mu < 1$.

$$q^t(v) = \mu n_1 - (1 - \mu) n_2 \tag{13}$$

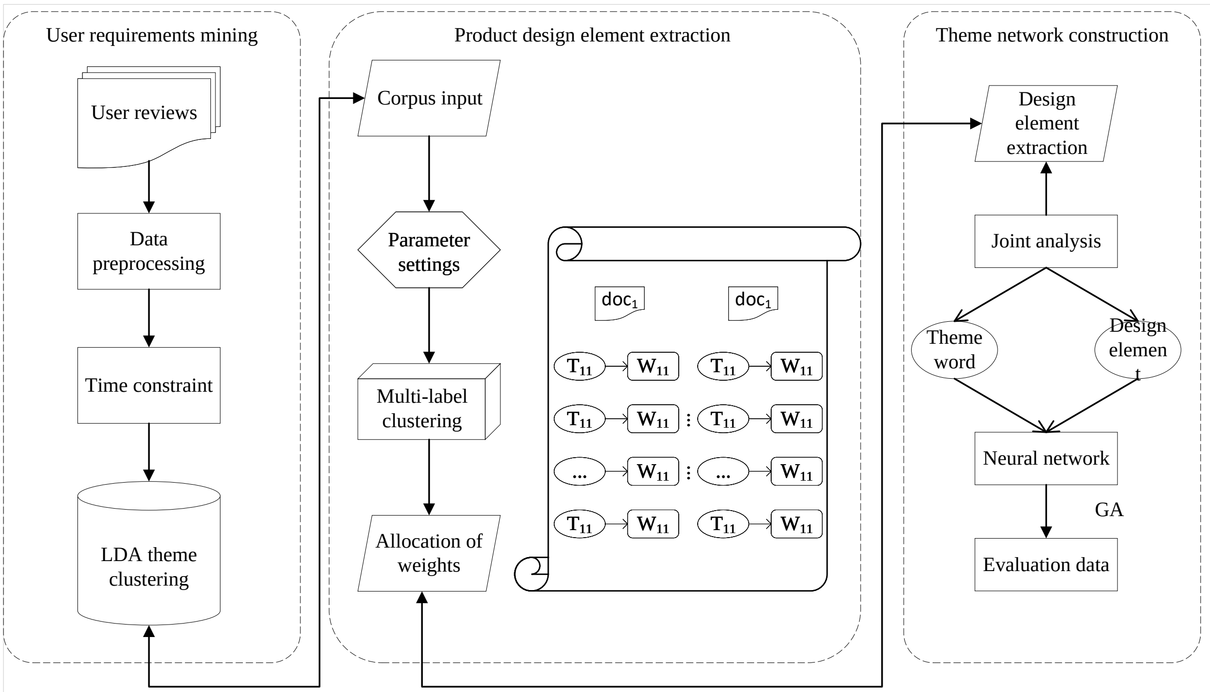


Figure 3. The overall flow of the suggested method

4.2. **Extraction of product design elements based on multi-label clustering.** After obtaining the subject terms of the product evaluation document, this paper starts from the polysemy of the words in the technical documents, introduces the cluster center vector $\{c_1, c_2, ..., c_l\}$ as the query vector of the attention mechanism, and uses the attention mechanism many times to assign weights to the words in the documents, and update the value of the cluster center vectors in the back propagation, so as to realize automatic extraction of the contexts.

The number of cluster center vectors is a hyperparameter in the improved LDA, and the initial values of the vectors are obtained by random initialization. After the integration of the attention mechanism, the features of the document are reorganized into several new elements $\{s_1, s_2, ..., s_l\}$ focusing on different contexts, which are formulated as bellow.

$$\mu(c_l, g_n) = W \tanh\left(V_1 c_k + V_2 g_n\right) \tag{14}$$

$$\eta_{l,n} = \frac{\exp[\mu(c_l, g_n)]}{\sum_{n'=1}^{l} \exp[\mu(c_l, g_{n'})]} \tag{15}$$

$$s_l = \sum_n \eta_{l,n} g_n \tag{16}$$

where $W$, $V_1$ and $V_2$ are the weight parameters in the attention mechanism, which are automatically updated by backpropagation, and $\eta_{l,n}$ is the weight of the implicit vectors of the words in the document on the cluster center vectors, which is multiplied by the implicit vectors of the words in the document, $g_n$, and the weight, $\eta_{l,n}$, and then summed up to get the new feature of each word on the cluster center. By setting the number of cluster centers, we can get the different elements of a document in each context, and combine these elements together to get the combination element $s_l$.

