

## Research Article

# Research on Patent-Knowledge Representation and Automatic Classification Based on Situation Mapping

Zilin Xu, Wenqiang Li , Yan Li, Jinlong Ma, and Qiyu Liu

*School of Mechanical Engineering, Sichuan University, Chengdu 610065, China*

Correspondence should be addressed to Wenqiang Li; [liwenqiang@scu.edu.cn](mailto:liwenqiang@scu.edu.cn)

Received 7 October 2019; Revised 19 January 2020; Accepted 11 May 2020; Published 27 July 2020

Academic Editor: L. J. García Villalba

Copyright © 2020 Zilin Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A patent is a type of long-term literature containing the most complete design information in most fields. Thus, it can provide designers with valuable guides for solving various design problems in different fields. Establishing the mapping relationship between the design problem and patent knowledge in different fields is the key to creating an incentive channel for the transfer and combination of multidisciplinary patent knowledge. According to the situation mapping between the design problem and patent knowledge, a method for patent-knowledge representation and automatic patent classification is proposed in this paper. The problem situation is described using four dimensions, namely, function, performance, relationship, and emotion, according to the design-problem types. Multigranular situation attributes are extracted from different design problems. A structured attribute database is established. The mapping relationship between the problem situation and patent knowledge is developed. To realize the effective utilization of the patent knowledge, this study investigates an automatic patent-classification method using the situation attributes as classification categories. An application system of patent knowledge is developed by this method. The system can support the search for patent knowledge related to the design problem and effectively assist designers in achieving an innovative design process.

## 1. Introduction

According to a study by the World Intellectual Property Organization, 90%–95% of inventions in the world are reported by patents, and 80% of them are not recorded in any other texts. Effective use of patent resources can shorten product development time by 60% and save 40% of the cost incurred on research and development [1]. Patent knowledge is not only a carrier for innovative achievement but also a significant resource to expand the knowledge space and promote the level of inventions. Therefore, it is vital to effectively exploit the tacit knowledge in patent literature in order to provide designers with more valuable design information for the innovative design of computer-aided products.

One problem is identifying the motive force of knowledge discovery, and it always starts with the representation of the problem situation. Constrained by limited short-term processing capacity and the focus on current complex tasks,

the human brain finds it difficult to shift from one mode of thinking to another [2–4]. Verifying every possible process for cognitive-information processing can be overwhelming, and the way in which people usually solve problems depends on analogical thinking [5]. Whether they solve familiar or new problems, people always start finding similarities between the current and previous problems and then activate the knowledge associated with them. The analogical-thinking mechanism of the brain represents a more efficient way of life, thereby enabling humans to almost always find effective solutions to problems rather than scanning the memory by trial and error. With regard to the exertion of analogical mechanism in the problem-solving process, current researchers believe that analogical problem-solving requires problem solvers to see some similarities between the current and existing problems in the memory and identify whether they are surfaces, features, semantics, relationships, structures, and so on [6]. The essence of analogical thinking is the transfer of design knowledge from one situation to

another using a mapping process. It aims to find a set of one-to-one correspondences (often incomplete) between one body of knowledge information and the aspects of another [7–10]. To accomplish the situation transfer between problems and knowledge, their mapping should be established, and relevant knowledge resources should be organized systematically.

According to user requirements, this paper holds that the design problem is divided into four dimensions: function, performance, relationship, and emotion. The design problem from functional dimension reflects user's description of product's functional requirements; the design problem from performance dimension reflects user's description of product's technical performance for functional requirements; the design problem from relationship dimension reflects user's description of the relationship between components of the product; and the design problem from emotional dimension reflects user's description of the product's appearance or manifestation. To accurately describe the main design problems, extracting different situation attributes from the design problem is necessary, and these attributes are helpful in clearly defining the design problems. Noh [11] and Cong [12] extracted the inventive principle of a design problem as a situation attribute in order to mine the technical information of patent knowledge and establish a unified representation model of a patent. Li [13] extracted the functional information of a product as a situation attribute of the design problem and realized effective mining of tacit knowledge in patent knowledge. Trappey [14, 15], Yu [16], and Lee [17] considered the trend in product technology evolution as situation attributes of design problems to forecast the direction of future technological improvement. Chen [18] considered scientific effects as situation attributes of design problems and deeply analyzed the working principle of technology in a patent. Cardillo [19] and Li [20] selected engineering parameters as situation attributes of design problems to analyze and solve technical conflicts in systems. However, the attributes of a complex system usually contain function, performance, relationship, and emotion. Integrity is lacking in establishing the mapping relationship between the design problems and patent knowledge from a single situation attribute. This limitation cannot support the systematic organization of patent knowledge from different aspects of a product design according to the need of the designers.

To support the complex mapping process between the design problems and patent knowledge, systematically organizing the patent knowledge according to different situation attributes is necessary. However, the workload of manual reading, labeling, and classifying patents according to the situation attributes is heavy, which greatly reduces the efficient utilization of patent knowledge. As a main tool for patent-text processing and automatic patent classification, computer natural-language-processing technology is an effective alternative to manual work to improve the utilization of patent knowledge. Ghareb [21], Labani [22], and Chen [23] proposed several methods for feature selection of patent texts, which could effectively support the attribute extraction of patent texts. Zhu [24] proposed an automatic

requirement-oriented patent-classification method to better meet various patent-management requirements. Wu [25] proposed an automatic classification method based on self-organizing maps and support vector machine (SVM), which can help in effectively analyzing the quality of a patent. Lai [26] proposed a new approach based on cocitation analysis of bibliometrics to assess the similarity of patents to support the establishment of a classification system. Liu [27] proposed a hybrid patent-classification method to analyze query patents and effectively predict their classes. Chen [28] proposed a novel three-phase categorization method that could classify patents down to the subgroup level with reasonable accuracy. However, most of the aforementioned studies were limited to the analysis of patent text or effective extraction of key words to improve the accuracy of automatic patent classification. On the other hand, there were only a few studies on how to improve the accuracy of automatic patent classification from multiple dimensions of the whole system, including the function, performance, relationship, and emotion.

A method for patent-knowledge representation and automatic patent classification based on situation mapping is proposed in this paper. According to the types of design problems, we described the problem situation in terms of four dimensions, namely, function, performance, relationship, and emotion, and extracted the granularity attributes from the different situation dimensions. We established a structured attribute database and a mapping relationship between the problem situation and patent knowledge based on the semantic similarity of the situation attributes. In addition, an automatic classification method of patent information based on situation attributes was proposed, and an experiment on automatic patent classification using situation attributes as categories was carried out using a computer. The established classifier in the experiment was used to classify a large number of unknown patent texts when the classification results satisfied the general use. The application system of the patent knowledge was developed by the proposed method, which realized effective use of patent knowledge in solving innovative design problems.

