Hindawi
International Journal of Aerospace Engineering
Volume 2022, Article ID 2583437, 17 pages
https://doi.org/10.1155/2022/2583437

Research Article

Identification of Key Factors and Mining of Association Relations in Complex Product Assembly Process

Huibin Zhang,[1] Mingwei Wang ,[1] Wei Deng,[1] Jingtao Zhou,[1] Lei Liu,[1] Jianfeng Li,[2] and Rong Li[2]

[1]*Key Laboratory of High Performance Manufacturing for Aero Engine (Northwestern Polytechnical University), Ministry of Industry and Information Technology, Xi'an 710072, China*
[2]*AECC Xi'an Aero-Engine Ltd., Xi'an 710021, China*

Correspondence should be addressed to Mingwei Wang; 1044439925@qq.com

Received 23 May 2022; Revised 2 August 2022; Accepted 4 August 2022; Published 23 August 2022

Academic Editor: Xingling Shao

There are many factors involved in the assembly process of complex products, but only a few of the key influencing factors really affect the performance of the product assembly. In order to ensure the performance of complex products, it is necessary to model the deviation transfer flow of the assembly process of complex products and identify key processes and characteristics and determine its scope of influence. This paper establishes the set of assembly performance influencing factors by analyzing the complex product assembly process and uses complex networks combined with entropy weight method to identify the influencing factors of product performance to achieve the identification of key influencing factors. Based on the key factor identification results using an Apriori algorithm to mine the association rules between key factors and performance to provide guidance for the assembly of complex products and achieve transparent and controllable performance. The assembly process of a certain type of aeroengine low-pressure fan rotor is verified as the research object. The results show that the proposed method can identify the key factors in the multistage assembly process and establish an association rule library to provide decision support for process adjustment.

1. Introduction

The assembly of complex products requires multiple stages of assembly, and there is a certain amount of assembly deviation on each stage. The errors in each assembly process are different in size and type (part manufacturing errors, positioning errors during assembly). Various errors are continuously transmitted, accumulated, and evolved with the advancement of assembly, resulting in the final assembly performance not meeting the requirements. By modeling the deviation flow transfer of the assembly process of complex products, we identify the key influencing factors and determine their influence range, which is a guide to adjust the influencing factors of the assembly performance and ensure the product assembly performance.

The assembly process of complex products includes not only multiple types of components but also the use of various types of equipment such as tools and fixtures. And the assembly process is more complex; assembly process chain is long. These factors are constantly changing and coupled with each other throughout the assembly process, which leads to the following difficulties in identifying and correlating key factors of complex products in the assembly process.

(1) There are many influencing factors, the factors are coupled, and the influencing mechanism is complex and unclear. Product assembly performance is affected by the quality of parts processing, assembly process, assembly quality, and other factors, involving many influencing factors. These influencing factors are dynamically changing and coupled with each other, and it is difficult to determine their influence range by mechanism analysis methods under the synergistic effect of multiple factors

(2) The accumulation of influencing factors has a cross-process effect. The assembly process of complex

products is generally divided into multiple processes. The errors generated by the current process will be passed to the next process, which has a certain cumulative effect

To address the above difficulties of key factor identification and association analysis mining of complex products, this paper proposes a key factor identification and association mining method based on the combination of complex network, entropy method, and Apriori method. Establish the set of assembly performance influencing factors by analyzing the complex product assembly process, and establish the key factor identification model by using complex network combined with entropy power method to identify the influencing factors of product performance and obtain the key influencing factors in the complex product assembly process. Based on the key factor identification results using the Apriori algorithm, we mine the association rules between key factors and performance, to achieve accurate regulation of the assembly process and provide decision support for subsequent assembly process adjustment.

This paper is divided into 6 subsections. The first part is the introduction, which introduces the complexity and difficulty of the research problem. The second part is a literature review to introduce the current research progress in key factor identification and association mining. The third part is a complex product key factor identification and association mining method, which introduces the details of the method proposed in this paper. The fourth part is an example analysis to validate the application of the method proposed in this paper in the low pressure rotor assembly process of an aeroengine. The fifth part is the conclusion. The sixth part is the references.

2. Literature Review

2.1. Key Factor Identification. In order to achieve efficient control of product performance, it is necessary to identify key influencing factors. At present, the commonly used methods of identifying key factors mainly include loss function method, risk analysis method, fuzzy theory, Bayesian network, complex network, and other methods.

Tang et al. [1] used the assembly directed graph to establish a candidate set of key characteristics and then calculated the influence degree of the candidate characteristics on the key characteristics of the upper layer based on the Taguchi quality loss method and then defined the key characteristics of the layer according to the degree of influence. Zhao et al. [2] based on the qualitative preidentification of key characteristics based on the cumulative relationship of error transmission and realized the identification of key characteristics through risk analysis. Xu et al. [3] studied the quality factor identification method based on Bayesian network. This method uses Bayesian inference to identify the key quality factors based on the product quality relationship of the Bayesian network. Whitney [4] analyzed the influencing factors of key characteristics through the combination of characteristic decomposition, tolerance analysis, and assembly sequence. Guo [5] uses Taguchi quality loss function to determine the quality loss caused by the coordination elements between assembly levels and then uses fuzzy theory to calculate the influence degree of each element to realize the identification of product elements. Raman et al. [6] used hypergraphs and rough sets to reduce attributes to form an optimal attribute subset and realize the selection of important attributes. Yang [7] and Zheng et al. [8] mentioned key quality characteristic identification methods based on risk analysis, quality loss function, principal component analysis, and historical data analysis. Zhong et al. [9] proposed a multiattribute fusion method that combines the topological attributes and diffusion attributes of nodes to adaptively obtain the ranking results, including two fusion methods based on attribute union (FU) and attribute-based ranking (FR). Based on principal component analysis (PCA), Jin et al. [10] proposed a new algorithm for calculating the importance of nodes in complex networks, combining attributes such as degree centrality, tight centrality, and eigenvector centrality. Zhao et al. [11] used the TOPSIS method to integrate degree centrality, mesoscopic centrality, aggregation coefficient, and proximity centrality to rank the importance of nodes and then delete the most important nodes to obtain the ranking results of node importance.

The above methods have achieved some results in the identification of key characteristics, but their identification mainly depends on the decomposition and transmission of various factors. The relationship between various factors is relatively clear, or more a priori knowledge is required to realize the identification of key factors. It is difficult to identify key characteristics for the problems of unclear action relationship and small number of factors. On the basis of clarifying nodes and node relationships, this paper realizes the identification of key nodes in complex networks by constructing an association relationship network between various characteristics. The identification process does not require experimental data, nor does it need to clarify the mechanism of action and transfer relationship among various factors.

2.2. Association Mining. The increase in the number of the quality of complex products and process feature quantities will increase the complexity of the association rule algorithm. In fact, the performance of complex products may only have a direct relationship with certain feature quantities, but the relationship with other feature quantities is not obvious. Commonly used association rule analysis algorithms include Apriori algorithm [12] and FP tree frequent itemset algorithm [13]. Guo et al. [14] introduced cluster analysis theory, modeled photovoltaic processing from the perspective of data mining, and used it to evaluate the reliability of photovoltaic power generation systems. Li et al. [15] proposed a new network loss assessment method based on hybrid cluster analysis using data mining and typical scenario simulation ideas. Li et al. [16] developed a set of equipment fault information management and analysis system using data mining technology to provide a basis for realizing the status maintenance of relay protection devices and provide decision support for the analysis and processing of power grid faults. Zhang et al. [17] proposed a data mining and analysis method for secondary equipment defects based on the Apriori algorithm based on the defect data of