To facilitate the interaction with the embedded representation of the label, the self-attention mechanism is used to transform the combination element $s_l$ into the contextual element of the product technical document, and the self-attention is implemented as bellow.

$$\mu(s_l) = W \tanh(V s_l + b) \tag{17}$$

$$\eta_l = \frac{\exp[\mu(s_l)]}{\sum_{l'=1} \exp[\mu(s_{l'})]} \tag{18}$$

$$s'_l = \sum_l \eta_l s_l \tag{19}$$

where $W, V$, and $b$ are the parameters of self-attention, which are automatically updated by backpropagation, and $\eta_l$ is the weight of the feature, multiplying the group element $s_l$ and its corresponding weight $\eta_l$ obtained in Equation (16) and then summing them up, we get the element $s'_l$ that represents the document in different contexts.

All tags $\{k_1, k_2, ..., k_m\}$ are passed through the tag embedding layer to obtain the embedded representation features $K$ of the product technical document tags, and the final document tag elements $S$ are obtained by dot-multiplying the feature matrices of the documents and tags, where $\delta = 1/(1 + e^{-x})$.

$$S = \delta\,(s'_l \cdot K) \tag{20}$$

4.3. **Innovative design of product form based on multi-label LDA clustering.** After obtaining the product evaluation text subject words and morphological design elements, it is necessary to calculate their similarity in order to obtain the final input variables, establish an adaptation evaluation neural network, and output the evaluation data given by the users to the design scheme, so as to provide design references for industrial designers.

To mine the semantic similarity between the subject word $y$ and the design element $S$, they are firstly processed by partitioning to form a collection of individual words. Assuming that the set of subject words is $A$ and the set of design elements is $B$, each word element in $A$ and each word element in $B$ are calculated for similarity, and then the average of the maximum value is taken as the similarity between the set of $A$ and the set of $B$ respectively.

$$Sim(A, B) = \frac{\sum_{y \in A \subseteq B}[MaxSim(a, b)] + \sum_{y \in B \subseteq A}[MaxSim(b, a)]}{|A| + |B|} \tag{21}$$

$A$ is then converted to a similarity value of 0 to 1 using the edit distance.

$$\lambda = 1 - \frac{Dist}{Max(length_A, length_B)} \tag{22}$$

where $Dist$ is the minimum edit distance between $A$ and $B$, and $length_A$ and $length_B$ are the lengths of the words to be computed, respectively.

Therefore, if the similarity between the subject term and the design element is between $[0, 1]$, the design element is used as the input variable $x_i$ of the neural network, and the user's evaluation score is used as the output data of the neural network, and the trained simulated neural network is used as the fitness function of the genetic algorithm.

A small number of individuals from the initial population are selected as the test sample set $C_{test}$, and the rest of the individuals are selected as the training sample set $C_{train}$.

$$\begin{cases} C_{train} = \{(x_i, f(x_i))\}, i = 1, 2, \ldots, n \\ C_{test} = \{(\tilde{x}_j, f(\tilde{x}_j))\}, j = 1, 2, \ldots, m \end{cases} \tag{23}$$

where $f(x_i)$ is the fitness value of $x_i$, $f(\tilde{x}_j)$ is the fitness value of the test individual $\tilde{x}_j$. $n$ and $m$ denote the total number of samples in the training and test sets, respectively.

The final evaluation value of the neural network output represents the fitness value of the offspring individual, and the larger the fitness value of the offspring individual, the more the individual meets the user's needs. The neural network generates an approximation model $\hat{f}(x_i)$ based on the learning of the training sample set. Individual $\tilde{x}_j$ in $C_{test}$ is inputted into the neural network, and the estimated individual fitness value $\hat{f}(\tilde{x}_j)$ is obtained, then the test error of the neural network is as follows.

$$\xi = \frac{1}{m} \sqrt{\sum_{j=1}^{m} \left( f(\tilde{x}_j) - \hat{f}(\tilde{x}_j) \right)^2} \tag{24}$$

Test error $\xi$ is an evaluation metric of the learning effectiveness of the neural network, which determines whether $\hat{f}(x_i)$ can play a role in the evaluation.