## 2. Mapping Process between Problem Situation and Patent Knowledge

Because of the diversity and complexity of the problems in different design phases and different fields, extracting the problem situation attributes from different dimensions is necessary to precisely establish the mapping relationship between the problem situation and patent knowledge. In addition, the process of creatively solving a problem involves concretization of the problem situation, which leads to the required patent knowledge that belongs to different abstract granularities in different design phases. Therefore, extracting the different granular attributes according to the abstraction level of the problem situation is significant in order to establish a complete mapping relationship in the solution process. Next, we present the mapping process between the problem situation and patent knowledge based on multi-dimensional and multigranular situation attributes.

*2.1. Mapping Process between the Problem Situation and Patent Knowledge Based on Multidimensional Attributes.* According to the user's requirements of design [29], we select function, performance, relationship, and emotion as the attributes of the problem situation. The attribute of the function reflects the ultimate purpose of the product design or the initial design requirements, and it reflects the embodiment of the value of the technology. The performance attribute focuses on solving the core problem to improve product performance and perfect the whole system. The relationship attribute aims to build up or cut off the connection among the product components to make the whole system more integrated and perfect. The emotion attribute is mainly used to satisfy the emotional or spiritual desire of the designers by changing the appearance or expression of the products. The mapping process between the problem situation and patent knowledge based on the situation attributes from these dimensions is shown in Figure 1.

*2.2. Mapping Process between the Problem Situation and Patent Knowledge Based on Multigranular Situation Attributes.* Granulation is a method of summarizing knowledge, and granularity is a measure of knowledge abstraction [30]. Multiple characterizations of the design problems represent the process of gradually clarifying the problem situation. Rough situation attributes are used to find inaccurate solutions in patent knowledge. When the problem constraints are more specific about the problem situation, the more refined situation attributes are used to search for more accurate solutions in the patent knowledge. This interactive process is continued between the problem and knowledge spaces until the exact solution is obtained. For example, the situation attribute can be considered as "separating substances" extracted from the design problem of "how to clean stains on clothes." According to different solutions, "separating substances" can be defined as "removing substances" or "decomposing substances." Further, "removing substances" can be separated into "removing solids," "removing liquids," "removing gas," and so on. The specific mapping process is shown in Figure 2.

*2.3. Normalized Representation of Problem Situation and Extraction of Attributes.* Because of the various representations of a problem situation, great difficulties are encountered in the extraction of situation attributes. Therefore, describing the problem situation in a standardized manner is necessary. According to the cognitive mechanism of human beings and the characteristics of the problem situation, the form "the object is  $\underline{?}$  and the operation to the object is  $\underline{?}$ " The combination of "operation + object" can be extracted as a situation attribute. For example, in the "Let clothes clean" event, the problem situation is described by "the object of operation is clothes, and the operation used to reach the goal is washing." The situation attribute is "washing + clothes." The performance dimension mainly aims to resolve the technical conflicts of the whole system to improve the product performance. Thus, the form "to improve the performance of products, the parameter needed to be

optimized is  $\underline{?}$ , and the parameter that becomes worse is  $\underline{?}$ " can describe the problem situation. The optimized and worse parameters are extracted as situation attributes. For example, in the event of "improving aircraft flight stability," the problem situation can be described as "to improve the performance of products, the parameter needed to be optimized is the intensity, and the parameter that can worsen the situation is the weight." The intensity and weight parameters are extracted as situation attributes. In the relationship dimension, because any unit of a system is composed of two components and the components interact, the method of material field analysis can be used to optimize the interaction among components of the products. "Component 1 (role sender) is  $\underline{?}$ , component 2 (role receiver) is  $\underline{?}$ , and the interaction between them is  $\underline{?}$ " is used to describe the problem situation. The two components and their interaction are extracted as situation attributes. For example, in the "bearing failure because of dust" event, "component 1 (role sender) is dust, component 2 (role receiver) is bearing, and the interaction between them is harmful effect" is used to describe the problem situation. Dust (component 1), bearing (component 2), and harmful effect are extracted as situation attributes. From the emotional dimension, a concise emotion vocabulary is provided to describe the appearance or expression of products, and they can accurately reflect the feelings of the designer about the product. "The feeling got from the product is  $\underline{?}$ " can be used to describe the problem situation, and the feeling is extracted as a situation attribute. For example, in the "designing a gorgeous dress" event, "the gorgeous feeling obtained from the product is  $\underline{?}$ " can be used to describe the problem situation, and gorgeous feeling is extracted as a situation attribute. The normalized expression of the problem situation and the extraction of situation attributes are shown in Figure 3.

### 3. Establishment of the Structured Attribute Database about the Problem Situation

To realize the complex mapping relationship between the problem situation and knowledge patent, establishing a structured attribute database from different dimensions and different abstract granularities is necessary. From the function dimension, the functional basis is selected as the function attributes according to Hertz's [31] combination and classification of functions and flows. As a type of standardized representation of a function, a functional basis is highly abstracted and integrated. It is divided into two layers. The first layer consists of 11 categories, namely, shift, regulate, absorb, combine, detect, stabilize, import, accumulate, output, produce, and separate. The second layer consists of 52 subcategories, which include stabilizing the motion parameters, stabilizing the process parameters, stabilizing the geometric parameters, and so on. Figure 4 shows these layers.