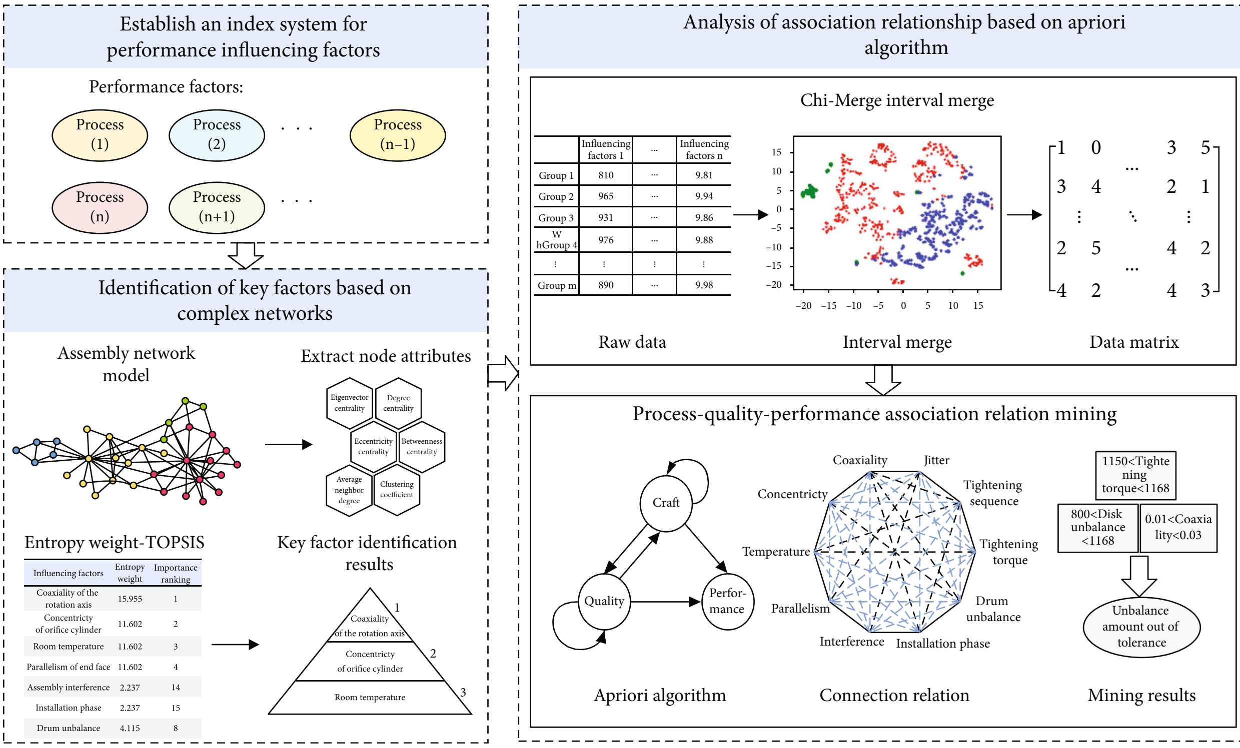

FIGURE 1: Method for key factor identification and association mining of complex products.

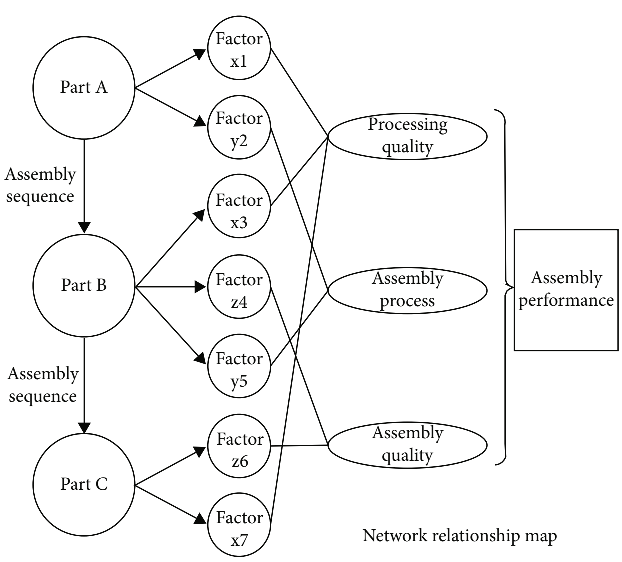

FIGURE 2: Complex relational network map.

secondary equipment, which improved the operation, maintenance, and control level of secondary equipment in the power system. Chen and Cao [18] proposed a multilayer association rule data mining algorithm and applied it to the analysis of massive data in commercial banking systems. Jia et al. [19] used the FP-Growth association rule mining algorithm to propose an association analysis for bird strikes in the aviation field and find out the association rules between the key inducing factors of civil aviation bird strikes.

The above-mentioned research has conducted certain research on the mining of association relations, but the above methods are all for the mining of association relations for discrete text data and cannot be mined for the multidimensional continuous data existing in the assembly process of complex products. Therefore, based on the identification of key factors, this paper uses the discretization algorithm to discretize the multidimensional continuous data, uses the association rule analysis method to mine the out of tolerance information data of complex product performance, analyzes the reliability between product performance and feature quantity, and reveals the correlation between product performance and feature. And one more advantage after discretization is that the distribution intervals of features can be obtained, and in the obtained association rules, it is a combination of multiple feature value intervals, which can provide clear data support for optimization adjustment.

3. Key Factor Identification and Association Mining of Complex Products

Firstly, this paper identifies all the factors affecting assembly performance in the assembly process of complex products, classifies the influencing factors, and establishes a product performance influencing factor index system. Then abstract the influencing factors as nodes and the assembly relationships as edges to build a complex network model. The properties of all factors are calculated based on the constructed complex network, and the Entropy weight-TOPSIS model is used to sort the influencing factors, in order to realize the identification of the key factors. Finally, based on the key factor identification results, the relevant data of the key

Table 1: Network characteristics and calculation methods of complex networks.

Network characteristics	Definition	Calculation method
Degree centrality	The degree of relevance of a node to other nodes	$C_D(N_i) = \sum_{J=1}^{g} x_{ij}$
Aggregation coefficient	Error propagation effect between nodes	$c_i = \frac{\sum_{r=1}^{k}\sum_{s=1}^{k} d(n_r, n_s)}{k(k-1)}$
Mesoscopic centrality	The number of the shortest paths through a node	$g(v) = \sum_{s\neq v\neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$
Proximity centrality	The proximity of a node to other nodes	$C(x) = \frac{1}{\sum_y d(y, x)}$
Centrifugal centrality	The distance from a given starting node to the farthest node	$e(i) = \max_{0\leq j\leq n} d_{ij}$
Eigenvector centrality	Number and importance of neighbor nodes	$\mathrm{EC}(i) = c\sum_{j=1}^{n} a_{ij}x_j$
Average neighbor degree	The average of the degrees of all nodes	$k_{nn,i} = \frac{1}{\lvert N(i)\rvert} \sum_{j\in N_{(i)}} k_j$

Table 2: Characteristic scale after multivalue discretization.

Feature amount	$F_{\mathrm{CM}}(x_i)$	Z^* element
x_1	0	z_1^{1*}
	1	z_2^{1*}
	2	z_3^{1*}
	⋮	⋮
⋮	⋮	⋮
x_n	0	z_1^{n*}
	1	z_2^{n*}
	⋮	⋮

Table 3: Database Z.

Label TID	Items
Qualified	$z_1^{1*}, z_2^{2*}, z_1^{3*}, z_2^{4*}, z_3^{5*}$
Failure	$z_1^{1*}, z_1^{2*}, z_2^{3*}, z_2^{4*}, z_1^{5*}$
Failure	$z_1^{1*}, z_2^{2*}, z_2^{3*}, z_2^{4*}, z_1^{5*}$
Qualified	$z_3^{1*}, z_1^{2*}, z_1^{3*}, z_2^{4*}, z_3^{5*}$
Qualified	$z_2^{1*}, z_2^{2*}, z_3^{3*}, z_2^{4*}, z_3^{5*}$

factors are collected, and the collected data are discretized. According to the discretization results of the collected data, the Apriori algorithm is used to mine the relationship between assembly influencing factors and assembly performance and determine the range of influence of key factors so as to provide for subsequent process adjustments, as shown in Figure 1.

3.1. Identification of Key Influencing Factors Based on Complex Network and Entropy Weight Method. Complex products have complex assembly relationships due to the complexity of the structure and the number of parts involved in the assembly. In the multistage assembly process of complex products, the machining quality of the parts themselves will affect the assembly performance of the products, and the assembly process and assembly quality will also affect the assembly performance of the products. In the assembly process, the assembly errors caused by the part processing quality, assembly process, and assembly quality are continuously transferred and accumulated among the parts, and there are also coupling relationships among the influencing factors. Therefore, the coupling relationship of complex product assembly process can be modeled and analyzed by means of complex network directed graph. The quality and process factors in the assembly process are abstracted as the nodes of a complex network, and the interactions between the influencing factors are abstracted as the edges of the complex network, and the connection relationships are determined by the assembly sequence.