Finally, according to the idea of genetic algorithm [26], the basic genetic operators such as selection, crossover and mutation are utilized to carry out the iterative evolutionary design of the product form, which provides a design reference for industrial designers.

## 5. **Performance testing and analysis.**

### 5.1. **Clustering performance analysis.**

In this article, the home product design dataset constructed in the literature [27] is used as an experimental object to evaluate the product form innovation design method and the dataset contains 1726 product evaluation texts and design texts. Before training, all the inputs and outputs are standardized, the maximum length of the training input sequence is 256, the batch size of the model training is 24, and the learning rate of the model is $3 \times 10^{-5}$. The toolbox in python is used to build the neural network with the evaluation data of the product design scheme as the output. Then train the model. The trained optimal program is obtained. The hardware parameters used are windows 10 64-bit operating system, Intel (R) Core (TM) i7-9750H CPU @ 2.60GHz, 16.0 GB of onboard memory (15.8 GB available), Python 3.8 programming language, and PyCharm development tool.

In this experiment, Accuracy, Precision, Recall, Pairwise F measure value (PWF) [28] and Normalized Mutual Information (NMI) [29] are used as evaluation metrics to assess the clustering performance of different product form innovation design methods. For the convenience of analysis, the method in this article is denoted as OURS, the method in

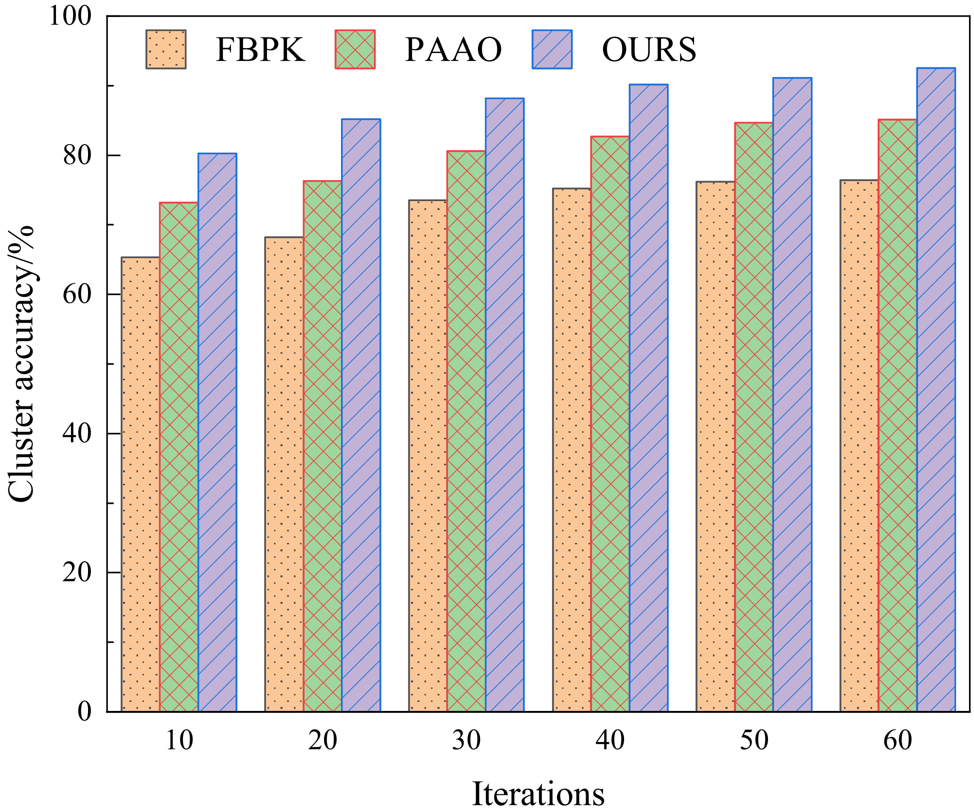


Figure 4. The comparison of the clustering accuracy of the various design methods

the literature [18] is denoted as FBPK, and the method in the literature [19] is denoted as PAAO. From Figure 4, it can be seen that the clustering accuracy of all algorithms is improving with the increase of the number of iterations, and the clustering accuracy of the OURS is higher than that of the FBPK and PAAO. When the number of iterations is 30, the accuracy of the OURS is 88.19%, the accuracy of the FBPK is 73.52%, and the accuracy of the PAAO is 80.58%. The OURS improves 14.67% and 7.61% compared to FBPK and PAAO, respectively. The OURS is superior because it does not require dimensionality reduction, clusters the subject words through the optimized LDA model, ensures the timeliness and efficiency of clustering.

