A Knowledge Content Matching Degree Calculation Method Supporting Process Knowledge Recommendation

Peilin Shao, Lihong Qiao, and Zhicheng Huang(✉)

School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China
zc_huang@buaa.edu.cn

Abstract. Product process design relies upon the process knowledge support acquired by process designers and enterprises. Process knowledge recommendation has been attached more importance by enterprises in that it can quickly provide process designers with accurate and appropriate process knowledge. In this paper, a "scene-label-classification" knowledge recommendation scheme for process knowledge graph is established, and a knowledge content matching degree calculation method supporting process knowledge recommendation is proposed. Design requirements are described by parameterized characteristics from various dimensions such as material, precision requirements. Based on knowledge coding, the attribute characteristics of process knowledge are uniformly identified and associated with requirement scene characteristics. A requirements-knowledge semantic vector space model is established, and a calculation method of matching degree between design requirements and knowledge content is proposed based on improved cosine distance. This scheme enables quickly positioning the corresponding process knowledge based on coding labels, and final process knowledge recommendation candidate set is obtained through matching degree calculation and filtering sort, which thus enables dynamic classification and appropriate recommendation of process knowledge.

Keywords: Knowledge recommendation · Knowledge coding · Matching degree calculation

1 Introduction

As a bridge connecting product design and product manufacturing, the current process design process relies upon the experience of process designers, as well as guide of existing process knowledge of the enterprise. During this process, inexperienced process designers frequently need to spend a lot of time searching for the required knowledge, resulting in the efficiency and quality difference of process design. In this case, due to the capacity to quickly provide appropriate process knowledge for process designers according to process design requirement, process knowledge recommendation technology has been widely used in machining [1], assembly [2], sheet metal and other different process.

L. Zhang et al. (Eds.): CINT 2022, CCIS 1714, pp. 181–191, 2022.
https://doi.org/10.1007/978-981-19-8915-5_16

With the continuous development and wide application, knowledge recommendation oriented to process knowledge graph has become a research focus, including process design requirement analysis, rapid positioning of requirements and knowledge, and knowledge matching algorithm design. According to the systematic analysis of process design requirements, Zhou et al. [3] obtained process requirements from five aspects such as users and design information, and converted them into corresponding process characteristics. Guo et al. [4] obtained hidden user requirements from user requirement data and established association mapping for product attributes. However, most of current process design requirements analysis is limited to the part information or user requirement, short of comprehensive analysis of design requirements. For the rapid positioning of requirements and knowledge, it is an effective method to construct the unified representation of knowledge attributes [5, 6]. For the matching between process knowledge and design requirements, currently semantic matching calculation methods such as vector space model-based matching calculation [7], ontology-based semantic matching calculation [8] and deep learn-based matching calculation are commonly used. Wang et al. [9] applied TF-IDF algorithm based on vector space model in the matching process of manufacturing knowledge and design requirements. Renu et al. [10] used the text matching algorithm to realize the knowledge retrieval and sharing of the text-based assembly process scheme.

To support the fast and high matching degree process knowledge recommendation, this paper proposes a method to map between process design requirements and process knowledge attributes by constructing the coding label of process knowledge graph. A requirements-knowledge semantic vector space model is established and the matching degree between requirements and knowledge is calculated based on improved cosine distance.

2 "Scene-Label-Classification" Process Knowledge Recommendation Framework

Oriented to knowledge graph, this paper proposed a process knowledge recommendation scheme-"scenario-label-classification", which is shown in Fig. 1. Different dimensions of process design requirements are described by parametric characteristics such as material, size characteristic, part type, precision, etc. At the same time, unified identification of process knowledge attributes are constructed using knowledge coding label, and parameterized requirement scenes are correlated with process knowledge attributes. Therefore, rapid positioning from process design requirements to process knowledge attributes is achieved and initial process knowledge recommendation candidate is generated. Process design requirement semantic vector is obtained by analyzing parameterized requirement scene. Meanwhile, process knowledge semantic vector of initial candidates is generated by TransE algorithm and dimensionality reduction to the same dimension as the process design requirement semantic vectors. Based on the improved cosine distance, the matching degree between them can calculated and filtrated according to matching degree threshold value. The final process knowledge recommendation candidate set is obtained so as to enable the dynamic classification of process knowledge under different requirement scenarios.