The performance dimension is divided into two layers according to the technical contradictions in TRIZ theory. The first layer consists of 39 categories, including speed, force, power, reliability, and so on. The second layer consists of 102 subcategories, including angular velocity, linear

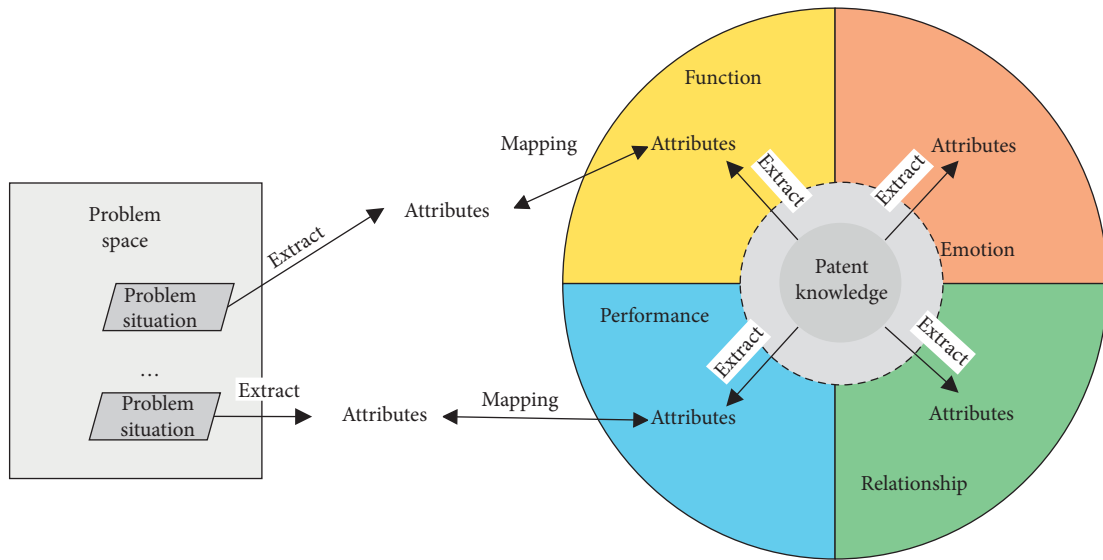


FIGURE 1: Mapping process between the problem situation and patent knowledge based on multidimensional situation attributes.

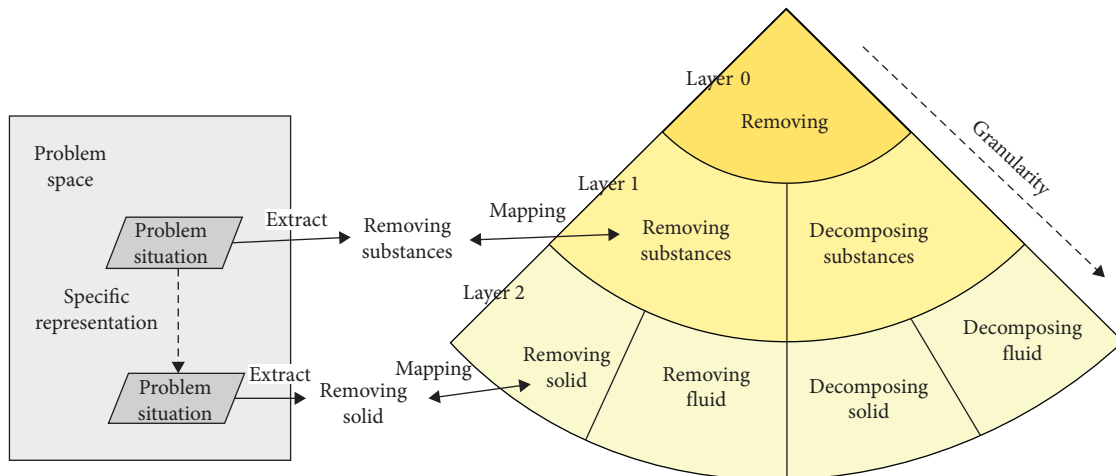


FIGURE 2: Mapping process between the problem situation and patent knowledge based on multigranular attributes.

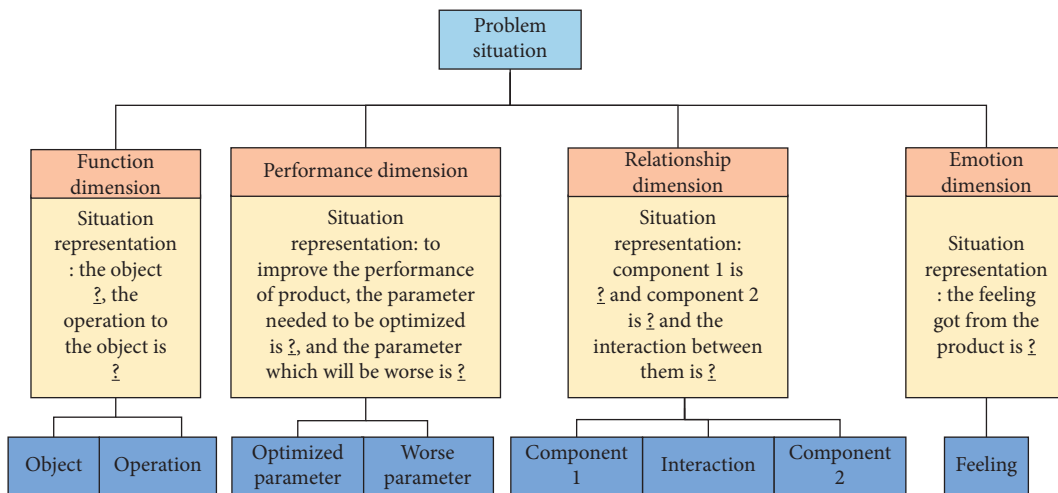


FIGURE 3: Normalized representation of the attributes based on the problem situation.

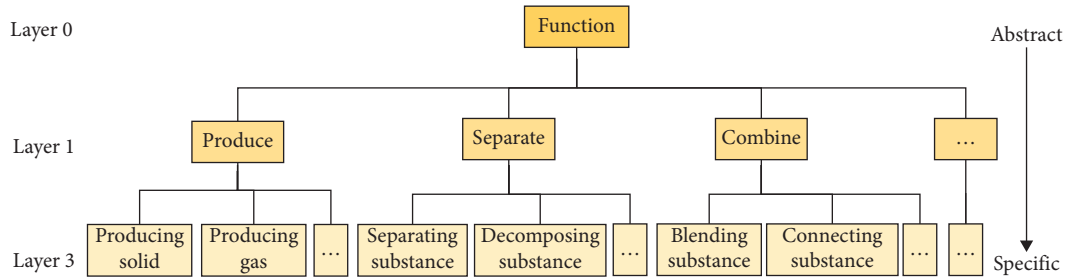


FIGURE 4: Situation attributes from the function dimension.

velocity, ampere force, Coulomb force, and so on. These layers are shown in Figure 5.

The relationship dimension is divided into two layers according to substance-field analysis in TRIZ theory. The first layer consists of four categories, namely, incomplete, harmful integrity, insufficient integrity, and excess integrity models. The second layer contains approximately 25 sub-categories, including the following: the model lacks component 1, the model lacks component 2, the model of component 1 is harmful to component 2, and so on. They are shown in Figure 6.