The set of influencing factors is the basis for building complex network models and analyzing key influencing factors. Therefore, it is necessary to clarify the influencing factors affecting the assembly performance from the assembly process of complex products and form a set of influencing factors for the assembly performance of complex products to provide a data basis for building complex networks. The assembly performance of complex products is influenced by the quality of parts processing, assembly process, and assembly quality, so it can be analyzed from these three aspects to form a set of assembly performance influencing factors, which is noted as M.

$$M_m = \{X^m, Y^m, Z^m\}, \tag{1}$$

$$X^m = \{x_1^m, x_2^m, \cdots, x_o^m\}, \tag{2}$$

Table 4: Frequent item set L_1.

Items	Support	Items	Support
z_1^{1*}	60%	z_3^{3*}	20%
z_2^{1*}	20%	z_1^{4*}	0
z_3^{1*}	20%	z_2^{4*}	100%
z_1^{2*}	40%	z_3^{4*}	0
z_2^{2*}	60%	z_1^{5*}	40%
z_3^{2*}	0	z_2^{5*}	0
z_1^{3*}	40%	z_3^{5*}	60%
z_2^{3*}	40%		

Table 5: Frequent item set L_2.

Items	Support	Items	Support
z_1^{1*}	60%	z_2^{3*}	40%
z_1^{2*}	40%	z_2^{4*}	100%
z_2^{2*}	60%	z_1^{5*}	40%
z_1^{3*}	40%	z_3^{5*}	60%

Table 6: Frequent item set L_3.

Items	Support	Items	Support
z_1^{1*}, z_1^{2*}	20%	z_2^{2*}, z_2^{3*}	20%
z_1^{1*}, z_1^{3*}	20%	z_2^{2*}, z_2^{4*}	60%
z_1^{1*}, z_2^{3*}	40%	z_2^{2*}, z_1^{5*}	20%
z_1^{1*}, z_2^{4*}	60%	z_2^{2*}, z_3^{5*}	40%
z_1^{1*}, z_1^{5*}	40%	z_1^{3*}, z_2^{4*}	40%
z_1^{1*}, z_3^{5*}	60%	z_1^{3*}, z_1^{5*}	0
z_1^{2*}, z_1^{3*}	20%	z_1^{3*}, z_3^{5*}	40%
z_1^{2*}, z_2^{3*}	20%	z_2^{3*}, z_2^{4*}	40%
z_1^{2*}, z_2^{4*}	40%	z_2^{3*}, z_1^{5*}	40%
z_1^{2*}, z_1^{5*}	20%	z_2^{3*}, z_3^{5*}	0
z_1^{2*}, z_3^{5*}	20%	z_2^{4*}, z_1^{5*}	40%
z_2^{2*}, z_1^{3*}	20%	z_2^{4*}, z_3^{5*}	60%

$$Y_m = \left\{y_1^m, y_2^m, \cdots, y_p^m\right\}, \quad (3)$$

$$Z_m = \left\{z_1^m, z_2^m, \cdots, z_q^m\right\}. \quad (4)$$

In the formula, M is the set of influencing factors. X is the processing quality influencing factor. Y is the assembly process influencing factor. Z is the assembly quality influencing factor. m means there are m processes.

The network relationship mapping between the factors is established according to the assembly order, as shown in Figure 2.

Table 7: Frequent item set L_4.

Items	Support	Items	Support
z_1^{1*}, z_2^{3*}	40%	z_1^{3*}, z_2^{4*}	40%
z_1^{1*}, z_2^{4*}	60%	z_1^{3*}, z_3^{5*}	40%
z_1^{1*}, z_1^{5*}	40%	z_2^{3*}, z_2^{4*}	40%
z_1^{1*}, z_3^{5*}	60%	z_2^{3*}, z_1^{5*}	40%
z_1^{2*}, z_2^{4*}	40%	z_2^{4*}, z_1^{5*}	40%
z_2^{2*}, z_2^{4*}	60%	z_2^{4*}, z_3^{5*}	60%
z_2^{2*}, z_3^{5*}	40%		

Complex networks, as a branch of complex system theory, are topological abstractions of real complex systems, and the theory and methods of complex networks can well describe the topological characteristics, functional properties, and interrelationships of systems using complex networks . Based on the theory and method of complex network, we construct a network model of association relationship reflecting the characteristics of the system and build a node importance evaluation index system with the node attributes of complex network.

The different connection relationships of the nodes in a complex network can make the node importance inconsistent, where important nodes are the special nodes in a complex network that affect the structure and function of the network [20]. The connection relationship of the nodes in the complex network reflects the inherent attribute information of the nodes. This paper selects seven network characteristics of each influencing factor to construct an evaluation index system for the importance of each process influencing factor. These network characteristics are degree centrality, aggregation coefficient, mesoscopic centrality, proximity centrality, centrifugal centrality, eigenvector centrality, and average neighbor degree. Table 1 shows the commonly used calculation methods for characteristic analysis in complex networks [21].

The complex network model of factors influencing the assembly performance of complex products is constructed in the following steps.

(Step 1) According to the analysis of complex product assembly performance influencing factors, the influencing factors are used as nodes of the complex network model. $V_m = \{v_i^m\}, (m = 1, 2, 3, 4; i = 1, 2, \cdots, N)$ represents the set of influencing factors for each process, where v_i^m represents the i-th node of the m-th process complex network model. m is the number of processes. N is the number of influencing factors of the m-th process.

(Step 2) Edge of complex network model based on the interaction influence relationship among the influencing factors. $E_m = \{e_{ij}^m\}, (m = 1, 2, 3, 4; i, j = 1, 2, \cdots, N)$ represents the set of influence relationships between the influencing factors

Table 8: Frequent item set L_5.

Items	Support	Items	Support
$z_1^{1*}, z_2^{3*}, z_2^{4*}$	40%	$z_2^{2*}, z_2^{4*}, z_1^{5*}$	20%
$z_1^{1*}, z_2^{3*}, z_1^{5*}$	40%	$z_2^{2*}, z_2^{4*}, z_3^{5*}$	40%
$z_1^{1*}, z_2^{3*}, z_3^{5*}$	0	$z_1^{3*}, z_2^{4*}, z_1^{5*}$	0
$z_1^{1*}, z_2^{4*}, z_1^{5*}$	40%	$z_1^{3*}, z_2^{4*}, z_3^{5*}$	40%
$z_1^{1*}, z_2^{4*}, z_3^{5*}$	20%	$z_2^{3*}, z_2^{4*}, z_1^{5*}$	40%
$z_1^{2*}, z_2^{4*}, z_1^{5*}$	20%	$z_2^{3*}, z_2^{4*}, z_3^{5*}$	0
$z_1^{2*}, z_2^{4*}, z_3^{5*}$	20%		

Table 9: Frequent item set L_6.

Items	Support	Items	Support
$z_1^{1*}, z_2^{3*}, z_2^{4*}$	40%	$z_2^{2*}, z_2^{4*}, z_3^{5*}$	40%
$z_1^{1*}, z_2^{3*}, z_1^{5*}$	40%	$z_1^{3*}, z_2^{4*}, z_3^{5*}$	40%
$z_1^{1*}, z_2^{4*}, z_1^{5*}$	40%	$z_2^{3*}, z_2^{4*}, z_1^{5*}$	40%

Table 10: Frequent item set L_7.

Items	Support	Items	Support
$z_1^{1*}, z_2^{3*}, z_2^{4*}, z_1^{5*}$	40%	$z_1^{1*}, z_2^{3*}, z_2^{4*}, z_3^{5*}$	0

Table 11: Frequent item set L_8.

Items	Support
$z_1^{1*}, z_2^{3*}, z_2^{4*}, z_1^{5*}$	40%

of each process. Where e_{ij}^m represents the existence of an interaction between the i-th node and the j-th node in the m-th process.

(Step 3) Draw complex network diagrams. Form a connection matrix based on the above two steps, and draw a complex network diagram based on the connection matrix.

(Step 4) The inherent properties of each node are calculated. Calculate the local and global information of each node as described in Table 1.

