In the product design stage, a large number of 3D CAD models are accumulated. This has led enterprises to pay increasing attention to the effective management and reuse of model resources in order to improve design efficiency and shorten the product life cycle. However, in the process of resource reuse, low efficiency is caused by the complexity of the assembly model structure, the designer's personal knowledge or inexperience, or unclear model function annotation. In addition, the high cost, the strong subjective nature, and the incompleteness of manual annotation affect the accuracy of model labeling, which decreases retrieval efficiency. To solve this problem, the automatic labeling of the model function has been a concern.

The automated annotation process uses the labeled models in a library to obtain the semantic information of an unknown model through a similarity comparison. Zhang et al. (2001) have proposed the active learning mechanism, which generates the semantic annotation of the model according to the geometric similarity with other models. Similar methods include the support vector machine (Ip and Regli, 2005; Wang et al., 2011) and the Bayesian recursive learning mechanism (Barutcuoglu and DeCoro, 2006). Leifman et al. (2005) used relevant feedback technology to record user intention during the model retrieval process, which alleviates the gap between geometry and semantics of the model to a certain extent. Zhang et al. (2010) and Li et al. (2013) presented the automatic labeling method, which uses the multiscale feature extraction method based on an attributed adjacency graph (AAG) and the local shape distribution histogram to obtain the shape information of the CAD model and calculate the shape similarity between different models. According to the known semantic classification information of the model in a sample database, a probabilistic annotation framework is constructed to make semantic annotation of the CAD model in order to establish a relationship functional semantics annotation of assembly model using the fusion of bag of relationships model and spectral technology

Li HAO*, Rong MO*, Zhoupeng HAN* and Binbin WEI**

* The Key Laboratory of Contemporary Design and Integrated Manufactory Technology, Ministry of education, Northwestern Polytechnical University, Xi’an 710072, China
E-mail: morong@nwpu.edu.cn

** School of Aeronautics, Northwestern Polytechnical University, Xi’an 710072, China

Received: 20 February 2019; Revised: 24 September 2019; Accepted: 12 November 2019

Abstract

Increasing attention has been paid to the effective management and reuse of CAD model resources. Aiming at the problems of low efficiency of model reuse and poor accuracy in the function labeling process of assembly models, this paper presents an effective probability-based model labeling strategy for complex assemblies. It is a proper method to realize automatic labeling of assembly functional semantics through active learning. Different from part model retrieval, an assembly model is described through graph theory and a bag-of-relationships model. The assembly relationships of assembly model and the information of key functional parts are considered synthetically. Then a two-tiered model retrieval mechanism is constructed to reduce the computation time cost and improve retrieval efficiency. Further, the concept of functional ontology is introduced to establish the normalized expression of the key functional semantics of the assembly model. The functional semantic annotation of the key parts of the CAD assembly model is carried out through a probability-based labeling framework to map the model shape structure to functional semantics, thus mitigating the “semantic gap” problem. Experimental results demonstrated that this method could improve the accuracy of functional semantic annotation of the assembly model, reduce the difficulty of labeling, and improve the reusability of 3D CAD models as a design resource.

Keywords: Automatic annotation, Key functional semantics, Assembly model labeling, CAD model retrieval, Model reuse

1. Introduction

In the product design stage, a large number of 3D CAD models are accumulated. This has led enterprises to pay increasing attention to the effective management and reuse of model resources in order to improve design efficiency and shorten the product life cycle. However, in the process of resource reuse, low efficiency is caused by the complexity of the assembly model structure, the designer's personal knowledge or inexperience, or unclear model function annotation. In addition, the high cost, the strong subjective nature, and the incompleteness of manual annotation affect the accuracy of model labeling, which decreases retrieval efficiency. To solve this problem, the automatic labeling of the model function has been a concern.

The automated annotation process uses the labeled models in a library to obtain the semantic information of an unknown model through a similarity comparison. Zhang et al. (2001) have proposed the active learning mechanism, which generates the semantic annotation of the model according to the geometric similarity with other models. Similar methods include the support vector machine (Ip and Regli, 2005; Wang et al., 2011) and the Bayesian recursive learning mechanism (Barutcuoglu and DeCoro, 2006). Leifman et al. (2005) used relevant feedback technology to record user intention during the model retrieval process, which alleviates the gap between geometry and semantics of the model to a certain extent. Zhang et al. (2010) and Li et al. (2013) presented the automatic labeling method, which uses the multiscale feature extraction method based on an attributed adjacency graph (AAG) and the local shape distribution histogram to obtain the shape information of the CAD model and calculate the shape similarity between different models. According to the known semantic classification information of the model in a sample database, a probabilistic annotation framework is constructed to make semantic annotation of the CAD model in order to establish a relationship
between the model’s shape information and its semantic information.