The emotion dimension is first divided into two layers in this paper. The first layer consists of 42 categories, which include color, harmony, stabilization, and so on. The second layer consists of 162 subcategories, which include fashion, simplicity, appropriateness, brightness, maturity, and so on. Figure 7 shows these layers.

To facilitate computer storage of the patent knowledge, structuralizing the situation attributes is necessary. The set of situation attributes for each patent can be represented as follows:

$$PC = \{F, P, R, E\}, \quad (1)$$

where  $PC$  is the patent knowledge and  $F$ (function),  $P$ (performance),  $R$ (relationship), and  $E$ (emotion) are the four dimensions of the situation attributes. The function attributes can be represented as  $F = [y_1, y_2, \dots, y_n]$  and  $y_i = (x_1, x_2, \dots, x_n)$ , where  $y_i$  and  $x_i$  are the situation attributes of the first and second layers, respectively. The attributes of performance, relationship, and emotion can be represented in the same manner.

#### 4. Mapping Relationship between the Problem Situation and Patent Knowledge Based on the Attributes

According to the attribute database, a single mapping relationship between the problem situation and patent knowledge is created. To provide the designers with accurate and abundant patent knowledge resources related to their problems, deeply analyzing the problem situation and patent knowledge is necessary to understand the design intentions and to accurately introduce the patent knowledge to the designers as much as possible. Because the design-problem situation and patent knowledge are described in the form of natural language, they are very

subjective and different. Establishing a coreference relationship among the situation attributes is an essential requirement to obtain the correlation of patent knowledge. Using the semantic characteristics of natural language is a reliable method for obtaining the relationship among the situation attributes by computing the semantic similarity. However, the selected semantic of words is closely related to the context, and some differences exist in different text environments. Thus, establishing the coreference relationship of the attributes may not possibly be obtained because of their context. Word2vec in Google was an efficient tool to represent words as numerical vectors in 2013 [32]. Its CBOW (Bag-of-words model) could generate high-dimensional feature vectors of a target word according to the frequency of the words near the target word in the text. A feature vector can effectively reflect the semantic weight of the target words in the text. Thus, the semantic relationship of the target words can be obtained by calculating the cosine similarity of the feature vectors of the different target words. Therefore, the Word2vec tool is used to establish the coreference relationship among the situation attributes to obtain the correlation of the patent knowledge and effectively expand its application scope.

The situation attributes of the design problem can be obtained from the designers. However, the situation attributes of the patent knowledge cannot be directly obtained from the patent knowledge until deep reading has been performed by an expert group. Because of the large number of patents and their quick updates, the manual deep-reading method and patent classification greatly affect the efficiency of patent-knowledge utilization. According to the characteristic of natural language, computer natural-language-processing technology is used to extract the representative feature words from the patent texts, and an appropriate classification algorithm is selected to automatically classify the patent text using situation attributes as categories. Because the situation attributes of products or systems include four categories, namely, function, performance, relationship, and emotion, they are described in the form of natural language, and all of them can be realized by computer natural-language-processing technology. Because of space constraints, this paper presents only the situation attributes from the function dimension as an example to show the experimental process of automatic classification of the patent text. Other attributes are processed in the same way.

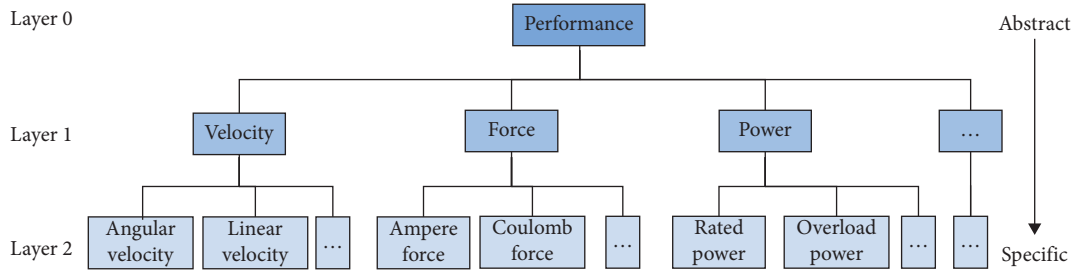


FIGURE 5: Situation attributes from the performance dimension.

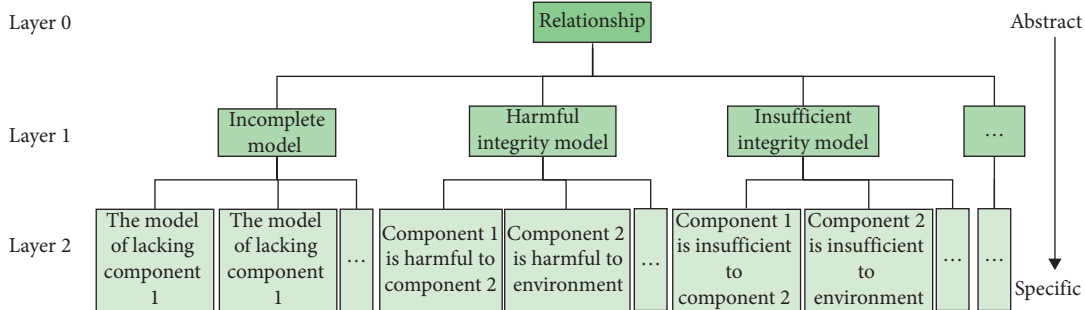


FIGURE 6: Situation attributes from the relationship dimension.

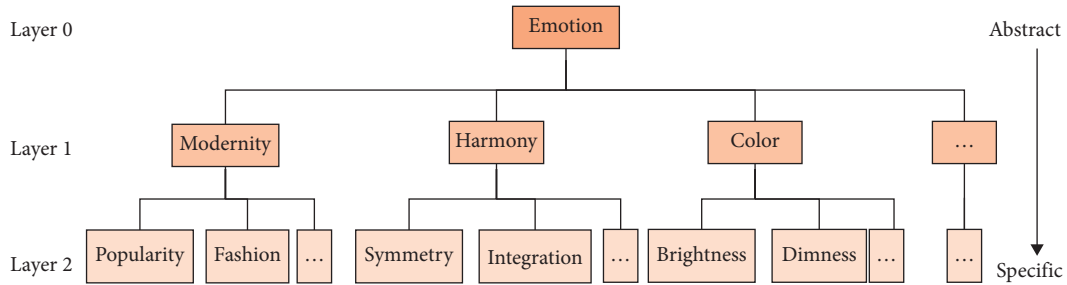


FIGURE 7: Situation attributes from the emotion dimension.