After the complex network model is built, the key nodes in it need to be identified. More research has been conducted on the identification of important nodes in complex networks, mainly including degree centrality, feature vector centrality, mediator centrality, and Paerank method [22]. Each of the above algorithms only evaluates the node importance from one aspect and has some limitations. Therefore, this paper adopts the entropy-weight-TOPSIS model to evaluate the node importance by considering the local attribute information as well as the global attribute information of the network and then identifies the important nodes of the complex network.

The entropy-TOPSIS model is a combination of the entropy and TOPSIS methods [23]. The TOPSIS method is a multi-indicator evaluation algorithm, which is a kind of ranking by judging the distance of each target from the ideal target. Since TOPSIS will involve multiple indicators in the operation, the weight of each indicator is not consistent; if the weight of each indicator is not determined, it will affect the accuracy of the ranking results. As an objective assignment method, the entropy method does not depend on human experience, so the entropy method is combined with TOPSIS method to achieve the identification of important nodes in complex networks. The specific steps for the identification of key influencing factors of product performance based on the entropy-weight-TOPSIS model are as follows:

(Step 1) The inherent properties of each influencing factor are calculated, and the corresponding primitive matrix is constructed based on the established complex network. Suppose with k indicators $X = \{X_1, X_2, \cdots, X_k\}$, where $X_i = \{x_{1i}, x_{2i}, \cdots, x_{ni}\}$. Indicator X_i has n sets of data. That is, there are n evaluation objects and **k** evaluation indicators. The original data matrix X is as follows:

$$X = \begin{bmatrix} x_{11}, x_{12}, \cdots, x_{1k} \\ x_{21}, x_{22}, \cdots, x_{2k} \\ \cdots \\ x_{n1}, x_{n2}, \cdots, x_{nk} \end{bmatrix}. \tag{5}$$

(Step 2) In data standardization, since the different meanings of the indicators will make the difference in the scale of each indicator and affect the evaluation results, the original data matrix needs to be standardized. There are positive and negative indicators in the standardization process. A positive indicator means that the larger the value, the more important the indicator is, and a negative indicator means that the lower the value, the more important the indicator is, so different methods are used to standardize data for different indicators. Where degree centrality, aggregation coefficient, mediator centrality, proximity centrality, eigenvector centrality, and average neighborliness are all positive indicators, so they are normalized according to equation (6). Centrifugal centrality is a negative indicator, normalized according to equation (7). The matrix $Y^i\, i = 1, 2, 3 \cdots$, after the standardization of each index is obtained, as shown in equation (8).

Table 12: Disc assembly process unbalance influence factor set.

Serial number	Influencing factors
1	Unbalance of the first-level disc
2	Cylindricity of the first-level disc shaft
3	Cylinder surface roughness of primary disc shaft
4	The flatness of the end surface of the first-stage disc relative to the rotation axis
5	End surface roughness of the first-level disc shaft
6	The degree of runout of the cylinder of the first-stage disc relative to the rotation axis
7	Cylinder surface roughness of the first-level disc spigot
8	The runout of the end face of the first-level disc spigot
9	Roughness of the end surface of the first-level disc
10	The actual size of the first-level disc spigot
11	Bolt quality of first and second grade
12	Unbalance of secondary disc
13	Cylindricity of the secondary disc shaft
14	Roughness of the secondary disc shaft
15	The flatness of the end surface of the secondary disc relative to the rotation axis
16	End surface roughness of the secondary disc shaft
17	The runout degree of the cylinder of the upper spigot of the secondary disc relative to the rotation axis
18	Cylinder surface roughness of matching spigot on the secondary disc
19	The runout degree of the end face of the matching spigot on the secondary disc relative to the rotating shaft
20	Roughness of the upper spigot end of the secondary disc
21	The actual size of the upper spigot of the secondary disc
22	Cylinder surface roughness of the bottom matching spigot of the secondary disc
23	Radial circle runout of the bottom matching stop cylinder of the secondary disc
24	The runout of the end face of the bottom matching stop of the secondary disc
25	Roughness of the bottom spigot of the secondary disc
26	The actual size of the bottom stop of the secondary disc
27	Second and third grade bolt quality
28	Three-level disc unbalance
29	Cylindrical runout of three-stage disc spigot
30	Cylindrical roughness of the three-stage disc with spigot
31	Three-stage disc with the end face runout degree of spigot
32	End surface roughness of three-stage disc spigot
33	Actual size of tertiary disc spigot
34	Bolt tightening torque during the first and second stage disc assembly process
35	The bolt tightening sequence during the first and second stage disc assembly process
36	The heating temperature of the secondary disc during the assembly process of the primary and secondary disc
37	Room temperature
38	Phase I and II plate disc installation phase
39	Bolt tightening torque during the assembly process of the second and third stage disc
40	The bolt tightening sequence in the second and third stage disc assembly process
41	The second and third stage disc installation phase
42	Coaxiality of the rotation axis
43	Parallelism of the end face of the first and second stage disc spigot
44	Cylinder coaxiality of the first and second stage disc spigot
45	The first and second level of the disc spigot assembly interference
46	Parallelism of the end face of the second and third stage disc spigot
47	Concentricity of cylindrical surface of second and third stage disc spigot
48	Second and third-level disc spigot assembly interference
49	The heating temperature of the third-level pan disc during the assembly process of the second and third-level pan disc

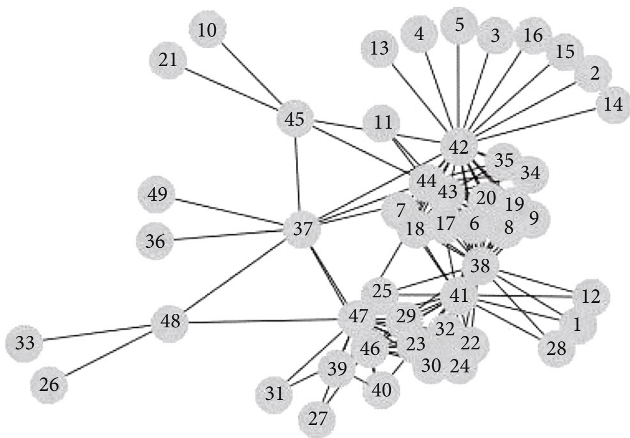

Figure 3: The complex network topology of the disc assembly process.

$$y_{ij} = \frac{x_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)}, \quad (6)$$

$$y_{ij} = \frac{\max(X_i) - x_{ij}}{\max(X_i) - \min(X_i)}, \quad (7)$$

$$Y = \begin{bmatrix} y_{11}, y_{12}, \cdots, y_{1k} \\ y_{21}, y_{22}, \cdots, y_{2k} \\ \cdots \\ y_{n1}, y_{n2}, \cdots, y_{nk} \end{bmatrix}. \quad (8)$$

(Step 3) Calculate the information entropy of each index. Calculate the information entropy of each index of different processes according to the formula of information entropy

$$E_i = -\ln(n)^{-1} \sum_{i=1}^{n} p_{ij} \ln p_{ij}. \quad (9)$$

In the formula, $p_{ij} = y_{ij}/\sum_{j=1}^{n} y_{ij}$. If $p_{ij} = 0$, then define $\lim_{p_{ij} \longrightarrow 0} p_{ij} \ln p_{ij} = 0$.

(Step 4) Calculate the weight of each indicator. Calculate the weight of each index W_k^i for each process according to the index weight calculation formula (10), $i = 1, 2, 3 \cdots$, $k = 1, 2, 3 \cdots$.

$$W_i = \frac{1 - E_i}{k - \sum E_i} \quad (i = 1, 2, \cdots, k). \quad (10)$$

(Step 5) Determine the positive ideal solution and the negative ideal solution. Based on the standardization of node attributes, the maximum value in each indicator data is the positive ideal solution A^+, and the minimum value in each indicator data is the negative ideal solution A^-. The details are as follows:

$$\begin{aligned} A^+ &= \{y_1^+, y_2^+, \cdots, y_k^+\}, \\ A^- &= \{y_1^-, y_2^-, \cdots, y_k^-\}. \end{aligned} \quad (11)$$

In the formula, $y_i^+ = \max_j y_{ij}$, $y_i^- = \min_j y_{ij}$.