Using a machine-learning algorithm to classify the unknown model, the existing 3D model semantic annotation methods consider the whole or local geometric shapes and topology information of the part model through active learning of the labeled model. To a certain extent, this achieves the automatic annotation of semantics. However, it is difficult to infer the thinking activities of designers during the product design process through semantic annotation using only shape similarity comparisons. If a functional analysis of 3D CAD assembly model structure is carried out, functional information of the assembly model structure can be extracted. Analyzing and determining the functional semantics that can be realized in the structure of each part of the assembly model will help map the relationship between the high-level semantic features and the low-level features (such as geometry, materials, and dimensions). This helps to mitigate the problem of the semantic gap significantly.

Function is the core that embodies the value of the product and the 3D model is the direct carrier of the design intention, and also the source of product innovation design (Xu et al., 2009). Functional semantics, as a semantic-oriented knowledge, is not only an abstraction of shape structures, but also a high-level semantic expression of the potential knowledge of the design intention, usage, and efficacy of the model, which plays an important role in the efficient reuse of the model.

Wang et al. (2013, 2014) established a functional semantic ontology of 3D CAD models, and conducted an entire functional semantic annotation of a 3D model. The 3D model retrieval was based on functional semantics aimed to meet the requirements of conceptual product design. This method focused on the functional semantic annotation of the part model and lacked the functional analysis and labeling at different granularity levels of the assembly model structure.

Compared with simple shape semantic annotation, the functional semantic annotation of an assembly model is more complex and difficult. As the basis and key step towards functional semantic retrieval of 3D models, functional semantic annotation still confronts many challenges.

The objective of this paper is to propose an automatic annotation method for CAD model assemblies. The method consists of the following steps: based on the analysis of the key functional parts of the assembly, find the main parts with key functions. Then, extract the bag-of-relationships (BoR) model of the assembly and the spectral vectors of the key functional parts. Subsequently, transform the similarity comparison of the assembly model into the best matching process of the weighted complete bipartite graph and map the function semantics with the shape structure of assembly model. Finally, introduce the automatic learning mechanism to the automatic annotation of the model, and provide a good solution strategy for the functional semantic annotation of the assembly model. Experimental results showed that this method improved the accuracy of the functional labeling of the assembly model, thus reducing the difficulty of labeling, and improving the reusability of the model as a design resource.

2. Identification and representation of functional structure

An assembly, which is a group of machine parts joined with a certain relationship, is a stable structure. The parts of the assembly can be divided into two categories: functional parts and connectors. Functional parts are components that enable the product to perform certain functions. The connector is an assembly part used as an auxiliary to implementing the product function. The connection relationships between parts in an assembly can be divided into two categories: one is the contact connection relationship between the parts’ mating surfaces, and the other is the fastening joint relationship. A complex assembly may contain hundreds of connectors, but its main function is borne by some or a particular simple structure of its functional parts.

2.1 Representation of assembly model and its parts

Assemblies can be defined as a set of parts and their combinations (sub-assemblies), assembly constraints, and assembly relationships. The parts in the assembly model fully realize the function of the product by transferring motion through contact. In this study, the representation of an assembly model is divided into two layers: the assembly layer and the part layer.

2.1.1 Assembly layer representation

A 3D CAD assembly model P consists of N parts, \( P = \{ p_1, p_2, \ldots, p_N \} \), where there are complex assembly constraint relationships between the different parts. According to graph theory, a composing part can be mapped as a node in a graph. The attribute of the part itself is mapped to the node attribute, and the assembly constraint relation
between parts is mapped to the connection edges and edge attributes between nodes. Thus, the assembly model can be expressed as the attributed adjacency graph of the part (P-AAG) $G$ (Bondy, 1976)

$$G = \{V_p, E_p, VA, EA\}$$

where $V_p = \{p_1, p_2, \ldots, p_N\}$ is the node set that represents the parts, $E_p = \{(p_i, p_j) | p_i, p_j \in V_p\}$ indicates the connection relationships between nodes, that is, the assembly constraint relationships between parts. $VA$ and $EA$ represent the node properties and edge properties in the P-AAG, respectively. $EA$ is a set of the attributes of edges between two adjacent nodes that contains the assembly constraints information between parts in assembly. For $VA = (p_{type,i}, p_{num,i}, p_{ras,i})$, which is a three-tuple, each item represents a property. $p_{type,i}$ represents the part type, such as a function part or a connector. $p_{num,i}$ represents the number of assembly features of part $i$, such as keyway number. $p_{ras,i}$ indicates the relative surface area of part $i$ in the assembly; if the surface area of the part is larger, the relevancy of its role is higher in the assembly process. $p_{ras,i}$ can be calculated by Eq. (1)

$$p_{ras,i} = \frac{A_{ras,i}}{\sum_{j=1}^{N} A_p} = \frac{A_{ras,i}}{\sum_{j=1}^{N} A_p}$$

(1)

where $N$ is the total number of parts contained in the assembly and $A_p$ represents the surface area of part $i$.