Table 1 shows the comparison of the clustering effect of different product form design methods. The PWF and NMI of OURS are 0.9067 and 0.8197, which are improved by 15.72% and 12.83% compared to FBPK and 6.78% and 7.14% compared to PAAO, respectively, and OURS method is better than FBPK and PAAO in terms of clustering effect. This is because FBPK only utilizes the meanshift clustering method to cluster the user sentiment factors in the product documents without further clustering the topic words in the documents and without feature extraction for the product design documents, which leads to poor clustering results. Although PAAO uses the LDA model to cluster the subject words in the evaluation documents, it does not optimize the traditional LDA model and does not consider the multi-label semantic information, which results in inaccurate semantic information. The OURS method not only optimizes the LDA model but also fully considers the multi-label semantic information of the product design documents, and therefore the clustering effect performs the best.

5.2. **Prediction accuracy analysis.** To evaluate the prediction accuracy of product morphology design methods, this paper uses MSE, MAE, RMSE, MAPE and coefficient

Table 1. Comparison of clustering effects of different design methods

| Method | Precision | Recall | PWF | NMI |
|---|---|---|---|---|
| FBPK | 0.7524 | 0.7466 | 0.7495 | 0.6914 |
| PAAO | 0.8326 | 0.8452 | 0.8389 | 0.7483 |
| OURS | 0.8931 | 0.9208 | 0.9067 | 0.8197 |

of determination $R^2$ [30] to compare and analyze the prediction performance of different methods. Figure 5 shows the comparison of MSE of different methods with epoch growth. It can be clearly seen that there is no obvious fluctuation in the training mean square error and test mean square error of the OURS method, and the trend is stable and small. In 1000 iterations, most of the values of the training mean square error of the OURS are between 0.0001 and 0.0035, which indicates that the trained model has a high degree of fit and can accurately reflect the mapping relationship between the product form and the user evaluation text.