(Step 6) Calculate the distance of each node from the ideal solution. The distance D_i^+ and D_i^- of each node evaluation index from positive ideal scenario A^+ and negative ideal scenario A^- is defined as follows:

$$\begin{aligned} D_i^+ &= \sqrt{\sum_{j=1}^{k} W_j \left(y_{ij} - y_j^+\right)^2}, \\ D_i^- &= \sqrt{\sum_{j=1}^{k} W_j \left(y_{ij} - y_j^-\right)^2}. \end{aligned} \quad (12)$$

(Step 7) Calculate the comprehensive evaluation index of each assessment object. Calculate the comprehensive evaluation index S based on the distance of each evaluation object from the positive and negative ideal solutions. The calculation formula is as follows:

$$S(i) = \frac{D_i^-}{D_i^+ + D_i^-}. \quad (13)$$

According to the comprehensive evaluation index of each influencing factor, the change curve of the importance of influencing factors is drawn and the key factors are selected. When there is a significant decrease in the weight, it indicates that there is a significant decrease in the importance of the later influencing factors relative to the preceding influencing factors. Therefore, when there is a significant decrease in the weight, the weight is used as the threshold for selecting the key influencing factors, and the factors greater than this threshold are the key influencing factors for performance.

3.2. Association Relationship Mining Based on Apriori Algorithm. In the research content of the previous section, the factors affecting the performance of complex products have been identified and the key factors have been extracted. In this part of the content, this paper proposes to use the Apriori method to analyze the credibility between the assembly performance of complex products and key

Table 13: The index value of the factors affecting the assembly process of the disc.

Influencing factors	Degree centrality	Clustering coefficient	Betweenness centrality	Proximity centrality	Eigenvector centrality	Eccentricity centrality	Average neighbor degree
1	0.042	1.000	0.000	0.079	23.500	0.400	4.000
2	0.021	0.000	0.000	0.033	23.000	0.393	4.000
3	0.021	0.000	0.000	0.033	23.000	0.393	4.000
4	0.021	0.000	0.000	0.033	23.000	0.393	4.000
5	0.021	0.000	0.000	0.033	23.000	0.393	4.000
6	0.104	0.500	0.003	0.165	19.600	0.471	4.000
7	0.104	0.500	0.003	0.165	19.600	0.471	4.000
8	0.104	0.500	0.003	0.165	19.600	0.471	4.000
9	0.104	0.500	0.003	0.165	19.600	0.471	4.000
10	0.021	0.000	0.000	0.008	5.000	0.310	4.000
11	0.042	0.000	0.000	0.053	14.000	0.364	4.000
12	0.042	1.000	0.000	0.079	23.500	0.400	4.000
13	0.021	0.000	0.000	0.033	23.000	0.393	4.000
14	0.021	0.000	0.000	0.033	23.000	0.393	4.000
15	0.021	0.000	0.000	0.033	23.000	0.393	4.000
16	0.021	0.000	0.000	0.033	23.000	0.393	4.000
17	0.104	0.500	0.003	0.165	19.600	0.471	4.000
18	0.104	0.500	0.003	0.165	19.600	0.471	4.000
19	0.104	0.500	0.003	0.165	19.600	0.471	4.000
20	0.104	0.500	0.003	0.165	19.600	0.471	4.000
21	0.021	0.000	0.000	0.008	5.000	0.310	4.000
22	0.083	0.500	0.001	0.117	18.750	0.436	4.000
23	0.083	0.500	0.001	0.117	18.750	0.436	4.000
24	0.083	0.500	0.001	0.117	18.750	0.436	4.000
25	0.083	0.500	0.001	0.117	18.750	0.436	4.000
26	0.021	0.000	0.000	0.004	4.000	0.293	4.000
27	0.042	0.000	0.000	0.038	14.000	0.372	4.000
28	0.042	1.000	0.000	0.079	23.500	0.400	4.000
29	0.083	0.500	0.001	0.117	18.750	0.436	4.000
30	0.083	0.500	0.001	0.117	18.750	0.436	4.000
31	0.042	0.000	0.000	0.038	14.000	0.372	4.000
32	0.083	0.500	0.001	0.117	18.750	0.436	4.000
33	0.021	0.000	0.000	0.004	4.000	0.293	4.000
34	0.104	0.800	0.001	0.143	16.000	0.457	4.000
35	0.104	0.800	0.001	0.143	16.000	0.457	4.000
36	0.021	0.000	0.000	0.015	9.000	0.350	4.000
37	0.188	0.139	0.171	0.139	10.000	0.533	3.000
38	0.500	0.181	0.183	0.402	6.583	0.578	4.000
39	0.083	0.833	0.000	0.084	13.750	0.417	4.000
40	0.083	0.833	0.000	0.084	13.750	0.417	4.000
41	0.479	0.146	0.134	0.348	5.826	0.527	4.000
42	0.479	0.103	0.390	0.311	5.478	0.640	3.000
43	0.271	0.141	0.041	0.231	6.538	0.485	3.000
44	0.313	0.238	0.075	0.272	7.533	0.552	3.000
45	0.104	0.300	0.082	0.078	9.800	0.444	3.000
46	0.271	0.128	0.054	0.163	5.538	0.490	3.000
47	0.313	0.114	0.197	0.200	6.600	0.571	3.000
48	0.083	0.167	0.082	0.036	6.500	0.410	3.000

Table 13: Continued.

Influencing factors	Degree centrality	Clustering coefficient	Betweenness centrality	Proximity centrality	Eigenvector centrality	Eccentricity centrality	Average neighbor degree
49	0.021	0.000	0.000	0.015	9.000	0.350	4.000

Table 14: The weights of different indexes in the assembly process of the disc.

Degree centrality	Clustering coefficient	Betweenness centrality	Proximity centrality	Eigenvector centrality	Eccentricity centrality	Average neighbor degree
0.134	0.106	0.305	0.062	0.026	0.040	0.326

Table 15: The positive and negative ideal plan of the disc assembly process.

Ideal solution	Degree centrality	Clustering coefficient	Betweenness centrality	Eccentricity centrality	Proximity centrality	Eigenvector centrality
Positive ideal solution	1	1	1	1	1	1
Negative ideal solution	0	0	0	0	0	0

influencing factors, reveal the correlation degree between the assembly performance and key influencing factors, determine the influence range of key factors, and form a correlation rule base to provide guidance for subsequent process optimization.

The association rule mining problem can be formally described as follows. Let $I = \{i_1, i_2, \cdots i_m\}$ be the set of all influencing factors. D is the set of all combinations of influencing factors. Each combination of influences T is the set of some range of values of influences. T is contained in I. Each combination of impact factors can be identified by a unique identifier TID. Let X_1, X_2 be the set of certain influencing factors. A combination of influences T is said to contain X_1 if $X_1 \subseteq T$. The association rule is expressed in the following form: implication of $(X_1 \subset T)X_1 \longrightarrow X_2($ $X_2 \subset T)$. Here, $X_1 \subset I$, $X_2 \subset I$, and $X_1 \cap X_2 = \Phi$. The rule $X_1 \longrightarrow X_2$ in the set D of the combination of influencing factors is bounded by the degree of support s and the degree of confidence c. The degree of confidence indicates the strength of the rule, and the degree of support indicates the frequency of occurrence in the rule. The support $s(X_1)$ of the data item set X_1 is the ratio of the number of combinations of influences containing X_1 in D to the total number of combinations of influences in D. The support s of rule $X_1 \longrightarrow X_2$ is defined as the proportion of the combination of influences containing $X_1 \cup X_2$ in D as s%, indicating the ratio of the number of combinations of influences containing both X_1 and X_2 to the total number of combinations of influences in D. The support s of rule $X_1 \longrightarrow X_2$ is defined as the degree to which c% of the combination of influences in D that contains X_1 also contains X_2, indicating how likely it is that the combination of influences in D that contains X_1 contains X_2.

Association rule mining is to find association rules with user-given minimum support minsup and minimum confidence minconf in the database D of influencing factor combinations. The association rule mining problem can be decomposed into two steps.