### 2.1.2 Part layer representation

A CAD part model $p$ is surrounded by multiple faces, $p = (f_1, f_2, \ldots, f_m)$. According to graph theory, a part can be mapped as an attributed adjacency graph (F-AAG) of its corresponding faces, $G_f = (V_f, E_f, VA_f, EA_f)$. Where $V_f$ is a node set that corresponds to the part face. $E_f$ is the connection relationship between nodes, that is, the contact relationship between the faces. $VA_f$ is the property of the node in the F-AAG, such as the face type or face concavity. $EA_f$ represents the properties of the edges, that is, the concavity of the boundaries between faces.

In the F-AAG $G_f$, the connection relationship between the faces can be represented by the attributed adjacency matrix:

$$A_f = [a_{st}]_{M \times M}$$

where,

$$a_{st} = \begin{cases} 1 & f_s \text{ and } f_t \text{ are contact} \\ 0 & f_s \text{ and } f_t \text{ are not contact} \end{cases}$$

### 2.2 Recognition of key functional parts

Taking the geared oil pump shown in Fig. 1 as an example. According to the method proposed by Han et al (2017, 2018), the key functional parts are selected according to the comprehensive importance of the parts topological information and the multi-source attribute information. The former includes degree, density, and number of parts. The latter includes the assembly feature type and quantity, relative surface area, and contact face type. As a result of selection, the key functional parts of the geared oil pump contain the drive gear shaft, the right end cover, the left end cover and the pump body, as shown in Table 1.

![Fig. 1 The explosion diagram of a geared oil pump](image_url)
Table 1 Key function parts of the geared oil pump

| Node number | Part name       | Picture | Function                                           |
|-------------|-----------------|---------|---------------------------------------------------|
| v₃          | drive gear shaft|         | • transmitting-torque                             |
|             |                 |         | • transforming-motion                             |
|             |                 |         | • transmitting –motion                             |
|             |                 |         | • changing-velocity                               |
| v₇          | right end cover |         | • sealing-liquid                                  |
|             |                 |         | • transmitting-energy                             |
| v₁          | left end cover  |         | • sealing-liquid                                  |
| v₆          | pump body       |         | • storing-liquid                                  |

2.3 The descriptor of assembly model

According to the identification method proposed by Han et al (2017), assembly model is represented based on complex network, which is map to an attributed adjacency graph. Extracting and computing the topological information (degree centrality, closeness centrality and betweenness centrality) and multi-source attributes of parts, taking use of a two-level evaluation model to evaluate importance of parts assembled, and the key function parts in assembly can be obtained by sorting result. After the identification and extraction of key functional parts, the assembly model can be reduced to a set of these parts, which are sorted according to the relevancy of their role in the assembly (Han et al, 2017). Through graph theory and a BoR model, vector representation of the assembly and the key functional parts is carried out in order to convert the assembly model into spatial vectors.

2.3.1 Assembly-level descriptor based on bag-of-relationships

The assembly relationship is an important aspect of assembly information. It directly represents the function of the product, and the part structure can be constrained to realize the design intention. The assembly relationship contains positioning relationships, connections, and motion relationships between the parts, as shown in table 2.

Table 2 The classification of assembly relationship

| Assembly relationship | Classification                     |
|-----------------------|------------------------------------|
| Positioning           | • Plying-up                        |
|                       | • Align                            |
| Connection            | • Threaded connection              |
|                       | • Riveted connection               |
|                       | • Weld bonding                     |
|                       | • Interference fit                 |
| Motion                | • Screw drive                      |
|                       | • Gear drive                       |
|                       | • Planar motion                    |
|                       | • Key drive                        |
|                       | • Chain drive                      |
|                       | • Rotating motion                  |
Bag-of-words (BoW) model first appeared in the field of natural language processing and information retrieval. In a BoW model, the grammar and sequence of text are ignored and the text is considered as a set of words unordered. The text or document could be described based on the statistics of words involved. An instance is shown in Fig.2, there are two texts, which were consisted of several words. Firstly, taking use of clustering method to build vocabulary dictionary with ten words. Then the two texts could be represented by two sets with unordered words. According to the vocabulary dictionary, vectors and histograms of two texts are obtained based on word frequency statistics, which could use for further computation.