## 5. Automatic Patent Classification Based on Function Attributes

According to the general method of automatic classification using computer natural-language-processing technology, we propose the automatic classification process based on the situation attributes, as shown in Figure 8. According to the international classification standard (IPC), Chinese patents related to engineering technology are downloaded, including operation, transportation, mechanical engineering, weapons, blasting, lighting, heating, electricity, and so on. Situation attributes are extracted from the patent and manually labeled in different dimensions and different abstract granularities. These labeled patents constitute the text set of automatic classification experiments, and the text set is divided into training and test sets using a certain separation ratio. Feature words, which are representative of the patent information, are selected from the training and test sets. Their feature vectors are obtained by calculating the word frequency in the text set. Then, appropriate classification-training methods are selected to develop a classifier. The test set is classified by the classifier, and the results are compared

with its manual labels. Thus, the classifier accuracy can be evaluated. When the classification results finally match the defined target, the classifier can be applied to the classification of any unknown patent. The details of each step will be explained in the later sections.

**5.1. Patent-Text Preparation.** Because the patent text contains a lot of information, including the patent number, title, abstract, claim, description, and so on, no unified conclusion can be arrived at regarding which parts are currently appropriate to be a representative of the patent text. The title and abstract of a patent summarize the main design problems and invention contents of the patent, covering the core creativity and novelty information of the patent. In addition, Zhang [33], Liang [34] have tested "title + abstract" as the sample of patent information extraction, and the results show that "title + abstract" can effectively express the main creative information of a patent. Therefore, this paper considers that the form of "title + abstract" can better meet the extraction of situation attributes. 843 samples of invention patents are randomly downloaded from a Chinese

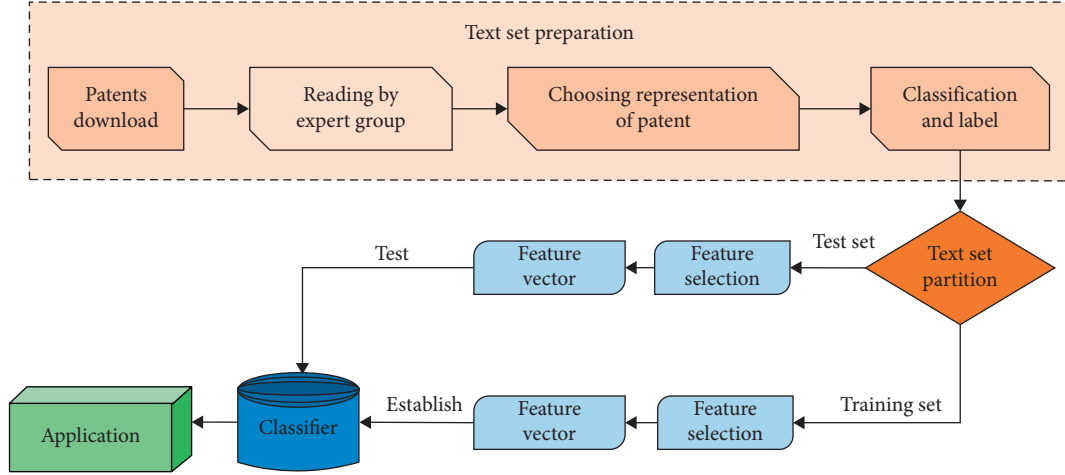


FIGURE 8: Process of automatic patent classification based on the situation attributes.

patent website. After the deep reading by the expert group, the function attributes of each patent text are defined and manually labeled as a sample set for the classification experiments.

**5.2. Feature Selection.** To select the feature words that can represent the text information, the text needs to be divided into a single word. The mature open-source Chinese lexical analysis system ICTCLAS [35] is used to segment the representative of the patent title and abstract. Considering the small number of patent samples and the limited length of each text, the feature selection method of the term frequency-inverse document frequency (TF-IDF) [36] is selected to calculate the weight of each feature word in the text. According to the sequence of word weights, the top 80% of the words are selected as feature words in the experiments. TF-IDF evaluates the importance of a word or phrase to a text by counting its frequency. If a word or phrase frequently appears in a text and rarely in the other texts, we consider that the word or phrase has a highly distinguished ability among the categories and is suitable for classification. The formula for calculating the eigenvalues of TF-IDF is expressed as follows:

$$w_{ij} = tf(t_j, d_i) \cdot idf(d_i), \quad (2)$$

where  $tf(t_j, d_i)$  is the term frequency and  $idf(d_i)$  is the inverse text frequency.

$$tf(t_j, d_i) = \frac{m_{ij}}{m_{ik}}, \quad (3)$$

where  $m_{ij}$  is the number of times feature word  $t_j$  appears in text  $d_i$ ,  $k$  is the text serial number, and  $m_{ik}$  is the total number of times the words appear in text  $d_i$ .

$$idf(d_i) = \log \frac{N}{n}, \quad (4)$$

where  $N$  is the number of all texts and  $n$  represents the number of texts that contains term  $t_j$ . TF-IDF tends to filter out the common words and reserve important words.

**5.3. Vectorization of the Patent Text.** After the feature words are selected, the eigenvalue calculated in the feature selection is used as the weight of the feature word. Thus, text  $d_i = (t_1, t_2, \dots, t_n)$ , which is expressed as  $d_i = (t_1, w_1; t_2, w_2; \dots; t_n, w_n)$ , is simply expressed as  $d_i = (w_1, w_2, \dots, w_n)$ . The feature vectors of all texts form the spatial eigenvector of the whole text set.

**5.4. Classification Experiment and Result Analysis.** The 11 function attributes, namely, shift, regulate, absorb, combine, detect, stabilize, import, accumulate, output, produce, and separate, are selected as classification categories, and the text set is divided into two types of ratios: 80% and 20% training and testing sets, respectively, and 70% and 30% training and testing sets, respectively. Different text-learning algorithms are used to establish the classifier in the training set. The classifier is tested in the test set, and the classification accuracy is calculated to analyze the experimental results. At present, various classification-learning algorithms are available, including the K-nearest neighbor, SVM [37], naive Bayesian, multilayer perceptron (MLP) [38], and so on. Because of the differences in the experimental sets, no consistent conclusion can be arrived at to identify which classification algorithm is more effective. The present study selects the MLP and SVM in carrying out the experiments. The harmonic mean F1-score of P (precision) and R (recall) is used as evaluation criteria. P refers to the number of correctly classified texts divided by the number of classified texts, and R refers to the number of correctly classified texts divided by the actual number of texts in the test set. The formula is expressed as follows:

$$F = \frac{2PR}{P + R} \quad (5)$$

The classification results in each category are shown in Figures 9 and 10.