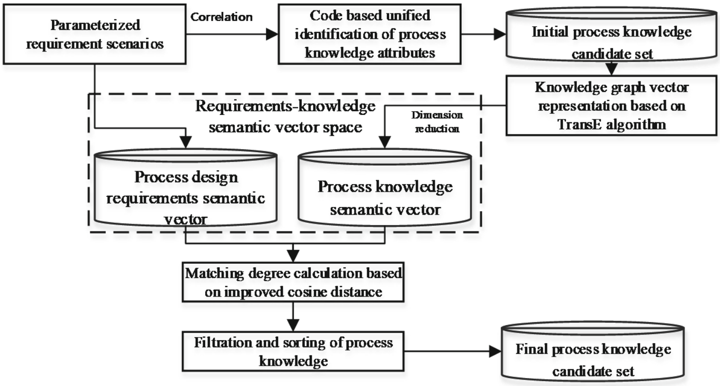

Fig. 1. "Scene-label-classification" process knowledge recommendation framework

3 Correlation Between Requirement Scenarios and Process Knowledge Attributes

The process design requirements of different dimensions are analyzed, and requirement scenarios of process knowledge recommendation are described through materials, dimension characteristics, part types and other parameterized characteristics. Meanwhile, the process knowledge attribute characteristics are identified uniformly by knowledge coding, and then requirement scene characteristics and process knowledge attribute characteristics are expressed in correlation.

According to the requirement scene of process knowledge recommendation, this paper divides process design requirement into three aspects: part information, product information and other information. Among them, part information contains the geometric information and non-geometric information including surface shape, size, precision information, part materials, technical requirements, etc. Compared with the part, product information contains the number of parts, the position relationship between parts, and the assembly relationship between parts, etc. Other information includes processing/assembly time limits, types of knowledge required, etc. Through parameterized characteristic description, the requirement scenarios set $R(R_1,R_2,\ldots,R_q)$ is established, where $R_1,R_2,\ldots,R_q$ represent the parameterized requirements characteristics, which facilitates the construction of association mapping between process design requirements and process knowledge attributes.

In this paper, process knowledge attributes are uniformly identified by constructing coding labels in the knowledge graph, enabling the quick positioning of knowledge attribute characteristics according to design requirements. However, when there are too many parameterized demand features, it may be impossible to find the process knowledge that meets all the demand characteristics. Therefore, the quantity threshold value S_n meeting the requirements of parameterized characteristics is set. And process knowledge that meets the conditions (including the number of requirement features $\geq S_n$) is selected according to the demand scenario, so as to generate an initial process knowledge recommendation candidate set $K(K_1,K_2,...,K_m)$, where $K_1,K_2,...,K_m$ represent contained process knowledge.

4 Construction of Requirements-Knowledge Semantic VSM

Vector space model (VSM) is an information retrieval model which is widely used and effective in recent years. The core idea of VSM is to represent text content with vectors and map it to n-dimensional vector space, thus transforming the similarity problem between texts into the similarity problem between vectors in multi-dimensional space. In this paper, semantic vector expression and dimensionality reduction are respectively conducted for requirements scenarios and initial process knowledge recommendation candidate set, thus obtaining requirement semantic vector and several process knowledge semantic vectors. Then the requirements-knowledge semantic VSM is established so as to calculate the matching degree between process design requirements and process knowledge.