Fig.2 An instance of bag-of-word model

Likewise, an assembly model is regarded as text; each connection relationship of the assembly is treated as a “word” (shown in table 3). Then, the bag-of-relationships (BoR) model is build, and dictionary D with connection relationships \( r_i \) as elements can be constructed for the assembly model, \( D = \{ r_1, r_2, \ldots, r_{\text{Num}} \} \), where Num is the number of assembly relationships.

| Text | Word | Vocabulary Dictionary | Corpus |
|------|------|------------------------|--------|
| Assembly model | Assembled relationship | Relationship Dictionary | Assembly model library |

The core concept of BoR is to treat the assembly model as a set of disordered keywords. The model is represented by the frequency of keywords appearing in the model. Actually, a keyword corresponds to an assembly relationship contained in each assembly. Because each assembly model contains a limited variety of assembly relationships, and the number of them can also be used as a parameter for differentiation, all assembly relationship numbers are counted as statistical objects when building spatial vectors. The specific steps are as follows:

1. Extract the assembly relationships of all assembly models in the model library;
2. According to all the extracted assembly relationships, construct a vocabulary dictionary \( \text{Dic} \). All assembly relationships are contained in the dictionary and have a unique index independent of the order;
3. Generate a histogram of the assembly relationships of each assembly based on the dictionary. The horizontal coordinate corresponds to the types of assembly relations, and the vertical coordinate to their frequency.

The difference between the BoR and BoW models is that the latter can distinguish different categories through “words”. However, in the former, for different kinds of assembly models, the type of assembly relationships contained is slightly different; therefore, more precise model descriptors are needed.

2.3.2 Part-level descriptor based on spectral vectors

In graph theory, algebraic properties of various graph matrices (such as adjacency matrix, Laplacian matrix, etc.) can be used to reflect the properties of the graph. Especially, the graph spectral vector is an ordered sequence of the eigenvalues of the graph adjacency matrix, which can describe the topological structure of the graph in the form of vectors (West, 2000). In this study, Laplacian matrices were used as the key function part surface attributed adjacency matrices of the assembly model to describe the topological properties of the identified functional parts. Laplacian matrix \( L \) can be calculated by Formula (2):
The specific steps are as follows:

(1) The first tier is a coarse-grained fast retrieval based on the BoR model in which the corresponding BoR model is constructed by extracting the assembly relationship between the assembly model to be compared and the assembly model in the model library. By calculating the distance between the BoR model of the assembly model to be compared...
with that of the assembly model in the model library, a top models are obtained, which constitutes the candidate set of the assembly model. The fast comparison between the assembly models $P$ and $Q$ employs the following distance formulas:

$$S_{BoR}(P, Q) = \min\left(d(P, Q)\right) = \min\left(\|BoR(P) - BoR(Q)\|_2\right)$$

(3)

(2) The second tier consists of accurate retrieval based on spectral vectors in which fine-grained matching of the assembly models in the candidate sets is performed. Firstly, the key functional parts contained in each assembly model are recognized. Secondly, the key functional parts are transformed into the F-AAG, and the Laplacian spectral vector of F-AAG is calculated. Thirdly, using each key functional part contained in the assembly model as a node, the spectral vector acts as a node attribute, so that the fine matching of models $P$ and $Q$ can be converted into the optimal matching process of the complete binary graph. Thus, the matching model based on the weighted complete bipartite graph is built, as shown in Fig.4:

![Weighted bipartite graph of functional parts matching of the assembly model](image)

According to Eq. (4), the similarity comparison of key functional parts contained in the assembly model can be transformed into the weight calculation of the complete bipartite graph.

$$\omega_{ij} = \text{dist}(Pad\left(\text{SpV}(P_i)\right), Pad\left(\text{SpV}(Q_j)\right)) = \|\text{Pad}\left(\text{SpV}(P_i)\right) - \text{Pad}\left(\text{SpV}(Q_j)\right)\|_2$$

(4)

where, $\text{dist}(\cdot, \cdot)$ represents the distance between two eigenvectors, that is, the individual distance comparison between the parts of the two assembly models. $\omega_{ij}$ indicates the similarity between the functional part $i$ in the assembly model $P$ and the functional part $j$ in model $Q$, that is, the weight between any two nodes of the complete binary graph.

The best match between all the key functional parts contained in model $P$ and all the key functional parts contained in model $Q$ is found. Thus, the similarity comparison problem can be transformed into an optimal matching problem of the complete binary graph, and the Kuhn-Munkres algorithm is used to calculate the similarity $S_f(P, Q)$ between the assembly models $P$ and $Q$ (West, 2000).