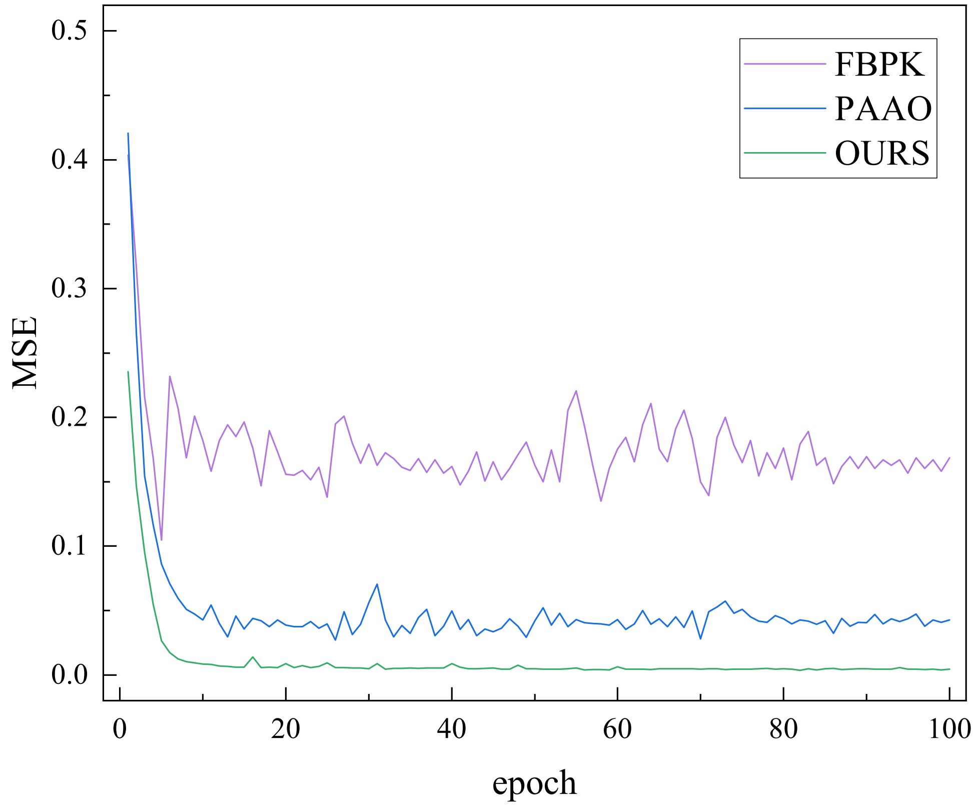


Figure 5. MSE experimental results of various methods

Comparison of the prediction performance indexes of different methods is implied in Table 2. The accuracy evaluation indexes of the OURS method are better than those of FBPK and PAAO. The MAPE of the OURS method is 0.0163, which is 0.0818 and 0.1769 lower than that of FBPK and PAAO, respectively, which indicates that the OURS method is more suitable for predicting the evaluation data of the product design scheme. Comparing the coefficient of determination $R^2$, the $R^2$ value of the OURS method is 0.9471, which is 20.11% and 8.59% higher compared to FBPK and PAAO, respectively, which indicates that the ORS method utilizes the optimized LDA model for keyword clustering and extracts the design elements based on multi-labels, and fuses the two as inputs to the neural network, which has a better fitting effect on the evaluation data prediction of product design form.

Table 2. Comparison of predictive performance metrics of different methods

| Method | MAE | RMSE | MAPE | $R^2$ |
|---|---|---|---|---|
| FBPK | 0.1729 | 0.2066 | 0.1932 | 0.7885 |
| PAAO | 0.0968 | 0.1492 | 0.0981 | 0.8722 |
| OURS | 0.0392 | 0.0517 | 0.0163 | 0.9471 |

6. **Conclusion.** Existing product form innovation design methods cannot well realize the matching between designers' and users' needs, resulting in low user satisfaction. Aiming at this problem, this paper designs a product form innovation design method based on multi-label LDA clustering. Firstly, binomial distribution is introduced to optimize the traditional LDA clustering model and increase the discriminative ability of word items. Then on this basis, we cluster the keywords of the product evaluation text, extract the theme words from the text, and add the distribution of the theme words under the time constraint. Secondly, we embed the multi-label feature representation of the document and introduce the cluster center vector as the query vector of the attention mechanism to realize the automatic extraction of design elements. Finally, the similarity between the evaluation text subject words and the product form design elements is calculated to obtain the final input variables, and an adaptive evaluation neural network is established to output the evaluation data given by users to the design scheme. The experimental results show that compared with the existing methods, the method proposed in this paper has lower MSE, MAE, RMSE, MAPE, and can efficiently realize the accurate prediction of user evaluation data.

**Acknowledgment.** This work is supported by the Key projects of Xuzhou Vocational College of Industrial Technology (No. XGY2023ZXZD03).

## REFERENCES

[1] M. Berkowitz, "Product shape as a design innovation strategy," *Journal of Product Innovation Management*, vol. 4, no. 4, pp. 274–283, 1987.
[2] F. Ricotta and M. Costabile, "Customizing customization: A conceptual framework for interactive personalization," *Journal of Interactive Marketing*, vol. 21, no. 2, pp. 6–25, 2007.
[3] P. Zheng, X. Xu, S. Yu, and C. Liu, "Personalized product configuration framework in an adaptable open architecture product platform," *Journal of Manufacturing Systems*, vol. 43, pp. 422–435, 2017.
[4] B. Montreuil and M. Poulin, "Demand and supply network design scope for personalized manufacturing," *Production Planning & Control*, vol. 16, no. 5, pp. 454–469, 2005.
[5] S.-H. Woo and S.-D. Kwon, "A study on personalized product demand manufactured by smart factory," *Management & Information Systems Review*, vol. 38, no. 1, pp. 23–41, 2019.
[6] M. Hassenzahl, S. Diefenbach, and A. Göritz, "Needs, affect, and interactive products–Facets of user experience," *Interacting with Computers*, vol. 22, no. 5, pp. 353–362, 2010.
[7] H.-C. Shen and K.-C. Wang, "Affective product form design using fuzzy Kansei engineering and creativity," *Journal of Ambient Intelligence and Humanized Computing*, vol. 7, no. 6, pp. 875–888, 2016.
[8] M. Goswami, Y. Daultani, and M. Tiwari, "An integrated framework for product line design for modular products: Product attribute and functionality-driven perspective," *International Journal of Production Research*, vol. 55, no. 13, pp. 3862–3885, 2017.
[9] J. Liu, Z. Cheng, and Y. Ma, "Product design-optimization integration via associative optimization feature modeling," *Advanced Engineering Informatics*, vol. 30, no. 4, pp. 713–727, 2016.
[10] J.-R. Chou, "A TRIZ-based product-service design approach for developing innovative products," *Computers & Industrial Engineering*, vol. 161, Art. 107608, 2021.