From the experimental data, the following conclusions are obtained.

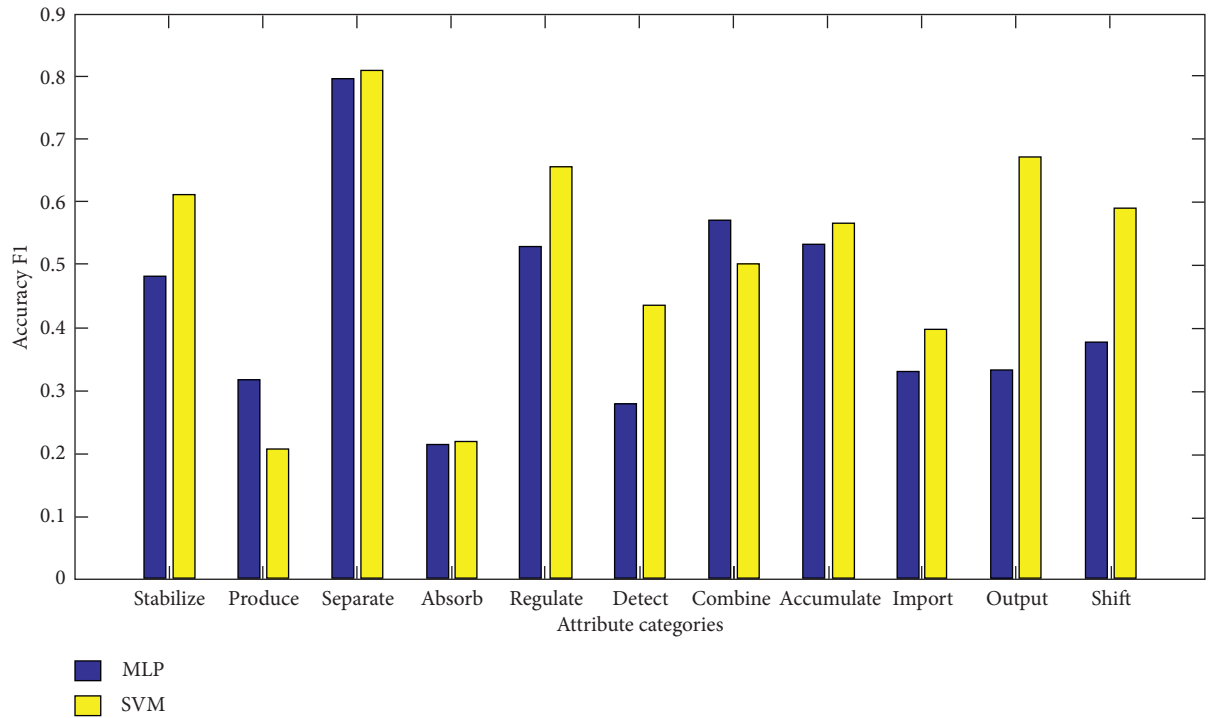


FIGURE 9: Text set divided into 70% training set and 30% test set.

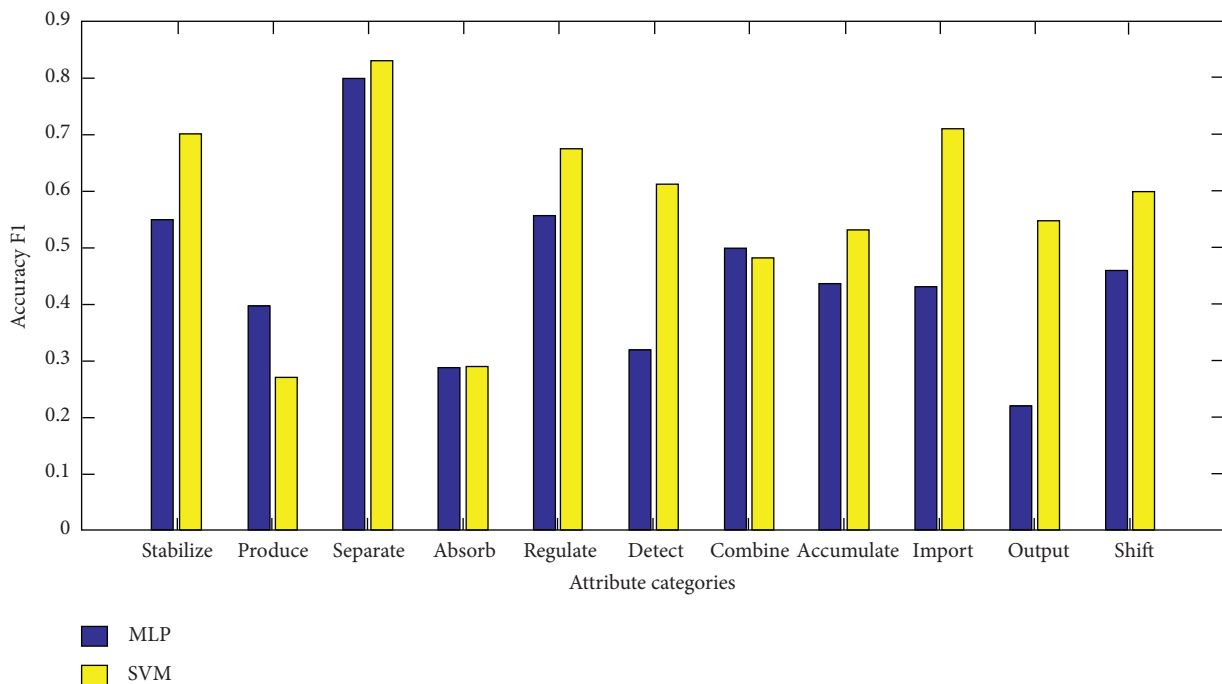


FIGURE 10: Text set divided into 80% training set and 20% test set.