4.1 Construction of Requirement Semantic VSM

The construction idea of semantic VSM is that text is regarded as a combination of several independent characteristic terms, and a high-dimensional space is constructed with these different characteristic terms. Each characteristic term is one dimension of this space, and text is regarded as a space vector. In this paper, parameterized characteristics such as material, dimension feature and part type are utilized to describe the process design requirements, so each parameterized characteristic is the characteristic term of requirement semantic VSM. Assuming that process design requirements consist of independent parameterized characteristics, it can be expressed as:

$$R = \{r_1, d_1; r_2, d_2; \ldots\ldots; r_n, d_n\} \tag{1}$$

Therein $r_i (1 \le i \le n)$ represents the parameterized characteristic name, and $d_i (1 \le i \le n)$ represents each parameter value/description content. For the parameterized characteristics such as technical requirements, the content is described in natural language, like "pay attention to temperature control and quenching transfer time of aluminum alloy materials in heat treatment process". In order to enable the matching between parameterized requirement characteristics described in natural language and the process knowledge content, it is necessary to generate text vector using NLP to calculate semantic matching degree.

Given different weights for each parameterized characteristic as the vector component, the text vector V_r used to represent the process design requirements can be expressed as:

$$V_r = \{Wr_1,\ d_1; W_{r2},\ d_2; \ldots\ \ldots; Wr_n, d_n\} \tag{2}$$

Therein $w_{ri} (1 \le i \le n)$ represents the weight of each parameterized characteristic, which is generally set by process designer according to the importance of each parameterized characteristic. Therefore, process design requirements can be represented by an N-dimensional characteristic vector.

4.2 Construction of Process Knowledge Semantic VSM

Aiming at each process knowledge graph in the process knowledge recommendation candidate set, the vector representation method of process knowledge graph based on TransE algorithm is adopted, so that each process knowledge could be represented by a p-dimension vector $V_{trans}(P \geq N)$. In order to calculate the matching degree between requirement vector and process knowledge vector, dimensionality reduction methods such as PCA is adopted to keep the dimension number and meanings of two vectors consistent. For the i^{th} process knowledge of the initial process knowledge recommendation candidate set, its semantic vector $V_{k\text{-}i}(1 \leq i \leq m)$ can be expressed as:

$$V_{k-i} = \left\{r_1, d_1'; \ r_2, d_2'; \ldots\ldots; r_n, d_n'\right\} \tag{3}$$

Therein $r_i(1 \leq i \leq n)$ represents n parameterized characteristics of process design requirements, and $d'_i(1 \leq i \leq n)$ represents the value/description content of this parameter (empty if no content). Different from requirement semantic vector, the weight of each parameterized characteristic of knowledge semantic vector is calculated based on statistics, among which the most commonly used method is TF-IDF method. Term Frequency (TF) weight indicates the number of times a characteristic item appears in this document. The more times it appears, the more important it is. However, for process knowledge, high appearance frequency of parameterized characteristic does not necessarily mean that it is more important. Therefore, in order to reduce the influence of TF weight, when the occurrence frequency of parameterized requirement characteristic is not 0, it can be expressed as:

$$\mathrm{TF}_{(i-k)} = 1 + log(1 + logj) \tag{4}$$

Therein j represents the number of occurrences of parameterized characteristic r_k in process knowledge K_i. If this parameterized requirement characteristic is not included in K_i, then $TF_{(i\text{-}k)} = 0$.

IDF (Inverse Document Frequency) weight refers to the frequency of all process knowledge contained in initial knowledge candidate set for one parameterized characteristic. If this parameterized characteristic appears in multiple process knowledge, it proves that its distinguishing ability is low, and therefore, IDF weight is expressed as:

$$\mathrm{IDF}_{(i-k)} = log(\frac{N}{n} + \alpha) \tag{5}$$

Therein α is a constant. In the similarity calculation between texts, if a keyword appears in all texts, its IDF value is extremely low. However, in the matching of requirements and process knowledge, if a parameterized characteristic appears in all process knowledge, it does not mean that this characteristic is unimportant. According to Formula (5), the larger α is, the weaker the distinguishing ability of this parameterized characteristic is. Therefore, α is set as 1 in this paper, N represents the number of process knowledge in the initial knowledge candidate set, and n represents the number of process knowledge with this parameterized characteristic. Finally, the weight of parameterized characteristic r_k in process knowledge K_i is:

$$w_{i-k} = \mathrm{TF}_{(i-k)} * \mathrm{IDF}_{(i-k)} \tag{6}$$

The semantic vector V_k used to represent a process knowledge can be expressed as:

$$V_k = \{w_{k1}, d'_1;\ w_{k2}, d'_2;\ \ldots\ldots;\ w_{kn}, d'_n\} \tag{7}$$

Therein $w_{ki}(1 \le i \le n)$ represents the weight of each parameterized characteristic obtained through Formula (6). For an initial process knowledge recommendation candidate set containing m process knowledge, a $m*n$ knowledge semantic VSM can be finally constructed:

$$V_{k-m} = \left\{ \begin{array}{c} w_{k11}, d'_{11}; \cdots; w_{k1n}, d'_{1n} \\ \cdots \\ w_{km1}, d'_{m1}; \cdots; w_{kmn}, d'_{mn} \end{array} \right\} \tag{8}$$

5 Knowledge Content Matching Degree Calculation

The calculation process of matching degree between requirements semantic vector and semantic vector of each process knowledge is shown in Fig. 2: Based on improved cosine distance, the matching degree of each process knowledge content in the initial candidate set and process design requirement is calculated. The matching degree threshold M_t and quantity threshold N_q is set in advance for comparison to filter out the process knowledge with low matching degree. The reserved process knowledge is sorted according to the matching degree value, and final process knowledge recommendation candidate set $K'(K_1,K_2,\ldots,K_q)$ is obtained and pushed to the process designer.

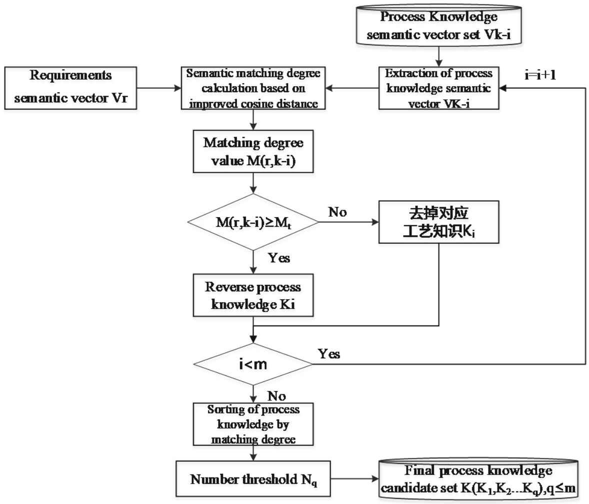

Fig. 2. The calculation process of matching degree

5.1 Matched-Degree Calculation Based on Improved Cosine Distance

After obtaining V_r and $V_{k\text{-}i}$, the semantic matching degree between requirements and process knowledge content can be expressed according to the matching degree between the vectors. In this paper, cosine distance between vectors is utilized:

$$M(i, k-i) = \cos\theta = \frac{V_r * V_{k-i}}{|V_r||V_{k-i}|} \tag{9}$$

For requirements semantic vector V_r and process knowledge semantic vector $V_{k\text{-}i}$, besides containing weights $w_{ri}(1 \le i \le n)$ and $w_{ki}(1 \le i \le n)$ of parameterized characteristics, descriptions $d_i(1 \le i \le n)$ and $d'_i(1 \le i \le n)$ of parameterized characteristics are also included. Compared with classical semantic calculation process based on cosine distance, this process also needs to calculate the matching degree between $d_i(1 \le i \le n)$ and $d'_i(1 \le i \le n)$. In this regard, this paper summarizes several situations that may occur when $d_i(1 \le i \le n)$ and $d'_i(1 \le i \le n)$ match:

a) Numerical matching. In this case, $d_i(1 \le i \le n)$ and $d'_i(1 \le i \le n)$ can be directly compared, which is applicable to parameterized requirement characteristics described by numerical values such as surface roughness and machining accuracy. The result has two cases: a match of 1 and a mismatch of 0.
b) Semantic matching. For parts material, parts type and other parameterized requirement characteristics described by simple text, $d_i(1 \le i \le n)$ and $d'_i(1 \le i \le n)$ can be directly compared, which is consistent with numerical matching. For parameterized requirement characteristics such as technical requirements that need to be processed by natural language, semantic matching degree between them should be calculated based on cosine distance. The matching result range is [0,1], where 0 indicates complete mismatch and 1 indicates complete match.

Therefore, based on classical cosine distance, this paper proposes a calculation method of matching degree between the requirements semantic vector V_r and the process knowledge semantic vector $V_{k\text{-}i}(1 \le i \le m)$ based on improved cosine distance. The calculation formula is:

$$M(r, k-i) = \frac{V_r * V_{k-i}}{|V_r||V_{k-i}|} = \frac{\sum_{j=1}^{n}\left[w_{r_j} * w_{(k-i)_j} * M\left(d_{r_j}, d_{(k-i)_j}\right)\right]}{(\sqrt{\sum_{j=1}^{n}(w_{r_j})^2} * (\sqrt{\sum_{j=1}^{n}(w_{(k-i)_j})^2}} \tag{10}$$

Therein $M(d_{rj}, d_{(k\text{-}i)j}, 1 \le j \le n)$ represents the matching degree between parameterized characteristics description contents of two vectors. The range of $M(r,k\text{-}i)$ is [0,1], and the larger $M(r,k\text{-}i)$ is, the higher matching degree between requirements and process knowledge is.

5.2 Candidate Process Knowledge Filtering Ranking

The matching degree *M(r,k-i)* between each knowledge and process design requirements is calculated and compared with threshold value M_t. If the matching degree value is greater than, the corresponding process knowledge is retained. Otherwise, corresponding process knowledge is eliminated. The reserved process knowledge is sorted from large to small according to matching degree value, and according to quantity threshold N_q, the final process knowledge recommendation candidate set is obtained and pushed to the designer.

6 Experimental Verification and Analysis

The validity of the proposed process knowledge recommendation scheme and knowledge content matching calculation method is verified by a shaft-hole part example of machining design. Firstly, according to the process specification of the part, the parameterized requirement characteristics information used to describe its process design requirements is summarized, as shown in Table 1.

Table 1. Parameterized requirement characteristics of a shaft-hole part

Requirement characteristics type	Requirement characteristics name	Description content
Part information	Part type	Shaft-hole
	Part material	45 steel
	Shape characteristic 1	Outer circle
	Dimensional information 1	∅40 mm × 180 mm
	Machining precision 1	IT7
	Surface roughness 1	1.6 μm
	Shape characteristic 2	Inner hole
	Dimensional information 2	∅20 mm × 180 mm
	Machining precision 2	IT8
	Surface roughness 2	3.2 μm
Non-part information	Technical requirement	Modulation hardness 220–250 HBW. Sharp edges blunt, remove edges and corners burrs

The corresponding process knowledge attributes are located in the knowledge base according to the knowledge coding, and the quantity threshold S_n value that meets the parameterized requirement characteristics is set as 8, and four characteristics including recommended knowledge category (machining route), part type (axle hole), and shape feature (1 is outer circle, 2 is inner hole) are must contained. An initial process knowledge recommendation candidate set containing 12 process knowledge is generated, and its description in "Machining Route" is shown in Table 2.