(1) Find all the itemsets in the influence factor combination database D that are greater than or equal to the user-specified minimum support. The set of combinations of influences with minimal support is called the set of frequent items

(2) Generate the required association rules using frequent item sets. For each frequent itemset A, find all nonempty subsets a of A. If the ratio support(A) /support$(a) >$ = minsup, generate association rule $a \longrightarrow (A - a)$. support(A)/support(a) is the confidence of association rule $a \longrightarrow (A - a)$

The Apriori algorithm is a classical algorithm for mining frequent itemsets of association rules. The algorithm uses an iterative method of layer-by-layer search to perform multiple scans of a combined database of influencing factors. First scan yields frequent 1-term set L_1. The result of the kth $(k > 1)$ scan is used before the kth scan (i.e., the frequent $k-1$ itemset) to generate the candidate k-item set C_k. The support of the elements in C_k is then determined during the scan. Finally, at the end of each scan, the frequent k-item set L_k is calculated. The algorithm ends when the candidate frequent k-item set C_k is empty. In this paper, the influencing factors are combined with variables corresponding to the performance of complex products to form a data item set Z in association rule analysis.

$$Z = \{x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_n\}. \tag{14}$$

The following correlation analysis of complex product

Table 16: Ranking results of key influencing factors of unbalanced quantity of disc assembly process.

Serial number	Influencing factors	Sort results	Weights (%)
1	Unbalance of the first level disc	11	2.447
2	Cylindricity of first-level disc shaft	33	1.396
3	Cylinder surface roughness of primary disc shaft	34	1.396
4	The flatness of the end surface of the first-stage disc relative to the rotation axis	35	1.396
5	End surface roughness of first-level disc shaft	36	1.396
6	The degree of runout of the cylinder of the first-stage disc relative to the rotation axis	18	1.906
7	Cylinder surface roughness of first-level disc spigot	19	1.906
8	The runout of the end face of the first-level disc spigot	20	1.906
9	Roughness of the end surface of the first-level disc	21	1.906
10	The actual size of the first-level disc spigot	46	0.107
11	First and second grade bolt quality	41	0.851
12	Unbalance of secondary disc	12	2.447
13	Cylindricity of secondary disc shaft	37	1.396
14	Roughness of secondary disc shaft	38	1.396
15	The flatness of the end surface of the secondary disc relative to the rotation axis	39	1.396
16	End surface roughness of secondary disc shaft	40	1.396
17	The runout degree of the cylinder of the upper spigot of the secondary disc relative to the rotation axis	22	1.906
18	Cylinder surface roughness of matching spigot on the secondary disc	23	1.906
19	The runout degree of the end face of the matching spigot on the secondary disc relative to the rotating shaft	24	1.906
20	Roughness of the upper spigot end of the secondary disc	25	1.906
21	The actual size of the upper spigot of the secondary disc	47	0.107
22	Cylinder surface roughness of the bottom matching spigot of the secondary disc	26	1.756
23	Radial circle runout of the bottom matching stop cylinder of the secondary disc	27	1.756
24	The runout degree of the end face of the lower matching spigot of the secondary disc relative to the rotating shaft	28	1.756
25	Roughness of the bottom spigot of the secondary disc	29	1.756
26	The actual size of the bottom stop of the secondary disc	48	0.000
27	Second and third grade bolt quality	42	0.844
28	Three-level disc unbalance	13	2.447
29	Cylindrical runout of three-stage disc spigot	30	1.756
30	Cylindrical roughness of three-stage disc with spigot	31	1.756
31	Three-stage disc with end face runout degree of spigot	43	0.844
32	End surface roughness of three-stage disc spigot	32	1.756
33	Actual size of tertiary disc spigot	49	0.000
34	Bolt tightening torque during the first and second stage disc assembly process	14	2.156
35	The bolt tightening sequence during the first and second stage disc assembly process	15	2.156
36	The heating temperature of the secondary disc during the assembly process of the primary and secondary disc	44	0.454
37	Room temperature	3	4.475
38	Phase I and II plate disc installation phase	9	3.459
39	Bolt tightening torque during the assembly process of the second and third stage disc	16	2.055
40	The bolt tightening sequence in the second and third stage disc assembly process	17	2.055
41	Second and third stage disc installation phase	10	3.100
42	Coaxiality of the rotation axis	1	5.838
43	Parallelism of the end face of the first and second stage disc spigot	5	4.043
44	Cylinder coaxiality of the first and second stage disc spigot	4	4.382
45	The first and second level of the disc spigot assembly interference	7	3.951

Table 16: Continued.

Serial number	Influencing factors	Sort results	Weights (%)
46	Parallelism of the end face of the second and third stage disc spigot	6	4.026
47	Concentricity of cylindrical surface of second and third stage disc spigot	2	4.824
48	Second and third-level disc spigot assembly interference	8	3.775
49	The heating temperature of the third-level pan disc during the assembly process of the second and third-level pan disc	45	0.454

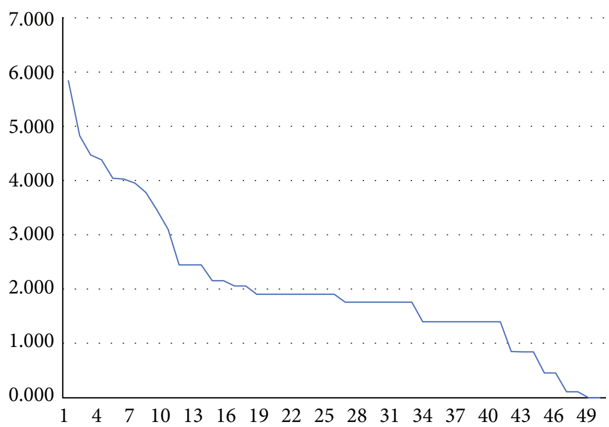

Figure 4: Changes in the weights of influencing factors in the disc assembly process.

performance and influencing factors is to analyze the correlation between the influencing factor set $X = \{x_1, x_2, \cdots, x_m\}$ and the performance set $Y = \{y_1, y_2, \cdots, y_n\}$. The credibility of the influencing factors is calculated to find $X \longrightarrow Y$.

However, association rule analysis algorithms are all for discrete data, and many factors in $x_1 \sim x_n$ are continuous values. Therefore, it is necessary to perform data preprocessing on the data of each influencing factor. Before mining association rules, all data is discretized, and continuous values are mapped into multiple discrete values.

There are many discretization methods for continuous values. Typical discretization algorithms are as follows: equal width or equal frequency method [13], C4.5 method [24], entropy method [25], and Chi-Merge algorithm [26]. This paper uses the Chi-Merge algorithm to discretize the data.

The Chi-Merge algorithm is a supervised discretization method based on the Chi-square distribution (represented by the symbol χ^2). Using a bottom-up strategy, the best neighboring interval is found recursively, and then they are merged to form a larger interval. The process is as follows:

(1) Sorting the data in ascending order

(2) Defining the initial interval so that each data is in a separate interval

(3) Repeating until the χ^2 of any two adjacent intervals is not less than the threshold determined by the specified confidence level

After Chi-Merge discretization preprocessing, the influence factor database Z with continuous values can be transformed into the influence factor database Z^* of Boolean type. The form of the data item set is as follows:

$$Z^* = \left\{ z_1^{1*}, z_2^{1*}, \cdots, z_1^{i*}, \cdots, z_j^{i*}, \cdots, z_1^{n*}, \cdots \right\}. \quad (15)$$

In the formula, Z^* denotes the database of influencing factors after Chi-Merge discretization. z_j^{i*} denotes the mapping value of each Chi-Merge discretized interval, as shown in Table 2. $F_{\mathrm{CM}}(x_i)$ in Table 2 indicates the mapping result of Chi-Merge discretization of x_i. i^* denotes the value of the interval mapping after discretization of the ith feature x_i. j denotes the ith feature x_i discretized into j interval mapping values.

After Chi-Merge discretization, all continuous attributes are discretized into discrete values and participate in the subsequent association rule mining as an item set.

Apriori association rule mining algorithm trial calculation process is as follows. Suppose there is a database Z with 5 transaction records, as shown in Table 3. Assume $n = 5$; that is, there are 5 eigenvolumes, each discretized into 3 intervals. The minimum support is set tominsup = 40%.

(i) Database Z

(ii) Scan database Z to obtain the frequent item set L_1, as shown in the following Table 4

(iii) Remove the set of items smaller than the minimum expenditure degree minsup to obtain the frequent item set L_2, as shown in Table 5

(iv) Scan frequent item set L_2 to obtain the frequent item set L_3, as shown in Table 6

(v) Remove the set of items smaller than the minimum expenditure degree minsup to obtain the frequent item set L_4, as shown in Table 7

Table 17: Sorting results of key influencing factors of disc assembly process.