The similarity comparison process of the assembly model is more complicated than the similarity comparison process of the part model. Combined with the results of fast retrieval and accurate matching, the comprehensive similarity evaluation of the assembly model is carried out according to Eq. (5), thus increasing the accuracy of the retrieval results.

$$S_d(P, Q) = \beta_1 S_{BoR}(P, Q) + \beta_2 S_f(P, Q)$$

(5)

where $S_d(P, Q)$ is the comprehensive similarity between assembly models $P$ and $Q$, and $\beta_1$ and $\beta_2$ are weights obtained by experiments.

3. Construction of functional semantic ontology

Ontology is used to standardize descriptions of concepts, terms, and their interrelationships, and offers a language that describes or expresses knowledge in a field (Uschold and Gruninger, 1996). Its inference rules are not only used for the organization and management of 3D model knowledge, but also play an important role in the extraction of advanced semantic knowledge of 3D models and semantic matching. In the process of product design, since designer's knowledge, experience, and design habits are different, the semantic expression of the model has some problems such as arbitrariness and semantic ambiguity, which affects the efficiency of model retrieval and reuse. Therefore, the
The establishment of the functional semantic ontology of the 3D model can standardize the knowledge expression of the model. Meanwhile, the functional terms have a unified and fixed form of expression, which could reduce ambiguity and inconsistent expressions, thus improving the reusability of the model.

The description of functional semantics needs to be expressed in the format of standardized terms and specifications. First, we should clearly define the concept categories and attribute relationships between categories in this field. This study uses a product function definition for “verb-object” (Stone and Wood, 2000). Here, “verb” is a verb word in functional classification and terminology vocabulary, and “object” is a concrete or flowing object of the 3D assembly model. (show in Fig.5)

Fig.5 Functional semantic classification

Constructing functional ontology knowledge and attaching functional semantics to the model can make low shape descriptors become semantic knowledge representations at a high levels, which is used for functional annotation to facilitate knowledge reuse.

4. Functional semantic annotation

In this study, through the analysis and learning of the labeled models in a small number of sample libraries, an automatic learning mechanism was constructed as shown in Fig.6. It contains two phases: (1) a small number of sample models are marked with functional semantics and become labeled models. (2) Extracting the feature of the sample model and the unknown model, which is used to similarity comparison. Taking use of the conditional probability method to label the most similar local parts between sample model and the unknown model. Then, the unknown model could be stored as labeled model, while achieving the purpose of extending the model library. That is to say, the functional semantic annotation of a large number of unknown assembly models was realized by using the probabilistic semantic annotation method.

Fig.6 Functional semantic labeling mechanism

The functional semantic annotation of an assembly model depends on the functional semantic annotation of its functional parts. Based on the key functional parts, we analyzed and extracted the functions performed by the functional parts, and labeled the corresponding functional semantics in order to generate the functional semantic annotation of the assembly model. Similar parts have different functions in different assemblies, so it is chiefly necessary to distinguish the assembly model categories in which the functional parts are located during the labeling process.
In this study, some known assembly model, including parts with key functions, were used to label the whole and local semantics of a large number of unknown assembly models automatically.

4.1 Automatic annotation based on conditional probability

In order to establish the relationship between the key functional structure of the assembly and the functional semantics, the functional semantics were labeled by extracting similar sub-assembly structures or parts. In this study, we first selected a small number of representative assembly models from the model library as a sample set, and built a framework of automatic labeling based on probability.

Suppose the sample set of the assembly model is \( T = \{ T_1^{K_1}, T_2^{K_2}, \ldots, T_n^{K_n} \} \), where \( T_i^{K_i} \) represents the \( i \)-th class of the models, such as the mechanical arm class, the gear pump class, the reducer class assembly model, and et al. The scale of each class is \( K_i \). The set of known categories in the library corresponds to the model sample set and is recorded as \( C = \{ c_1, c_2, \ldots, c_n \} \). The functional semantic label set of the assembly model is \( L = \{ l_1^{K_1}, l_2^{K_2}, \ldots, l_n^{K_n} \} \), \( l_i^{K_i} \) is a set of functional semantic labels contained in the corresponding \( i \)-th class model, and \( \Omega_i \) is the number of functional label sizes.

A labeling process based on conditional probabilities is built as follows:

1. First, the probability of the model to be labeled as the \( i \)-th class model is calculated, namely, the probability of the \( m^{*} \) model is marked as the \( i \)-th class model:

\[
p(c_i | m^{*},T_i^{K_i}) = \frac{1}{K} \sum_{i=1}^{K} Sd(m^{*},T_i^{K_i})
\]

where \( Sd(m^{*},T_i^{K_i}) \) is the similarity value between model \( m^{*} \) and each model of the \( i \)-th class in model library \( T \) in turn.