Table 2. Information of initial knowledge recommendation candidate set $K(K_1,K_2,\ldots,K_{12})$

Serial number	Process knowledge information
1	Rough turning - Semi-finish turning - Drilling - Reaming - Semi-fine hinge - Finish turning
2	Rough turning - Semi-finish turning - Drilling - Reaming - Semi-fine hinge - Rough grinding – Semi-finish grinding
3	Rough turning - Semi-finish turning - Drilling - Semi-fine hinge - Finish hinge - Finish turning
4	Rough turning - Semi-finish turning - Drilling - Semi-fine hinge - Finish hinge - Rough grinding – Semi-finish grinding
5	Rough turning - Semi-finish turning - Drilling - Semi-fine hinge - Rough grinding hole - Rough grinding outer circle
6	Rough turning - Semi-finish turning - Drilling - Semi-fine hinge – Rough grinding hole - Finish turning
7	Rough milling - Semi-finish milling - Drilling - Reaming - Semi-fine hinge - Finish milling
8	Rough milling - Semi-finish milling - Drilling - Semi-fine hinge - Finish hinge - Finish milling
9	Rough milling - Semi-finish milling - Drilling - Semi-fine hinge - Rough grinding hole - Finish milling
10	Rough turning - Semi-finish turning - Rough pulling - Finish pulling – Finish turning
11	Rough turning - Semi-finish turning - Rough grinding - Rough pulling - Finish pulling – Semi-finish grinding
12	Rough milling - Semi-finish milling - Rough pulling - Finish pulling – Finish milling

According to Table 1 and the generated initial process knowledge recommendation candidate set, the requirements semantic vector $V_r = \{w_{r1},d_1;w_{r2},d_2;\ldots;w_{r12},d_{12}\}$ and knowledge semantic vector $V_{k\text{-}i} = \{w_{r1},d'_1;w_{r2},d'_2;\ldots;w_{r12},d'_{12}\}$ are respectively established. For requirements semantic vector V_r, its weight coefficient represents the key degree of this characteristic. Under the condition that the knowledge attributes of the parts type, material and shape characteristics must meet the requirement, it is assumed that the machining accuracy, surface roughness and technical requirements are the focus of the process designer to pay attention to whether the requirements and knowledge match. The weight coefficients of the machining accuracy, surface roughness and technical requirements are set as 0.2, and the weight coefficients of the other parameterized requirements are set as 0.1.

After the semantic vector weight coefficient of each process knowledge $V_{k\text{-}i}$ in initial knowledge candidate set is calculated, the matching degree threshold $M_t = 0.7$ and quantity threshold $N_q = 6$ are set. By the matching degree calculation method based on improved cosine distance, the matching degree between requirements and each process knowledge is calculated. The results are shown as Fig. 3.

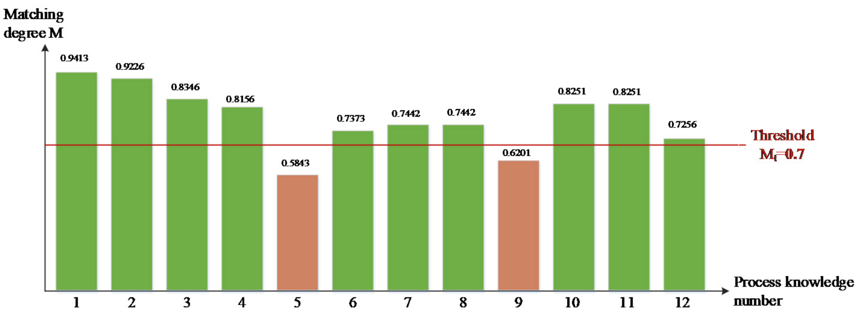

Fig. 3. Matching degree results between requirements and each process knowledge

Table 3. Information of final knowledge recommendation candidate set $K'(K_1,K_2,\ldots,K_6)$