Serial number	Influencing factors	Weights	Classification
1	Coaxiality of the rotation axis	5.838	Assembly quality
2	Concentricity of cylindrical surface of second and third stage disc spigot	4.824	Assembly quality
3	Room temperature	4.475	Processing quality
4	Cylinder coaxiality of the first and second stage disc spigot	4.382	Assembly quality
5	Parallelism of the end face of the first and second stage disc spigot	4.043	Processing quality
6	Parallelism of the end face of the second and third stage disc spigot	4.026	Processing quality
7	The first and second level of the disc spigot assembly interference	3.951	Assembly quality
8	Second and third-level disc spigot assembly interference	3.775	Assembly quality
9	Phase I and II plate disc installation phase	3.459	Assembly process
10	Second and third stage disc installation phase	3.100	Assembly process
11	Unbalance of the first level disc	2.447	Processing quality
12	Unbalance of secondary disc	2.447	Processing quality
13	Three-level disc unbalance	2.447	Processing quality
14	Bolt tightening torque during the first and second stage disc assembly process	2.156	Assembly process
15	Bolt tightening torque during the assembly process of the second and third stage disc	2.055	Assembly process
16	The degree of runout of the cylinder of the first-stage disc relative to the rotation axis	1.906	Processing quality
17	Cylinder surface roughness of first-level disc spigot	1.906	Processing quality
18	The runout of the end face of the first-level disc spigot	1.906	Processing quality
19	Roughness of the end surface of the first-level disc	1.906	Processing quality
20	The runout degree of the cylinder of the upper spigot of the secondary disc relative to the rotation axis	1.906	Processing quality
21	Cylinder surface roughness of matching spigot on the secondary disc	1.906	Processing quality
22	The runout degree of the end face of the matching spigot on the secondary disc relative to the rotating shaft	1.906	Processing quality
23	Roughness of the upper spigot end of the secondary disc	1.906	Processing quality
24	Cylinder surface roughness of the bottom matching spigot of the secondary disc	1.756	Processing quality
25	Radial circle runout of the bottom matching stop cylinder of the secondary disc	1.756	Processing quality
26	The runout degree of the end face of the lower matching spigot of the secondary disc relative to the rotating shaft	1.756	Processing quality

Table 17: Continued.

Serial number	Influencing factors	Weights	Classification
27	Roughness of the bottom spigot of the secondary disc	1.756	Processing quality
28	Cylindrical runout of three-stage disc spigot	1.756	Processing quality
29	Cylindrical roughness of three-stage disc with spigot	1.756	Processing quality
30	End surface roughness of three-stage disc spigot	1.756	Processing quality

Table 18: Key factors of unbalanced out-of-tolerance.

Variable	The key factor
x_1	Coaxiality of the rotation axis
x_2	Concentricity of cylindrical surface of second and third stage disc spigot
x_3	Room temperature
x_4	Cylinder coaxiality of the first and second stage disc spigot
x_5	Parallelism of the end face of the first and second stage disc spigot
x_6	Parallelism of the end face of the second and third stage disc spigot
x_7	The first and second level of the disc spigot assembly interference
x_8	Second and third-level disc spigot assembly interference
x_9	Phase I and II plate disc installation phase
x_{10}	Second and third stage disc installation phase
x_{11}	Unbalance of the first level disc
x_{12}	Unbalance of secondary disc
x_{13}	Three-level disc unbalance
x_{14}	Bolt tightening torque during the first and second stage disc assembly process
x_{15}	Bolt tightening torque during the assembly process of the second and third stage disc
x_{16}	The degree of runout of the cylinder of the first-stage disc relative to the rotation axis
x_{17}	Cylinder surface roughness of first-level disc spigot
x_{18}	The runout of the end face of the first-level disc spigot
x_{19}	Roughness of the end surface of the first-level disc
x_{20}	The runout degree of the cylinder of the upper spigot of the secondary disc relative to the rotation axis
x_{21}	Cylinder surface roughness of matching spigot on the secondary disc
x_{22}	The runout degree of the end face of the matching spigot on the secondary disc relative to the rotating shaft
x_{23}	Roughness of the upper spigot end of the secondary disc
x_{24}	Cylinder surface roughness of the bottom matching spigot of the secondary disc
x_{25}	Radial circle runout of the bottom matching stop cylinder of the secondary disc
x_{26}	The runout degree of the end face of the lower matching spigot of the secondary disc relative to the rotating shaft
x_{27}	Roughness of the bottom spigot of the secondary disc
x_{28}	Cylindrical runout of three-stage disc spigot
x_{29}	Cylindrical roughness of three-stage disc with spigot
x_{30}	End surface roughness of three-stage disc spigot

Table 19: Discretization results of multivalue continuous attributes.

The key factor	Interval range	Quantity	Mapped value	The key factor	Interval range	Quantity	Mapped value
	0.000~0.005	41	0		0.000~0.008	23	0
	0.005~0.007	13	1		0.008~0.017	45	1
x_1	0.007~0.010	32	2	x_2	0.017~0.019	18	2
	0.010~0.012	25	3		0.019~0.024	33	3
	0.012~0.015	39	4		0.024~0.030	31	4
⋮							
⋮							
	0.00~0.10	15	0		0.00~0.19	34	0
	0.10~0.35	47	1		0.19~0.31	26	1
x_{29}	0.35~0.53	31	2	x_{30}	0.31~0.39	14	2
	0.53~0.72	36	3		0.39~0.62	43	3
	0.72~0.80	21	4		0.62~0.80	33	4

Table 20: Strong association rules when the imbalance is out of tolerance.

$F_{CM}(x_i)$	Support:0.2	Confidence
$64791 < x_3 < 65141$		
$0.0044 < x_{10} < 0.0058$	$\Longrightarrow$ Unbalanced amount out of tolerance	0.9
$9.84 < x_{13} < 9.94$		
$0.0571 < x_{11} < 0.0852$		
$65151 < x_3 < 67177$	$\Longrightarrow$ Unbalanced amount out of tolerance	0.8
$9.84 < x_{13} < 9.94$		
⋮	⋮	⋮
$75 < x_{13} < 80$		
$64791 < x_3 < 65141$	$\Longrightarrow$ Unbalanced amount out of tolerance	0.7
$49.02 < x_4 < 50.00$		

(vi) Scan frequent item set L_4 to obtain the frequent item set L_5, as shown in Table 8

(vii) Remove the set of items smaller than the minimum expenditure degree minsup to obtain the frequent item set L_6, as shown in Table 9

(viii) Scan frequent item set L_6 to obtain the frequent item set L_7, as shown in Table 10

(ix) Remove the set of items smaller than the minimum expenditure degree minsup to obtain the frequent item set L_8, as shown in Table 11

When the frequent item set L_4 is scanned again by the Apriori algorithm, the frequent item set L_4 is empty and the algorithm trial process is ended. The association rules and their support and confidence levels are obtained through algorithmic trial calculations. By setting the minimum support and confidence level and retaining the association rules that meet the requirements, a library of association rules is obtained for guiding the plant to make process adjustments.

4. Case Analysis

This paper analyzes an example of a certain assembly process of a certain type of aeroengine low-pressure fan rotor. The optional influencing factors are processing quality, assembly process, and assembly quality, a total of 49 items, and the performance index is the unbalance of the fan rotor.

This article uses Python's open source software library NetworkX to model complex networks. The library contains visualization and analysis algorithms for complex networks, which can visualize complex networks and analyze data. Taking the assembly of the primary and secondary discs of the low pressure fan rotor as an example, the set of influencing factors is shown in Table 12.

The complex network visualization model of the correlation model of the low-pressure rotor unbalance influencing factors is shown in Figure 3.

For the above established correlation relationship model of each process unbalance influence factors, the attribute information of each node is calculated and the results are shown in Table 13.

According to the calculation formulas (10), (11), and (13), the calculation results are shown in Tables 14 and 15, and the weight of each influencing factor is shown in Table 16.

It can be seen from the curve in Figure 4 that the threshold of the disc assembly process is 1.760, and the key influencing factor identification results are shown in Table 17.