Considering the influence of the whole sample set model on the results, this method is used to represent the similarity probability between the models in order to eliminate the randomness of the similarity comparison of a single model.

2. After obtaining the class label of the assembly model \( m^{*} \), the assembly model with the maximum similarity \( T_i^{*} \) is obtained by comparing model \( m^{*} \) with the \( i \)-th class model in the known sample model library. If the function label corresponding to the model \( T_i^{*} \) is \( l_i^{K_i} \), the matching results between the key functional parts of the assembly model to be marked are calculated, as well as those of model \( T_i^{*} \) with maximum similarity. Once the matching result is determined, the probability for the part of the model to be marked, labeled as \( l_i^{*} \), is:

\[
p(l_i^{*} | m_i^{*},T_i^{*}) = Sd(m_i^{*},T_i^{*}) p(l_i^{*},T_i^{*})
\]

3. The probability that the model \( m^{*} \) is labeled with a functional label \( l_i^{*} \) is calculated:

\[
p(l_i^{*},m^{*}) = p(m^{*},T_i^{*}) p(l_i^{*} | m_i^{*},T_i^{*}) = \max \left( Sd(m^{*},T_i^{K_i}) \cdot Sd(m_i^{*},T_i^{*}) \cdot p(l_i^{*},T_i^{*}) \right)
\]

4.2 Automatic screening of matching results based on k-near neighbor algorithm

The k-near neighbor algorithm (k Nearest Neighbor-KNN) is a simple machine learning method proposed by Cover and Hart (1968). The steps to screen the results using the KNN algorithm are as follows:

1. Calculate the distance between each model in the model library with known class and the model to be tagged;
2. Sort the calculated distances of step (1) in ascending order;
3. Select the \( \alpha \) top model with the smallest distance;
4. Determine the frequency of occurrence of the class \( k \) of the \( \alpha \) top model;
5. Return the class \( k \) with the highest frequency of the \( \alpha \) top model as the class label of the model to be labeled;
6. Select the model with maximum similarity in class \( k \) and obtain its functional semantics \( l_i^{*} \). According to the matching results of local similarity comparison, the model is labeled.

5. Results and discussions

5.1 Example validation

In order to verify the method proposed in this paper and the automatic annotation results of a CAD model, we constructed a model library containing three hundred 3D CAD assembly models; most of them downloaded from the website, and some from the team project. Each model in the model library was divided into 6 categories, including the
pump class, mechanical arm class, and the reducer class. Moreover, 70% of the models of each class were chosen as the sample set, and the rest as a test set. As shown in table 4:

| Classes         | Models                                      |
|-----------------|---------------------------------------------|
| Reducer         | ![Reducer Models](image1)                   |
| Gear oil Pump   | ![Gear oil Pump Models](image2)             |
| Centrifugal pump| ![Centrifugal pump Models](image3)          |
| Air cylinder    | ![Air cylinder Models](image4)              |
| Lifter          | ![Lifter Models](image5)                    |
| Mechanical arm  | ![Mechanical arm Models](image6)            |

Taking the gear pumps shown in Fig. 3 as an example, Model P is the model to be labeled, and model Q is the model in the model library. The model functional semantic labeling process is performed as follows:

Step 1: Extracting and calculating the BoR model of the assembly model.

The assembly relationships of model P and the model in the library are extracted, and the assembly relationship dictionary \( D_{rel} \) is constructed.

\( D_{rel} = \{ \text{threaded connection}, \text{riveted connection}, \text{pin connection}, \text{key connection}, \text{coupling joint}, \text{welding}, \text{bonding} \} \)

The assembly relationship vectors \( V_P \) and \( V_Q \) of models P and Q are respectively extracted. Then, the distance between the two vectors is calculated according to Eq. (3). European distance was used in this study.

\[
V_P = (14, 0, 3, 2, 0, 0, 0), \quad V_Q = (12, 0, 3, 2, 0, 0, 0)
\]

Step 2: Extracting the key functional parts of the assembly model.

The key functional parts of the model are extracted and the assembly model is transformed into a set of local parts, as described by Han (2017),

\[
P = \{ p_1, p_2, p_3, p_4 \}, \quad Q = \{ \text{drive gear shaft}, \text{right end cover}, \text{left end cover}, \text{pump body} \}
\]

Step 3: Calculating the Laplacian spectral vector.

According to Eq. (3), the Laplacian spectral vectors of models P and Q are represented as follows:
Then, the distance between each part of the two models is calculated, and the weight matrix $\omega$ is obtained as follows:

\[
\omega = \begin{bmatrix}
1.46 & 4.23 & 4.98 & 1.46 \\
4.15 & 1.33 & 2.29 & 3.08 \\
4.04 & 1.58 & 2.48 & 2.93 \\
2.21 & 4.23 & 4.96 & 1.16 \\
\end{bmatrix}
\]

Step 4: Similarity comparison of the assembly model.