Serial number	Process knowledge information	Matching degree value
1	Rough turning - Semi-finish turning - Drilling - Reaming - Semi-fine hinge - Finish turning	0.9413
2	Rough turning - Semi-finish turning - Drilling - Reaming - Semi-fine hinge - Rough grinding – Semi-finish grinding	0.9226
3	Rough turning - Semi-finish turning - Drilling - Semi-fine hinge - Finish hinge - Finish turning	0.8346
4	Rough turning - Semi-finish turning - Rough pulling - Finish pulling – Finish turning	0.8251
5	Rough turning - Semi-finish turning - Rough grinding - Rough pulling - Finish pulling – Semi-finish grinding	0.8251
6	Rough turning - Semi-finish turning - Drilling - Semi-fine hinge - Finish hinge - Rough grinding – Semi-finish grinding	0.8156

According to the matching degree threshold $M_t = 0.7$, 5th and 9th process knowledge are filtered out. Process knowledge numbered 6th, 7th, 8th and 12th process knowledge are filtered out according to the required process knowledge quantity threshold $N_q = 6$. The final process knowledge recommendation candidate set was obtained after reordering according to the matching degree from large to small, as shown in Table 3. After verification, the candidate set of process knowledge meets the machining requirements of the shaft-hole part.

7 Conclusion

In this paper, the "scene-label-classification" knowledge recommendation scheme for process knowledge graph is established to enable dynamic classification of process knowledge for different requirements scenarios. The multi-dimensional requirements for process knowledge recommendation are fully considered, and correlation mapping between requirements and process knowledge attributes was established by parameterized requirement characteristics and coding labels. The demand-knowledge semantic vector space model was constructed by taking parameterized requirement characteristics as dimensions of the space. The matching degree calculation method based on improved cosine distance is proposed, which considered both parameterized requirement characteristics and description content. A verification example with shaft-hole part showed that based on specific requirement scenarios, the proposed method achieved the process knowledge recommendation with strong pertinence and flexible number of knowledge candidates.

Acknowledgement. The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this paper: This work is supported by the National Key Research and Development Program (Grant 2021YFB1716200).

References

1. Geng, J., Zhang, S., Hui, W.: Knowledge Recommendation method for concept development of manufacturing technology using morphological similarity. Int. J. Softw. Eng. Knowl. Eng. **28**(08), 1121–1150 (2018)
2. Wu, Z., Li, L., Liu, H.: Process knowledge recommendation system for mechanical product design. IEEE Access PP(99), 1 (2020)
3. Hu, Y., Ding, Y., Xu, F., Liu, J., Xu, W., Feng, H.: Knowledge recommendation system for human-robot collaborative disassembly using knowledge graph. In: Proceedings of the ASME 2021 16th International Manufacturing Science and Engineering Conference, MSEC2021, vol. 2, pp. 1–6. ASME (2021)
4. Guo, Q., Xue, C., Yu, M., Shen, Z.: A New user implicit requirements process method oriented to product design. J. Comput. Inf. Sci. Eng. **19**(1), 011010.1–011010.11 (2019)
5. Kudryavtsev, D., Gavrilova, T.: From anarchy to system: a novel classification of visual knowledge codification techniques. Knowl. Process. Manag. **24**(01), 3–13 (2016)
6. Liu, D., Ray, G., Whinston, A.B.: The interaction between knowledge codification and knowledge-sharing networks. Inf. Syst. Res. **21**(04), 892–906 (2010)
7. Kadowaki, N., Kishida, K.: Empirical comparison of word similarity measures based on co-occurrence, context, and a vector space model. Korea Inst. Sci. Technol. Inf. **8**(02), 06–17 (2020)
8. Wen, Y., Gao, C., Pan, H.: Research on concept semantic similarity computation based on ontology. In: 2011 IEEE 2nd International Conference on Computing, Control and Industrial Engineering, vol. 2, pp. 284–287. IEEE, Wuhan (2011)
9. Wang, K., Liu, T., Tong, S., Zhang, X.: Research on matching of manufacturing knowledge and design requirements based on virus adsorption mechanism. Manuf. Automat. **36**(15), 116–120 (2014). (In Chinese)
10. Renu, R., Mocko, G.: Computing similarity of text-based assembly processes for knowledge retrieval and reuse. J. Manuf. Syst. **39**, 101–110 (2016)