As shown in Table 18, this paper extracts 30 key influencing factors related to the imbalance out-of-tolerance according to the key factor identification in the previous part, and the corresponding variables are $x_1 \sim x_{30}$. According to the definition of association rules, they are all important parameters that characterize the performance of low-pressure rotors, and they are all related to each other.

A total of 150 sets of data with out-of-tolerance imbalances have been collected in this paper. The results of using the Chi-Merge discretization algorithm to discretize the continuous data are shown in Table 19.

Taking the discrete data as a new item set, using Apriori association rule analysis, some more detailed association rules can be obtained, and these rules can also be used for process adjustment in the assembly process. Table 20 is a part of the association rules with higher credibility after association rule mining.

Obviously, the obtained association rules with higher credibility can be used to make a preliminary judgment on whether the imbalance is out of tolerance and to provide decision support for process adjustment based on this, so as to improve the success rate of one-time assembly of the low-pressure rotor.

5. Conclusion

In this paper, we propose a key influence factor identification method based on the combination of complex network and entropy power method to achieve the identification of key influence factors for complex product assembly performance and the scope of influence. Based on the key factor identification results, the Apriori algorithm is used to mine the association rules between key influencing factors and the assembly performance of complex products. Several association rules have been calculated and analyzed through examples, and a library of association rules has been formed, which can provide decision support for process adjustment and has been used with good results in the actual application in aeroengine assembly plants. The method in this paper is not limited to the adjustment and optimization of the assembly performance of complex products; it can also be combined with MEMS [27], mobile robots [28], and quadrotors [29] in the subsequent research process to better optimize their control by identifying key factors and mining association rules. However, when the Apriori algorithm is used for association rule mining, the database needs to be scanned every time the set of candidate items is generated, and the size of the database used for association rule mining is usually relatively large, which makes the algorithm inefficient. The efficiency of the algorithm will be improved by improving the Apriori algorithm during the subsequent research.

Data Availability

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the National Key R&D Program of China (No. 2019YFB1703802).

References

[1] W. B. Tang, J. F. Yu, Y. Li, and S. L. Tang, "Application of quantitative identification and decomposition methods for product key characteristics," *Computer Integrated Manufacturing System*, vol. 17, no. 11, pp. 2383–2388, 2011.

[2] S. Zhao, S. L. Xie, Z. P. Deng, and X. Huang, "Research on key characteristic identification and control method based on assembly process," *Aviation Manufacturing Technology*, vol. 8, pp. 56–59, 2016.

[3] L. Xu and X. Su, "Identification method of key quality elements of complex quality structure product system based on Bayesian network," *Journal of Jiangsu University of Science and Technology (Natural Science)*, vol. 29, no. 2, pp. 193–198, 2015.

[4] D. E. Whitney, "The role of key characteristics in the design of mechanical assemblies," *Assembly Automation*, vol. 26, no. 4, pp. 315–322, 2006.

[5] F. Y. Guo, *Research on Aircraft Digital Assembly Coordination Technology, [Ph D Thesis]*, Northwestern Polytechnical University, China, 2015.

[6] M. G. Raman, K. Kirthivasan, and V. S. Sriram, "Development of rough set–hypergraph technique for key feature identification in intrusion detection systems," *Computers & Electrical Engineering*, vol. 59, pp. 189–200, 2017.

[7] F. Yang, "A key characteristics-based model for quality assurance in supply chain," in *2011 IEEE 18th International Conference on Industrial Engineering and Engineering Management*, pp. 1428–1432, Changchun, China, 2011.

[8] L. Y. Zheng, C. A. McMahon, L. Li, L. Ding, and J. Jamshidi, "Key characteristics management in product lifecycle management: a survey of methodologies and practices," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 222, no. 8, pp. 989–1008, 2008.

[9] L. Zhong, C. Gao, Z. Zhang, N. Shi, and J. Huang, "Identifying influential nodes in complex networks: a multiple attributes fusion method," in *International conference on active media technology*, D. Ślęzak, G. Schaefer, S. T. Vuong, and Y. S. Kim, Eds., vol. 8610 of Lecture Notes in Computer Science, , pp. 11–22, Springer, Cham, 2014.

[10] J. Jin, K. Xu, N. Xiong, Y. Liu, and G. Li, "Multi-index evaluation algorithm based on principal component analysis for node importance in complex networks," *IET Networks*, vol. 1, no. 3, pp. 108–115, 2012.

[11] Z. Y. Zhao, X. R. Meng, and R. N. Sun, "Nodes importance ranking method based on multi-attribute evaluation and deletion," *Computer Engineering*, vol. 44, no. 6, pp. 62–67, 2018.

[12] Y. C. Ma, "Improvement of Apriori algorithm in mining association rules," *Computer Applications and Software*, vol. 21, no. 11, pp. 82-83, 2004.

[13] U. D. Alan and D. Birant, "Server-based intelligent public transportation system with NFC," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 1, pp. 30–46, 2018.

[14] Y. F. Guo, H. L. Gao, and J. Tian, "Photovoltaic output modeling by introducing clustering analysis and its application in reliability evaluation," *Automation of Electric Power Systems*, vol. 40, no. 23, pp. 93–100, 2016.

[15] Y. H. Li, J. X. Wang, and X. L. Wang, "A power system network loss evaluation method based on hybrid clustering analysis," *Automation of Electric Power Systems*, vol. 40, no. 1, pp. 60–65, 2016.

[16] X. Li, Q. Gong, Q. Yang, S. Luo, and S. Li, "Fault information management and analysis system based on data mining technology for relay protection devices," *Electric Power Automation Equipment*, vol. 31, no. 9, pp. 88–91, 2011.

[17] Y. Zhang, C. Hu, S. Huang, S. Feng, and G. Lin, "Apriori algorithm based data mining and analysis method for secondary device defects," *Automation of Electric Power Systems*, vol. 41, no. 19, pp. 147–151, 2017.

[18] S. Y. Chen and M. Cao, "Research and application of multi-association rule mining algorithm," *Computer Engineering and Design*, vol. 31, no. 4, pp. 885–888, 2010.

[19] Y. Jia, Q. Shao, and J. S. Zhang, "Correlation analysis of civil aviation bird strike events based on FP growth algorithm," *Journal of Safety and Environment*, vol. 16, no. 1, pp. 110–114, 2016.

[20] H. Ming, Z. Yanli, L. Mingyue, L. Zhihui, and G. Zheng, "Critical node identification of a power grid based on multi-attribute decision," *Complex Systems and Complexity Science*, vol. 17, no. 3, p. 11, 2020.

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: structure and dynamics," *Complex Systems and Complexity Science*, vol. 424, no. 4-5, pp. 175–308, 2006.

[22] M. Hongming, "Identification method for nodes importance in multi-attribute based complex networks," *Identification Method for Nodes Importance in Multi-Attribute Based Complex Networks*, vol. 32, no. 5, pp. 68–74, 2018.

[23] L. Ren, H. T. Zhang, and M. Z. Wei, "Research on the evaluation of development level of Smart City based on entropy TOPSIS model," *Information Studies:Theory & Application*, vol. 42, no. 7, pp. 113–118+125, 2019.

[24] H. Q. Liang, "Research on the discretization methods for numerical attributes," *Information Technology*, vol. 5, pp. 99–101, 2008.

[25] Z. Jingxian, N. Chunpeng, and Z. Yuanrui, "Efficient discretization algorithm for continuous attributes," *Systems Engineering and Electronics*, vol. 31, no. 1, pp. 195–199, 2009.

[26] R. Kerber, "Chimerge-discretization of numeric attributes," in *Proceedings of the tenth national conference on Artificial intelligence*, pp. 123–128, San Jose, California, 1992.

[27] X. Shao, Y. Shi, and W. Zhang, "Input-and-measurement event-triggered output-feedback chattering reduction control for MEMS gyroscopes," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 9, pp. 5579–5590, 2021.

[28] X. L. Shao, J. T. Zhang, and W. D. Zhang, "Distributed cooperative surrounding control for mobile robots with uncertainties and aperiodic sampling," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2022.

[29] W. Zhang, X. Shao, W. Zhang, J. Qi, and H. Li, "Unknown input observer-based appointed-time funnel control for quadrotors," *Aerospace Science and Technology*, vol. 126, p. 107351, 2022.