(1) When a classifier is established using the SVM algorithm, the accuracy of F, except that in the three categories of produce, absorb, and combine, is approximately 50%. The accuracy of F in the other category is as high as 83%. Compared with the random classification, F has significantly higher accuracy, which indicates the feasibility of the

computer natural-language-processing technology and the big advantage of automatic classification of the patent text.

(2) The accuracy of F is higher when the text set is divided into 80% training set and 20% test set than that when the text set is divided into 70% training set and 30% test set, except for the combine, accumulate,



and output categories. This result indicates that properly increasing the proportion of training set in the process of automatic patent-text classification can effectively improve the classification accuracy.

- (3) Comparison of the accuracy of the two classification algorithms clearly shows that the accuracy of the SVM algorithm is significantly higher than that of the MLP algorithm except for the produce and combine categories. This result indicates that the SVM algorithm is better for automatic classification of the patent text.

At present, the accuracy of the patent automatic classification is not ideal. But the impact of different classification algorithms on the classification results will provide the basis for the subsequent establishment of a larger range of corpus, so as to effectively improve the accuracy of patent classification.

### 6. Development of Patent-Knowledge System Based on Situation Attributes

A total of 65548 patent texts (in the form of “title + abstract”) are automatically obtained from a professional patent database using a Web crawler program. Using the situation attributes (from the four dimensions of function, performance, relationship, and emotion) as categories, these patent texts are classified by the automatic classification method proposed in this paper, and a classification set is established.

To analyze the problem situation and retrieve the analogical knowledge in the design process, a patent-knowledge-retrieval system based on situation attributes is developed using Java in the Eclipse rapid development platform. The users can describe the problem situation using natural language from the four dimensions: function, performance, relationship, and emotion. For example, the users can input the situation attribute in the form of “operation and object” about the design problem from the function dimension. When the situation attribute input matches one of all the situation attributes in the attribute database, the system pushes the designer patent texts that belong to the situation attribute. When the situation attribute input cannot match any situation attribute in the attribute database, the system immediately calculates the semantic similarity between the input and each situation attribute in the database and pushes three to five situation attributes that are most similar to the input. The users can choose one situation attribute, and the system pushes the patent texts that belong to the selected situation attribute to the users. For example, in the matter of the problem of Sewage treatment in the manufacturing industry, the situation attribute of “treat + sewage” can be extracted from the function dimension. Then, the user inputs “treat” and “sewage” to text boxes, respectively. Next, the user needs to choose one situation attribute from five as required. Finally, the tool provides 810 patents for the user and the patents’ categories include lighting, heating, electricity, and so on, which can effectively inspire the user to complete the innovative design process. The main interfaces of the tool’s application process are shown in Figures 11 and 12.

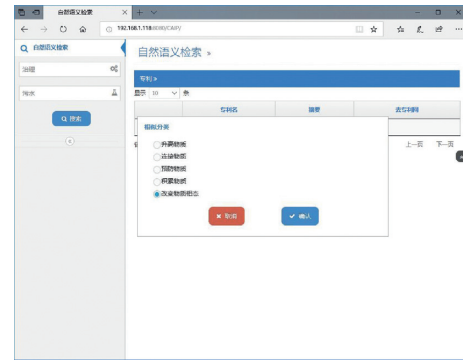


FIGURE 11: The interface of attribute choice.



FIGURE 12: The interface of search results.

### 7. Conclusion

This study investigates a method of patent-knowledge representation and automatic classification based on situation mapping. The method can effectively assist designers to obtain patent knowledge related to their design problem so that they can develop an innovative design process. The main content includes the following: (1) On the basis of situation attributes, the mapping process between the problem situation and patent knowledge is developed, and the multidimensional and multigranular attribute database is established. (2) The mapping relationship between the problem situation and patent knowledge is formed, and the coreference relationship among the situation attributes is established. (3) An automatic patent-classification method based on situation attributes is proposed, and using the classification method, an application system of patent knowledge is developed.

Because of the complexity and diversity of problem situations, the mapping relationship between the problem situation and patent knowledge needs to be improved further. Related research will be carried out on the following four aspects: (1) Further research is to establish more abundant situation attributes and coreference relationship, so as to form various mappings between the design problem and patent knowledge. (2) It is necessary to propose a more reasonable and efficient automatic classification method, so as to improve the classification accuracy of Chinese patent or English patent. (3) The next step of research will consider how to get the weight distribution of situation attributes according to user requirements.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors have no conflicts of interest to declare.

## Acknowledgments

This work was supported by the Science and Technology Ministry Innovation Method Program of China (No. 2017IM040100) and by the Applied Basic Research Funding Project of Sichuan (No. 2018JY0119).