The best match of the complete binary graph is constructed, and the best matching results between the parts of the model are calculated by using the Kuhn-Munker method as follows,

\[
R = \{(p_i, \text{drive\ gear\ shaft}, 0.8795), (p_i, \text{right\ end\ cover}, 0.8772), (p_i, \text{left\ end\ cover}, 0.7748), (p_i, \text{pump body}, 0.9073)\}
\]

\[
Sd(P, Q) = \omega_1 S_{\text{abcd}}(P, Q) + \omega_2 S_f(P, Q) = 0.863
\]

where $\omega_1 = 0.2$, $\omega_2 = 0.8$.

Step 5: Results annotation.

According to Eqs. (4)-(6), the functional semantics and category annotation results of the model are obtained as follows:

| Model                  | Key functional part          | Category label and its probability | Functional semantic and its probability         |
|------------------------|------------------------------|------------------------------------|------------------------------------------------|
| Gear pumps [0.863]     | Drive gear shaft [0.8795]    | • transmitting-torque [0.7590]     |                                                 |
|                        |                              | • transforming-motion [0.7590]      |                                                 |
|                        | Left end cover [0.8772]      | • sealing-liquid [0.7570]          |                                                 |
|                        | Right end cover [0.7748]     | • sealing-liquid [0.6687]          |                                                 |
|                        | Pump body [0.9073]           | • transmitting-energy [0.783]      | • support-part [0.783]                         |

5.2 Automatic annotation results

Table 6 shows the annotation results of the six categories of the assembly model in the test set using this method.
Table 6 The annotation results of the typical assembly model

| Model     | Key functional part | Category label and its probability | Functional semantic and its probability |
|-----------|---------------------|-----------------------------------|----------------------------------------|
| Reducer   | Tank cover          | [0.8711]                          | • sealing-parts [0.7260]                |
|           | [0.8640]            |                                   | • fix-parts [0.7201]                    |
|           | Base                | [0.7201]                          | • support-parts [0.7201]                |
|           | Main large gear shaft | [0.8840]  | • transmitting-torque [0.7367]         |
|           | Main small gear shaft | [0.8817]  | • transmitting-motion [0.7367]         |
|           | Main small gear shaft | [0.8705]  | • changing-velocity [0.7348]           |
|           | Piston rod          | [0.7929]                          | • transmitting-energy [0.6053]          |
|           | Stator              | [0.7741]                          | • storing-gas [0.5909]                  |
|           | Drive gear shaft    | [1.0000]                          | • transmitting-velocity [0.8485]        |
|           |                    |                                   | • transforming-motion [0.8485]          |
|           |                    |                                   | • transmitting-motion [0.8485]          |
|           |                    |                                   | • changing-liquid [0.8485]              |
| Gear pumps| Right end cover     | [1.0000]                          | • sealing-liquid [0.8485]               |
|           | Left end cover      | [1.0000]                          | • sealing-liquid [0.8485]               |
|           | Pump body           | 1.0000                            | • transmitting-energy [0.8485]          |
|           |                    |                                   | • supporting-parts [0.8485]             |
|           | Driven gear shaft   | [1.0000]                          | • transmitting-torque [0.8485]          |
|           |                    |                                   | • transmitting-motion [0.8485]          |

The classification label and functional semantic label of the assembly model and their probabilities are listed below the model to be labeled. From the results, we can see that this method can effectively classify assemblies, and can
provide designers with the main functional semantics to help them better reuse the models.

5.3 Evaluation of retrieval and annotation

In order to verify the effectiveness of this algorithm in 3D CAD assembly model labeling, a recall-accuracy (recall-precision) curve was used as the evaluation criterion in this study. The method proposed by Li (2013), a multiscale feature characterization method based on local shape distribution is used to retrieve and annotate the model. The algorithm proposed in this paper takes into account the assembly relationship, geometric information, and functional semantics of parts (as shown in Fig.7). Its accuracy of retrieval and annotation is better than that of the multiscale shape distribution algorithm proposed by Li (2013). The retrieval and annotation accuracies of the 3D CAD assembly model have been further improved.