[11] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining," *Sustainability*, vol. 11, no. 15, Art. 4235, 2019.

[12] S.-W. Hsiao and H.-C. Huang, "A neural network based approach for product form design," *Design Studies*, vol. 23, no. 1, pp. 67–84, 2002.

[13] H.-C. Tsai, S.-W. Hsiao, and F.-K. Hung, "An image evaluation approach for parameter-based product form and color design," *Computer-Aided Design*, vol. 38, no. 2, pp. 157–171, 2006.

[14] H. Quan, S. Li, and J. Hu, "Product innovation design based on deep learning and Kansei engineering," *Applied Sciences*, vol. 8, no. 12, Art. 2397, 2018.

[15] C. A. Nizinski, C. Ly, C. Vachet, A. Hagen, T. Tasdizen, and L. W. McDonald IV, "Characterization of uncertainties and model generalizability for convolutional neural network predictions of uranium ore concentrate morphology," *Chemometrics and Intelligent Laboratory Systems*, vol. 225, Art. 104556, 2022.

[16] P. Li, Y. Ren, Y. Yan, and G. Wang, "Intelligent product-gene acquisition method based on K-means clustering and mutual information-based feature selection algorithm," *AI EDAM*, vol. 33, no. 4, pp. 469–483, 2019.

[17] M. Kashkoush and H. ElMaraghy, "Product design retrieval by matching bills of materials," *Journal of Mechanical Design*, vol. 136, no. 1, Art. 11002, 2014.

[18] L. Liu, Y. Li, Y. Xiong, and D. Cavallucci, "A new function-based patent knowledge retrieval tool for conceptual design of innovative products," *Computers in Industry*, vol. 115, Art. 103154, 2020.

[19] C. Wan, Y. Peng, K. Xiao, X. Liu, T. Jiang, and D. Liu, "An association-constrained LDA model for joint extraction of product aspects and opinions," *Information Sciences*, vol. 519, pp. 243–259, 2020.

[20] S.-W. Hsiao and H.-P. Wang, "Applying the semantic transformation method to product form design," *Design Studies*, vol. 19, no. 3, pp. 309–330, 1998.

[21] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19–46, 2024.

[22] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10, Art. 2339, 2023.

[23] T.-Y. Wu, H. Li, and S.-C. Chu, "CPPE: An Improved Phasmatodea Population Evolution Algorithm with Chaotic Maps," *Mathematics*, vol. 11, no. 9, Art. 1977, 2023.

[24] A. E. Gelfand, "Gibbs sampling," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1300–1304, 2000.

[25] A. Gelman and C. R. Shalizi, "Philosophy and the practice of Bayesian statistics," *British Journal of Mathematical and Statistical Psychology*, vol. 66, no. 1, pp. 8–38, 2013.

[26] M. T. Bhoskar, M. O. K. Kulkarni, M. N. K. Kulkarni, M. S. L. Patekar, G. Kakandikar, and V. Nandedkar, "Genetic algorithm and its applications to mechanical engineering: A review," *Materials Today: Proceedings*, vol. 2, no. 4–5, pp. 2624–2630, 2015.

[27] S. Bell and K. Bala, "Learning visual similarity for product design with convolutional neural networks," *ACM Transactions on Graphics*, vol. 34, no. 4, Art. 1, 2015.

[28] A. Alferaidi, K. Yadav, S. Yasmeen, Y. Alharbi, W. Viriyasitavat, G. Dhiman, and A. Kaur, "Node multi-attribute network community healthcare detection based on graphical matrix factorization," *Journal of Circuits, Systems and Computers*, vol. 33, no. 05, Art. 2450080, 2024.

[29] A. Amelio and C. Pizzuti, "Correction for closeness: Adjusting normalized mutual information measure for clustering comparison," *Computational Intelligence*, vol. 33, no. 3, pp. 579–601, 2017.

[30] P. Zou, J. Yang, J. Fu, G. Liu, and D. Li, "Artificial neural network and time series models for predicting soil salt and water content," *Agricultural Water Management*, vol. 97, no. 12, pp. 2009–2019, 2010.