## References

- [1] Y.-H. Tseng, C.-J. Lin, and Y.-I. Lin, "Text mining techniques for patent analysis," *Information Processing & Management*, vol. 43, no. 5, pp. 1216–1247, 2007.
- [2] F. Gobet, "Chunking models of expertise: implications for education," *Applied Cognitive Psychology*, vol. 19, no. 2, pp. 183–204, 2010.
- [3] P. Chandler and J. Sweller, "Cognitive load while learning to use a computer program," *Applied Cognitive Psychology*, vol. 10, no. 2, pp. 151–170, 2010.
- [4] G. Sala and F. Gobet, "Experts' memory superiority for domain-specific random material generalizes across fields of expertise: a meta-analysis," *Memory & Cognition*, vol. 45, no. 2, pp. 183–193, 2017.
- [5] A. Carmeli, G. Z. Sheaffer, and T. R. ShimoniReiter-Palmon, "Transformational leadership and creative problem-solving: the mediating role of psychological safety and reflexivity," *The Journal of Creative Behavior*, vol. 48, no. 2, pp. 115–135, 2014.
- [6] B. M. Rottman, D. Gentner, and M. B. Goldwater, "Causal systems categories: differences in novice and expert categorization of causal phenomena," *Cognitive Science*, vol. 36, no. 5, pp. 919–932, 2012.
- [7] C. R. Wolfe and V. F. Reyna, "Semantic coherence and fallacies in estimating joint probabilities," *Journal of Behavioral Decision Making*, vol. 23, no. 2, pp. 203–223, 2010.
- [8] S. B. Day and R. L. Goldstone, "The import of knowledge export: connecting findings and theories of transfer of learning," *Educational Psychologist*, vol. 47, no. 3, pp. 153–176, 2012.
- [9] I. J. Bennett and D. J. Madden, "Disconnected aging: cerebral white matter integrity and age-related differences in cognition," *Neuroscience*, vol. 276, pp. 187–205, 2014.
- [10] F. Günther, C. Dudschig, and B. Kaup, "LSAfun-an R package for computations based on Latent Semantic Analysis," *Behavior Research Methods*, vol. 47, no. 4, pp. 930–944, 2015.
- [11] H. Noh, Y. Jo, and S. Lee, "Keyword selection and processing strategy for applying text mining to patent analysis," *Expert Systems with Applications*, vol. 42, no. 9, 2015.
- [12] H. Cong and L. H. Tong, "Grouping of TRIZ Inventive Principles to facilitate automatic patent classification," *Expert Systems with Applications*, vol. 34, no. 1, pp. 788–795, 2008.
- [13] W.-q. Li, Y. Li, J. Chen, and C.-y. Hou, "Product functional information based automatic patent classification: method and experimental studies," *Information Systems*, vol. 67, pp. 71–82, 2017.
- [14] C. V. Trappey, A. J. C. Trappey, H.-Y. Peng, L.-D. Lin, and T.-M. Wang, "A knowledge centric methodology for dental implant technology assessment using ontology based patent analysis and clinical meta-analysis," *Advanced Engineering Informatics*, vol. 28, no. 2, pp. 153–165, 2014.
- [15] A. J. C. Trappey, P. P. J. Chen, C. V. Trappey, and L. Ma, "A machine learning approach for solar power technology review and patent evolution analysis," *Applied Sciences*, vol. 9, no. 7, p. 1478, 2019.
- [16] H. Yu and D. Fan, "Man-made boards technology trends based on TRIZ evolution theory," *Physics Procedia*, vol. 33, no. 6, pp. 221–227, 2012.
- [17] C. Lee, Y. J. Jeon, and Y. Park, "Monitoring trends of technological changes based on the dynamic patent lattice: a modified formal concept analysis approach," *Technological Forecasting and Social Change*, vol. 78, no. 4, pp. 690–702, 2011.
- [18] J. W. Chen, X. D. Xu, and X. Q. Shen, "Review of computer aided product innovation design based on knowledge," *Applied Mechanics and Materials*, vol. 713–715, pp. 2335–2340, 2015.
- [19] A. Cardillo, F. G. Cascini, and F. Rotini, "Computer-aided embodiment design through the hybridization of mono objective optimizations for efficient innovation process," *Computers in Industry*, vol. 62, no. 4, pp. 384–397, 2011.
- [20] Z. Li, M. Atherton, and D. Harrison, "Identifying patent conflicts: TRIZ-led patent mapping," *World Patent Information*, vol. 39, pp. 11–23, 2014.
- [21] A. S. Ghareb, A. A. Bakar, and A. R. Hamdan, "Hybrid feature selection based on enhanced genetic algorithm for text categorization," *Expert Systems with Applications*, vol. 49, pp. 31–47, 2016.
- [22] M. Labani, P. Moradi, F. Ahmadizar, and M. Jalili, "A novel multivariate filter method for feature selection in text classification problems," *Engineering Applications of Artificial Intelligence*, vol. 70, pp. 25–37, 2018.
- [23] K. Chen, Z. Zhang, J. Long, and H. Zhang, "Turning from TF-IDF to TF-IGM for term weighting in text classification," *Expert Systems with Applications*, vol. 66, pp. 245–260, 2016.
- [24] F. Zhu, D. X. Wang, and Y. Liu, "A supervised requirement-oriented patent classification scheme based on the combination of metadata and citation information," *International Journal of Computational Intelligence Systems*, vol. 8, no. 3, pp. 502–516, 2015.
- [25] J.-L. Wu, P.-C. Chang, C.-C. Tsao, and C.-Y. Fan, "A patent quality analysis and classification system using self-organizing maps with support vector machine," *Applied Soft Computing*, vol. 41, pp. 305–316, 2016.
- [26] K.-K. Lai and S.-J. Wu, "Using the patent co-citation approach to establish a new patent classification system," *Information Processing & Management*, vol. 41, no. 2, pp. 313–330, 2005.
- [27] D.-R. Liu and M.-J. Shih, "Hybrid-patent classification based on patent-network analysis," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 2, pp. 246–256, 2011.
- [28] Y.-L. Chen and Y.-C. Chang, "A three-phase method for patent classification," *Information Processing & Management*, vol. 48, no. 6, pp. 1017–1030, 2012.
- [29] L. I. Yan and W. Li, *Method to Creative Design*, Science Press, Beijing, China, 2012.
- [30] W. Pedrycz, B. Russo, and G. Succi, "Knowledge transfer in system modeling and its realization through an optimal allocation of information granularity," *Applied Soft Computing*, vol. 12, no. 8, 2012.

- [31] J. Hirtz, D. A. Stone, S. Szykman, and K. L. Wood, "A functional basis for engineering design: reconciling and evolving previous efforts," *Research in Engineering Design*, vol. 13, no. 2, pp. 65–82, 2002.
- [32] C. Cerisara, P. Král, and L. Lenc, "On the effects of using word2vec representations in neural networks for dialogue act recognition," *Computer Speech & Language*, vol. 47, pp. 175–193, 2018.
- [33] Y. Zhang, L. L. Shang, G. A. L. ZhangPorter, and D. Zhu, "A hybrid similarity measure method for patent portfolio analysis," *Journal of Informetrics*, vol. 10, no. 4, pp. 1108–1130, 2016.
- [34] Y. Lu, R. Tan, and J. Ma, "Study on patent text classification for product innovative design[J]," *Computer Integrated Manufacturing Systems*, vol. 19, no. 2, pp. 382–390, 2013.
- [35] D. Zeng, D. Wei, M. Chau, and F. Wang, "Domain-specific Chinese word segmentation using suffix tree and mutual information," *Information Systems Frontiers*, vol. 13, no. 1, pp. 115–125, 2011.
- [36] Y. Lu, Z. M. Liang, and L. Cao, "Improved particle swarm optimization algorithm and its application in text feature selection," *Applied Soft Computing*, vol. 35, pp. 629–636, 2015.
- [37] D. Zhang, H. Xu, Z. Su, and Y. Xu, "Chinese comments sentiment classification based on word2vec and SVMperf," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1857–1863, 2015.
- [38] N. Talebi, Nasrabadi, M. Ali, Mohammad-Rezazadeh, and Iman, "Estimation of effective connectivity using multi-layer perceptron artificial neural network," *Cognitive Neurodynamics*, vol. 12, no. 1, pp. 1–22, 2018.