![Fig.7 The comparison results of algorithm's recall-accuracy curve](image)

6. Conclusion

In order to mitigate the problem of low efficiency during manual functional labeling of 3D CAD assembly models, this paper proposes a probability-based method of automatic annotation of key functional semantics. First, the feature characterization is carried out for assembly topology and part attributes, and a model descriptor based on a BoR model and the part spectral vector is constructed. The concept of functional ontology is introduced to establish the normalized expression of the key functional semantics of the assembly model. Additionally, the accuracy of assembly model retrieval is improved by the two-tiered similarity comparison mechanism. The automatic probability-based annotation process was established so that the assembly model to be labeled obtained simultaneously the category label and the function semantic label. Mapping the model shape structure with functional semantics offers an efficient method to mitigate the “semantic gap” problem. By analyzing and learning the labeled models in a small sample database, an automatic learning mechanism was constructed, and the functional semantic migration labeling of a large number of unknown assembly models was realized using the probability-based semantic annotation method. The experimental results show that:

1. This paper provides a reasonable assembly annotation strategy.
2. To a certain extent, the “semantic gap” problems existing between the structure and function of 3D CAD assembly models are alleviated.
3. The annotation results obtained by using this method for assembly model labeling are more accurate than previous reported methods.

Acknowledgements

We thank the anonymous reviewers for their constructive comments. The support from the National Natural Science Foundation of China (51775445), and the Natural Science Basic Research Plan in the Shaanxi Province of China (Program No. 2016JM5040) toward to this research is gratefully acknowledged.
Reference

Barutcuoglu, Z. and DeCoro, C., Hierarchical Shape Classification Using Bayesian Aggregation. International Conference on Shape Modeling and Applications (2006), pp.44-50.

Bondy, J. A., Graph theory with applications, Journal of the Operational Research Society, Vol.28, No.1 (1977), pp.237-238.

Cover, T. and Hart, P., Nearest Neighbor Pattern Classification, IEEE Transactions on Information Theory, Vol.13, No.1 (1967), pp. 21-27.

Han, Z., Mo, R., Chang, Z., Hao, L. and Niu, W., Key assembly structure identification in complex mechanical assembly based on multi-source information, Assembly Automation, 2017, Vol.37, No.2 (2017), pp. 208-218.

Han, Z., Mo, R., Yang, H. and Hao, L., Module partition for mechanical CAD assembly model based on multi-source correlation information and community detection, Journal of Advanced Mechanical Design Systems & Manufacturing Vol.12, No.1(2018), pp.JAMDSM0023-JAMDSM0023.

Ip, C. Y. and Regli, W. C., Manufacturing Classification of CAD Models Using Curvature and SVMs, International Conference on Shape Modeling and Applications. IEEE (2005), pp.361-365.

Leifman, G., Meir, R. and Tal, A., Semantic-oriented 3d shape retrieval using relevance feedback, Visual Computer, Vol.21, No.8-10 (2005), pp.865-875.

Li, L., Zhang, S. and Bai, X. and Huang, R., Auto-tagging algorithm of 3D CAD models, Computer Integrated Manufacturing Systems, Vol.19, No.7 (2013), pp.1484-1489.

McWherter, D., Peabody, M. and Regli, W.C., An approach to indexing databases of solid models, Technical Report DU-MCS-01-02, Philadelphia (2001), PA: Drexel University.

Stone, R. B. and Wood, K. L., Development of a functional basis for design, Journal of Mechanical design, Vol.122, No.4(2000), pp.359-370.

Uschold, M. and Grueninger, M., Ontologies: Principles, methods and applications, The Knowledge Engineering Review, Vol.11, No.2 (1996), pp.93-136.

Wang, X., Qin Z., Zhuang, C. and Zhang, X., 3D models retrieval method based on collaboration semantic annotation of discrepant features, Journal of Computer-Aided Design & Computer Graphics, Vol.23, No.1 (2011), pp:152-160.

Wang, Z., and Tian, L., Function-based 3D model retrieval system, Journal of Computer-Aided Design and Computer Graphics, Vol. 25, No.12 (2013), pp.1877-1885.

Wang, Z., Tian, L. and Duan, W., Annotation and Retrieval System of CAD Models Based on Functional Semantics, Chinese Journal of Mechanical Engineering, Vol. 27, No.6 (2014), pp.1112-1124.

West, D. B., Introduction to graph theory, 2nd ed. Upper Saddle River, N. J., USA: Prentice Hall (2000).

Xu, R., Tang, D., Zhu, R., Chen, X. and Zhang, T., Research on re-creative design strategy based on function ontology, Journal of Engineering Design, Vol.16, No.6(2009), pp.395-394.

Zhang, C. and Chen, T., Indexing and Retrieval of 3D Models Aided by Active Learning. ACM Multimedia (2001), pp. 615-616.

Zhang, X., Mo, R., Gong, Z. and Lian, X., Automatic semantic tagging of CAD models for retrieval, Journal of Computer-Aided Design & Computer Graphics, 2010, Vol.22, No.12 (2010), pp.2162-2167.